ICAROB 2025

PROCEEDINGS OF THE 2025 INTERNATIONAL CONFERENCE ON ARTIFICIAL LIFE AND ROBOTICS

February 13 to 16, 2025
J:COM HorutoHall, Oita, Japan
30th AROB International Meeting Series

Editor-in-Chief
Masanori Sugisaka
Editors: Yingmin Jia, Takao Ito, Ju-Jang Lee
ISBN 978-4-9913337-2-9

Proceedings of The 2025 International Conference on

ARTIFICIAL LIFE AND ROBOTICS

(ICAROB2025)

February 13 to 16, J:COM HorutoHall, Oita, Japan, 2025
30th AROB International Meeting Series

Editor-in-Chief
Masanori Sugisaka
Editors: Yingmin Jia, Takao Ito, Ju-Jang Lee
ISBN 978-4-9913337-2-9

Contents

SPONSERED

ALife Robotics Corporation Ltd.

ORGANIZED BY

International Steering Committee of International Conference on Artificial Life and Robotics (ICAROB)

SUPPORTED BY

IEEE Fukuoka Section (Japan)

CO-ORGANIZED BY

Chinese Association for Artificial Intelligence (CAAI, P. R. China)

ADVISORY COMMITTEE CHAIRS

Kazuo Ishii (Kyushu Institute of Technology, Japan)
Kensuke Harada (Osaka University, Japan)
Takashi Kohno (University of Tokyo, Japan)
Eiji Hayashi (Kyushu Institute of Technology, Japan)
Jeffrey Johnson (The Open University, UK)
Katia Passerini (Seton Hall University, USA)

ADVISORY COMMITTEE

Bruce Eisenstein (Drexel University, Former IEEE President, USA)
Hidenori Kimura (RIKEN & Wasada University, Japan)
Joshua M. Epstein (The Johns Hopkins University, USA)
Ken-ichi Tanaka (Executive Fellow, Mitsubishi Electric Corporation, Japan)
Masato Nakagawa (Executive Fellow, DENSO CORPORATION, Visiting Professor of Hiroshima University, Japan)
Moshe Kam (New Jersey Institute of Technology, Former IEEE President, USA)
Paul Kalata (Drexel University, USA)
Paul Oh (Drexel University, USA)
Peter Herczfeld (Drexel University, USA)

GENERAL CHAIR

Masanori Sugisaka
(Alife Robotics Corporation Ltd., Japan)

CO-GENERAL CHAIRS

Yingmin Jia (Beihang Univesty, P. R. China)
Takao Ito (Hiroshima University, Japan)
Ju-Jang Lee (Honorary professor, KAIST, Korea)

VICE GENERAL CHAIR

Ang, Chun Kit (University of UCSI, Malaysia)
Henrik. H. Lund (Technical University of Denmark, Denmark)
John. L. Casti (International Institute for Applied Systems Analysis, Austria)
Luigi Pagliarini (Technical University of Denmark, Denmark)
(Academy of Fine Arts of Macerata, Italy)
Mohd Rizon bin Juhari (University of UCSI, Malaysia)
Katia Passerini (Seton Hall University, USA)
Norrima Mokhtar (University of Malaya, Malaysia)

PROGRAM CHAIRMAN

Makoto Sakamoto (University of Miyazaki, Japan)

CO-PROGRAM CHAIR

Marion Oswald (Technische Universität Wien, Austria)

INTERNATIONAL ORGANIZING COMMITTEE

Akira Nakamura (Saitama Institute of Technology, Japan)
Akinori Abe (Chiba University, Japan)
Caoli Wang (University of Shanghai for Science and Technology, P. R. China)
Chan Gook Park (Seoul National University, Korea)
Evgeni Magid (Kazan Federal University, Russia)
Fuzhong Wang (Henan Polytechnic University, P. R. China)
Hazry Desa (University of Malaysia, Perlis, Malaysia)
Hidehiko Yamamoto (Gifu University, Japan)
Hiroki Tamura (The University of Miyazaki, Japan)
Hiroshi Kage (Mitsubishi Electric Corporation, Japan)
Hiroshi Matsuno (Yamaguchi University, Japan)
Jiwu Wang (Beijing Jiaotong University, P. R. China)
Jovana Jovic (CNRS-AIST JRL, Japan, France)
Junping Du (Beijing University of Posts and Telecommunications, P. R. China)
Katsunori Shimohara (Doshisha University, Japan)
Kazuo Ishii (Kyushu Institute of Technology, Japan)
Kenichi Tanaka (Nagasaki Institute of Applied Science, Japan)
Kenji Hashimoto (Waseda University, Japan)
Kunikazu Kobayashi (Aichi Prefectural University, Japan)
Kuo-Hsien Hsia (National Yunlin University of Science and Technology, Taiwan)
Kyungho Park (U.S. ARMY, Japan and USA)
Masao Kubo (National Defense Academy of Japan, Japan)
Maxim Sokolov (Innopolis University, Russia)
Mehta Rajiv (New Jersey Institute of Technology, USA)
Minoru Kumano (University of Miyazaki, Japan)
Peter Sapaty (Ukrainian Academy of Science, Ukraine)
Pierre Parrend (University of Strasbourg, France)
Qiang Cai (Beijing Technology and Business University, P. R. China)
Qu Yanbin (Harbin Institute of Technology, P. R. China)
Singo Mabu (Yamaguchi University, Japan)
Takashi Kohno (The University of Tokyo, Japan)
Takashi Ogata (Iwate Prefectural University)
Tetsuro Hattori (Kagawa University, Japan)
Thi Thi Zin (University of Miyazaki, Japan)
Thomas S. Ray (University of Oklahoma, USA)
Toru Yamamoto (Hiroshima University, Japan)

Victor Berdonosov (Komsomolsk-on-Amur State University of Technology, Russia)
Yasunari Yoshitomi (Kyoto Prefectural University, Japan)
Yi Chai (Chongqing University, P. R. China)
Yoshifumi Morita (Nagoya Institute of Technology, Japan)
Yoshiro Imai (Kagawa University, Japan)

INTERNATIONAL PROGRAM COMMITTEE

Abdul Rahman bin Dullah (Universiti Teknikal Malaysia Melaka, Malaysia)
Akinori Abe (Chiba University, Japan)
Akihiro Yamaguchi (Fukuoka Institute of Technology, Japan)
Akihito Kanai (Hosei University, Japan)
Ali Selamat (University of Technology of Malaysia, Malaysia)
Aminurafiuddin bin Zulkifli (Multimedia University, Malaysia)
Amornphun Phunopas (King Mongkut's University of Technology, North Bankok, Thailand)
Andre Rosendo (ShanghaiTech University, P. R. China)
Anne Jeannin-Girardon (University of Strasbourg, France)
Anton Shiriaev (Norwegian University of Science and Technology, Norway)
Artur Sagitov (Kazan Federal University, Russia)
Ashi Gautam (The University of Tokyo, Japan)
Atsuya Tange (The University of Tokyo, Japan)
Ayumu Tominaga (Kitakyushu National College of Technology, Japan)
Bin Zhang (Beijing University of Posts and Telecommunications, P. R. China)
Cecilia Zanni-Merk (INSA-Rouen, France)
Chaoli Wang (University of Shanghai for Science and Technology, P. R. China)
Chia-Nan Ko (Nan kai University of Technology, Taiwan)
Chung-Wen Hung (National Yunlin University of Science & Technology, Taiwan)
Congdao Han (Shanghai Institute of Technology, P. R. China)
Cui Zhang (Beihang University, P. R. China)
Cynthia Matuszek (University of Maryland, USA)
Donglian Qi (Zhejiang University, P. R. China)
Duangiai Jitkongchuen (Dhurakij Pundit University, Thailand)
Dunwei Gong (China University of Mining and Technology, P. R. China)
Jiao Jia (Beihang University, P. R. China)
Endra Joelianto (Bandung Institute of Technology, Indonesia)
Fabio Guigou (University of Strasbourg, France)
Fei Hao (Beihang University, P. R. China)
Fuzhong Wang (Henan Polytechnic University, P. R. China)
Haibo Li (Royal Institute of Technology, Sweden)
Haisheng Li (Beijing Technology and Business University, P. R. China)
Haruhisa Okuda (Mitsubishi Electric Corporation, Japan)
Haruka Tsuboi (The University of Miyazaki, Japan)
Haruna Matsushita (Kagawa University, Japan)
Heeje Kim (Pusan National University, Korea)
Hidetsugu Suto (Muroran Institute of Technology, Japan)
Hiroyuki Iizuka (Osaka University, Japan)
Hongbo Li (Tsinghua University, P. R. China)

Hongji Ma (Shandong University of Science and Technology, P. R. China)
Hongjiu Yang (Yanshan University, P. R. China)
Hongpeng Yin (Chongqing University, P. R. China)
Hussein Abbass (University of New South Wales, and ADFA, Australia)
I-Hsien Liu (National Cheng Kung Univrsiity, Taiwan)
Ilya Afanasyev (Innopolis University, Russia)
Istvan Harmati (Budapest Institute of Technology and Economics, Hungary)
Ivan Tanev (Doshisha University, Japan)
Jiandong Zhao (Beijing Jiaotong University, P. R. China)
JJ Merelo (University of Granada, Spain)
Joono Cheong (Korea University, Korea)
Julio Navarro Lara (University of Strasbourg, France)
Jun Kobayashi (Kyushu Institute of Technology, Japan)
Jung-Shian Li (National Cheng Kung University, Taiwan)
Keiji Kamei (Nishinippon Institute of Technology, Japan)
Keisuke Watanabe (Tokai University, Japan)
Keita Honda (Gifu University, Japan)
Ke Zhang (Chongqing University, P. R. China)
Kenichi Tanaka (Nagasaki Institute of Applied Science, Japan)
Kensuke Ogata (Osaka City University Japan)
Khairul Salleh bin Mohamed Sahari (Universiti Tenaga Nasional, Malaysia)
Khoyratee Farad (The University of Tokyo, Japan)
Kouichi Takeuchi (Okayama University, Japan)
Konstantin Yakovlev (Russian Academy of Sciences, Higher School of Economics, Russia)
Kui Xiang (Wuhan University of Technology, P. R. China)
Leonid Freidovich (Umea University, Sweden)
Levi Timothée (The University of Tokyo, Japan)
Liming Chen (Beihang University, P. R. China)
Lin Li (University of Shanghai for Science and Technology, P. R. China)
Lin Zhao (Qingdao University, P. R. China)
Mamoru Yoshimura (The University of Miyazaki, Japan)
Manabu Yamada (Nagoya Institute of Technology, Japan)
Masahide Ito (Aichi Prefectural University, Japan)
Masahiro Ushio (Kyushu Institute of Technology, Japan)
Masahiro Yokomichi (The University of Miyazaki, Japan)
Masamichi Hori (The University of Miyazaki, Japan)
Masanori Takahashi (Tokai University, Japan)
Masayoshi Kano (Chukyou University, Japan)
Masayoshi Tabuse (Kyoto Prefectural University, Japan)
Masaomi Hatakeyama (university of Zurich, Switzerland)
Max Talanov (Kazan Federal University, Russia)
Meng Duan (Beihang University, P. R. China)
Moeko Tominaga (Kyushu Institute of Technology, Japan)
Mohammad Al-Shabi (University of Sharjah, United Arab Emirates)
Mohammad Biglarbegian (University of Guelph, Canada)
Mou Chen (Nanjing University of Aeronautics and Astronautics, P. R. China)
Nan Xiao (Northeastern University at Qinhuangdao, P. R. China)

Nicolas Monmarché (university of Tours, France)
Noritaka Sato (Nagoya Institute of Technology, Japan)
Norrima Mokhtar (University of Malaya, Malaysia)
Palakorn Tantrakool (King Mongkut's Institute of Technology, North Bankok, Thailand)
Ping Wang (North China Electric Power University, P. R. China)
Pierre David (University of Strasbourg, France)
Pierre Willaume (University of Strasbourg, France)
Rizauddin bin Ramli (Universiti Kebangsaan Malaysia, Malaysia)
Roman Lavrenov (Kazan Federal University, Russia)
Ruztamreen bin Jenal (Universiti Teknikal Malaysia Melaka, Malaysia)
Ryohei Anshi (Kyushu Institute of Technology, Japan)
R.K.P.S. Ranaweera (University of Moratuwa, Sri Lanka)
Satoshi Ikeda (The University of Miyazaki, Japan)
Sanjay S. Joshi (University of California, USA)
Seong-Ik Han (Pusan National University, Korea)
Shahriman Abu Bakar (Universiti Malaysia Perlis, Malaysia)
Shanbi Wei (Chongqing University, P. R. China)
Shihao Sun (Beihang University, P. R. China)
Shin-ichi Asakawa (Tokyo Woman's Christian University, Japan)
Shin Wakitani (Hiroshima University)
Shumin Fei (Southeast University, P. R. China)
Shyi-Ming Chen (National Taichung University of Education, Taiwan)
Stephen Wilkerson (York College of Pennsylvania, USA)
Takashi Kuremoto (Nippon Institute of Technology, Japan)
Takayoshi Yamada (Gifu University, Japan)
Takuya Fuginaga (Fukuoka University, Japan)
Takuya Nanami (The University of Tokyo. Japan)
Taishiro Kishimoto (Keio University, Japan)
Taisuke Akimoto (Kyushu Institute of Technology, Japan)
Tarik bin Abd Latef (Universiti Malaya, Malaysia)
Taro Shibanoki (Okayama University, Japan)
Tetsuro Katayama (The University of Miyazaki, Japan)
Thomas Noel (University of Strasbourg, France)
Thunyaseth Sethaput (Thammasat University, Thailand)
Tianping Zhang (Yangzhou University, P. R. China)
Tomohiko Takagi (Kagawa University, Japan)
Toru Hiraoka (University of Nagasaki, Japan)
Toshihiro Inukai (DENSO Wave Incorporated, Japan)
Toshinori Nawata (Kumamoto National College of Technology, Japan)
Tsunehiro Yoshinaga (Tokuyama National College of Technology, Japan)
Ussath Martin (The University of Potsdam, Germany)
Wan Khairunizam Wan Ahmad (Universiti Malaysia Perlis, Malaysia)
Weicun Zhang (University of Science and Technology Beijing, P. R. China)
Wenhao Zheng (Beihang University, P. R. China)
Wenlin Li (Beihang University, P. R. China)
Wisanu Jiviriya (King Mongkut's University of Technology North Bankok, Thailand)
Xiaocan Wang (Technical University Munich, Germany)

Xiaofeng Su (Beijing Institute of Astronautical Systems Engineering, P. R. China)
Xiaoyan Chen (Tianjin University of Science and Technology, P. R. China)
Xiaoyan Fu (Capital Normal University, P. R. China)
Xuemei Ren (Beijing Institute of Technology, P. R. China)
Xuhui Lu (Beihang University, P. R. China)
Yan Cui (Shanxi Normal University, P. R. China)
Yasunori Takemura (Nishinippon Institute of Technology, Japan)
Yo Horikawa (Kagawa University, Japan)
Yongqiang Qi (China University of Mining and Technology, P. R. China)
Yoshihiro Kita (University of Nagasaki, Japan)
Youji Kawamura (Kindai University, Japan)
Yu-an Zhang (Qinghai University, P. R. China)
Yue Lin (Beijing Institute of Control Engineering, P.R. China)
Yueqian Liang (China Academy of Elecronics and Information Technology, P. R. China)
Yuji Minami (National Institute of Technology, Ube College, Japan)
Yunju Chen ((Shiga University, Japan)
Yunzhong Song (Henan Polytechnic University, P. R. China)
Zacharie Mbaitiga National Institute of Technology, Okinawa College, Japan)
Zakri bin Ghazali (Universiti Malaysia Pahang, Malaysia)
Zengqiang Chen (Nankai University, P. R. China)
Zhao Peng (Huazhong University of Science and Technology, P. R. China)
Zhengquan Yang (Civil Aviation University of China, P. R. China)
Zhongxin Liu (Nankai University, P. R. China)
Zuradzman bin Mohamad Razlan (Universiti Malaysia Perlis, Malaysia)

LOCAL ARRANGEMENT COMMITTEE

Makoto Sakamoto (University of Miyazaki, Japan)
Masanori Sugisaka (ALife Robotics Corporation Ltd., Japan)
Takao Ito (Hiroshima University, Japan)

HISTORY

The International Conference on Artificial Life and Robotics (ICAROB) resulted from the AROB-symposium (International Symposium on Artificial Life and Robotics) whose first edition was held in 1996 and the eighteenth and last edition in 2013. The AROB symposium was annually organized by Oita University and ALife Robotics Corporation Ltd., under the sponsorship of the Science and Technology Policy Bureau, the Ministry of Education, Science, Sports, and Culture (Monbusho), presently, the Ministry of Education, Culture, Sports, Science, and Technology (Monkasho), Japanese Government, Japan Society for the Promotion of Science (JSPS), the Commemorative Organization for the Japan World Exposition ('70), Air Force Office of Scientific Research, Asian Office of Aerospace Research and Development (AFOSR/AOARD), USA. I would like to express my sincere thanks to not only Monkasho (annually fund support from 1996 to 2013) but also JSPS, the Commemorative Organization for the Japan World Exposition ('70), and various other Japanese companies for their repeated support. The old symposium (this symposium has been held every year at B-Con Plaza, Beppu, Oita, Japan except in Oita, Japan (AROB 5th ’00) and in Tokyo, Japan (AROB 6th ’01).) was organized by the International Organizing Committee of AROB and was co-operated by the Santa Fe

Institute (USA), RSJ, IEEJ, ICASE (Now ICROS) (Korea), CAAI (P. R. China), ISCIE, IEICE, IEEE (Japan Council), JARA, and SICE. The old AROB-symposium expanded much by absorbing much new knowledge and technologies into it. This history and character of the former AROB symposiums are passed on the current ICAROB conference and to these journals, Journal of Robotics, Networking and Artificial Life (JRNAL)(vol1-8) & Journal of Robotics, Networking and Artificial Life (JRNAL)(vol9-) & Journal of Advances in Artificial Life Robotics (JAALR). From now on, ALife Robotics Corporation Ltd. is in charge of management of both the conference and the journals. The future of the ICAROB is brilliant from a point of view of yielding new technologies to human society in the 21st century. We also expect to establish an international research institute on Artificial Life and Robotics in the future with the help of Japanese Government and ICAROB. This conference invites you all.

AIMS AND SCOPE

The objective of this conference is the development of new technologies for artificial life and robotics which have been recently born in Japan and are expected to be applied in various fields. This conference presents original technical papers and authoritative state-of-the-art reviews on the development of new technologies concerning robotics, networking and artificial life and, especially computer-based simulation and hardware for the twenty-first century. This conference covers a broad multidisciplinary field, including areas such as:
Artificial intelligence & complexity
Artificial living
Artificial mind research
Artificial nervous systems for robots
Artificial sciences
Bipedal robot
Brain science and computing
Chaos
Cognitive science
Computational Molecular biology
Computer graphics
Data mining
Disasters robotics
DNA computing
Empirical research on network and MOT
Environment navigation and localization
Evolutionary computations
Facial expression analysis, music recommendation and augmented reality
Foundation of computation and its application
Fuzzy control
Genetic algorithms
Human-welfare robotics
Image processing
Insect-like aero vehicles
Intelligence in biological systems
Intelligent control
Management of technology
Medical surgical robot
Micro-machines
Multi-agent systems
Nano-biology
Nano-robotics
Networking
Neural circuits
Neuro-computer
Neuromorphic Systems
Neuroscience
Pattern recognition
Quantum computing

Reinforcement learning system & genetic programing
Robotics
Software development support method
System cybernetics
Unmanned underwater vehicles
Unmanned Aerial Systems Technologies
Unmanned Aerial Systems designing, controls and navigation
Unmanned Aero vehicles
Virtual reality
Visualization
Hardware-oriented submissions are particularly welcome. This conference will discuss new results in the field of artificial life and robotics

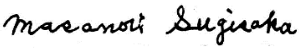
COPYRIGHTS

Accepted papers will be published in the proceeding of The 2025 International Conference on Artificial Life and Robotics (ICAROB2025) by ALife Robotics Corp. Ltd. Copyright belongs to ALife Robotics Corp. Ltd. Some of high-quality papers in the proceeding will be requested to re-submit their papers for the consideration of publication in Journal of Robotics, Networking and Artificial Life (JRNAL)(vol1-8) & Journal of Robotics, Networking and Artificial Life (JRNAL)(vol9-) & Journal of Advances in Artificial Life Robotics(JAALR) under agreement of both Editor-in- Chief Dr. Masanori Sugisaka and 3 reviewers. All correspondence related to the conference should be addressed to ICAROB Office.

ICAROB Office
ALife Robotics Corporation Ltd.
4-Go, 8-Ban, Higshi 2Cyome, Handadai, Oita 870-1108, JAPAN
TEL/FAX：+81-97-597-7760
E-MAIL：icarob@alife-robotics.co.jp
Home Page：https://alife-robotics.co.jp/

MESSAGES

Masanori Sugisaka
General Chair
(President, ALife Robotics Corp., Ltd, Japan)



Masanori Sugisaka

General Chair of ICAROB

It is my great honor to invite you all to The 2025 International Conference on Artificial Life and Robotics (ICAROB 2025) to be held at J:COM HorutoHall, Oita, Japan, 2025. This Conference is changed as the old symposium from the first (1996) to the Eighteenth (2013) annually which were organized by Oita University and ALife Robotics Corporation Ltd. under the sponsorship of the Science and Technology Policy Bureau, the Ministry of Education, Science, Sports, and Culture (Monbusho), presently, the Ministry of Education, Culture, Sports, Science, and Technology (Monkasho), Japanese Government, Japan Society for the Promotion of Science (JSPS), The Commemorative Organization for the Japan World Exposition ('70), Air Force Office of Scientific Research, Asian Office of Aerospace Research and Development (AFOSR/AOARD), USA. I would like to express my sincere thanks to not only Monkasho (annually fund support from 1996 to 2013) but also JSPS, the Commemorative Organization for the Japan World Exposition ('70), Japanese companies for their repeated support. The old symposium was organized by International Organizing Committee of AROB and was co-operated by the Santa Fe Institute (USA), RSJ, IEEJ, ICASE (Now ICROS) (Korea), CAAI (P. R. China), ISCIE, IEICE, IEEE (Japan Council), JARA, and SICE. The old AROB symposium was growing up by absorbing many new knowledge and technologies into it. This history and character was inherited also from ICAROB2014(The 2014 International Conference on Artificial Life and Robotics, included a series of ICAROB proceedings indexed by SCOPUS and CPCI-Web of Science now. From now on, ALife Robotics Corporation Ltd. is in charge of management. This year we have The 2025 International Conference on Artificial Life and Robotics (ICAROB2025) (30th AROB Anniversary). The future of The ICAROB is brilliant from a point of view of yielding new technologies to human society in 21st century. I have founded Robot Artificial Life Society in 2017/12/07 together with Professor at Hiroshima University Takao Ito and Professor at University of Miyazaki Makoto Sakamoto. I hope that fruitful discussions and exchange of ideas between researchers during Conference (ICAROB2025) will yield new merged technologies for happiness of human beings and, hence, will facilitate the establishment of an international joint research institute on Artificial Life and Robotics in future.

Yingmin Jia
Co-General Chair
(Professor, Beihang University, P.R. China)

Yingmin Jia

Co-General Chair of ICAROB

It is my great pleasure to invite you to The 2025 International Conference on Artificial Life and Robotics (ICAROB 2025), will be held at J:COM Horuto Hall, Oita, JAPAN, from February 13 to 16, your understanding and support will be the strongest driving force for us to organize the meeting well.

ICAROB develops from the AROB that was created in 1996 by Prof. Masanori Sugisaka and will celebrate her 30th Anniversary in 2025. So far many important results have been presented at the past meetings and have a profound impact on artificial life and robotics. Doubtless, it is really one of the most famous international conferences in the field of artificial intelligence and attract wide interests among scientist, researchers, and engineers around the world, and effectively promotes the unprecedented popularity of artificial intelligence. Especially, we feel very happy for Prof. Sugisaka's good health and would like to express our sincerest congratulations to him.

For a successful meeting, many people have contributed their great efforts to the ICAROB. Here, I would like to express my special thanks to all authors and speakers, and the meeting organizing team for their excellent works. Looking forward to seeing you at the ICAROB2025.

Takao Ito
Co-General Chair
(Professor Hiroshima University, Japan)

Takao Ito

Co-General Chair of ICAROB

It is my great honor and pleasure to invite you all to the 2025 International Conference on Artificial Life and Robotics (ICAROB 2025).

The ICAROB has its long history. First launched in 1996 as ISAROB, this former organization of ICAROB, was developed under the strong leadership and yeoman efforts of the President—the internationally famous Professor Masanori Sugisaka, who is widely acknowledged as the father of our AROB conference. Our conference has brought together many research scholars, faculty members, and graduate students from all over the world, and published numerous manuscripts in high-quality proceedings as well as highly reputed journals every year.

Over the years, dramatic improvements have been made in the field of artificial life and its applications. The ICAROB has provided a foundation for unifying the exchange of scientific information on the studies of man-made systems that exhibit the behavioral characteristics of natural living systems, including software, hardware, and wetware. Our conference shapes the development of artificial life, extending our empirical research beyond the territory circumscribed by life-as-we-know-it and into the domain of life-as-it-could–be. It will provide us a good place to present our new research results, innovative ideas, and valuable information about artificial intelligence, complex systems theories, robotics, and management of technology.

The conference site is Horuto Hall, one of the most famous international convention centers in Oita City, Japan. You can find many fantastic scenic spots and splendid historical places in Oita City. Please enjoy your stay!

I eagerly look forward to personally meeting you during the ICAROB 2025 and to sharing a most pleasant, interesting, and fruitful conference with you. Do come and make this conference a fruitful, productive as well as enjoyable event!

Ju-Jang Lee
Co-General Chair
(Honorary professor, KAIST)

Ju-Jang Lee

Co-General Chair of ICAROB

The First International Conference on Artificial Life and Robotics (ICAROB) was held in Oita City, Oita, Japan from Jan. 11th to 13th, 2014. This year's Conference will be held amidst the high expectation of the increasingly important role of the new interdisciplinary paradigm of science and engineering represented by the field of artificial life and robotics that continuously attracts wide interests among scientist, researchers, and engineers around the globe.
Distinguished researchers and technologists from around the world are looking forward to attending and meeting at ICAROB. ICAROB is becoming the annual excellent forum that represents a unique opportunity for the academic and industrial communities to meet and assess the latest developments in this fast-growing artificial life and robotics field. ICAROB enables them to address new challenges, share solutions, discuss research directions for the future, exchange views and ideas, view the results of applied research, present and discuss the latest development of new technologies and relevant applications.
In addition, ICAROB offers the opportunity of hearing the opinions of well-known leading experts in the field through the keynote sessions, provides the bases for regional and international collaborative research, and enables to foresee the future evolution of new scientific paradigms and theories contributed by the field of artificial life and robotics and associated research area. The twenty-first century will become the century of artificial life and intelligent machines in support of humankind and ICAROB is contributing through wide technical topics of interest that support this direction.
It is a great honor for me as a Co-General Chair of the 12th ICAROB 2025 to welcome everyone to this important event. Also, I would like to extend my special thanks to all authors and speakers for contributing their research works, the participants, and the organizing team of the 12th ICAROB.
I'm looking forward to meeting you at the 12th ICAROB at J:COM HorutoHall in Oita city and wishing you all the best.

GENERAL SESSION TOPICS

GS1 Machine Learning & Autonomous Driving (5)	GS2 Image Processing I (4)
GS3 Image Processing II (4)	GS4 Robotics (5)
GS5 Applications I (6)	GS6 Applications II (5)

ORGANIZED SESSION TOPICS

OS1 Human-Machine Interface (7)	OS2 Computer and Information Engineering (3)
OS3 Robot Path Planning (6)	OS4 Robot Images and Motion (3)
OS5 Intelligent Life and Cybersecurity (7)	OS6 Intelligent Algorithm Application (7)
OS7 Intelligent Robotics and Systems (6)	OS8 Intelligent Systems and Applications (8)
OS9 Pattern Recognition and Control I (7)	OS10 Pattern Recognition and Control II (7)
OS11 Industrial Artificial Intelligence Robotics (7)	OS12 Advances in Field Robotics and Their Applications (9)
OS13 Research Towards the Renewable Energy and the Sustainable Development Goals (SDG's) / Part A (5)	OS14 Research Towards Renewable Energy and the Sustainable Development Goals (SDG's)/ Part B (4)
OS15 Robotic Manipulation (5)	OS16 Natural Computing (3)
OS17 Artificial Intelligence for Embedded Systems and Robotics (10)	OS18 Robotics in Biophilic-Designed Space Toward Co-existence of Humans, Robots, and Plants (4)
OS19 Intelligent Control (4)	OS20 Applications (6)
OS21 Applications of Information Technology for Better Living (5)	OS22 Navigation and Tracking (3)
OS23 Mathematical Informatics (10)	OS24 New Media Interactions (3)
OS25 Robotic and Communications (6)	OS26 Navigating the Digital Frontier: Innovations in the Age of Industry Revolution 4.0 (11)
OS27 Industrial Revolution (4)	

TIME TABLE (2/13)

Local time in Japan

2/13(Thu.) 17:30-19:30	Welcome Party (Oita Century Hotel)
2/16(Sun) 15:00-16:00	Farewell Party

TIME TABLE (2/14)

2/14(Fri.)	**Room 403**	**Room 404**	**Room 405**	**Room 406** **Remote Session** (ZOOM ID: 835 1207 4893)	**Room 407** **Remote Session** (ZOOM ID: 810 2039 1137)
9:40-	Registration (407)				
10:00-11:15	OS7 Intelligent Robotics and Systems (6) Chair: Kuo-Hsien Hsia will be ended before 11:30	OS19 Intelligent Control (4) Chair: Yingmin Jia	GS6 Applications II (5) Chair: Marion Oswald	OS14 Research Towards Renewable Energy and the Sustainable Development Goals (SDG's)/ Part B (2) Chair: Firas Basim	
11:15-11:30	Coffee Break				
11:30-12:00	Chair: Marion Oswald (Room 302. 303) Opening Ceremony				OS25-1 Robotic and Communications (2) Chair: Mastaneh Moayef
12:00-13:00	Lunch				
13:00-14:00	Chair: Takao Ito (Room 302, 303) Plenary Speech PS2 Firas Basim Ismail (Universiti Tenaga Nasional (UNITEN), Malaysia)				OS25-2 Robotic and Communications (4) Chair: Mastaneh Mokayef
14:00-14:30	Coffee break				
14:30-15:45	OS5-1 Intelligent Life and Cybersecurity (5) Chair: I-Hsien Liu	OS8-1 Intelligent Systems and Applications (5) Chair: Chun-Liang Liu	OS17-1 Artificial Intelligence for Embedded Systems and Robotics (5) Chair: Hakaru Tamukoh	OS6-1 Intelligent Algorithm Application (5) Chair: Peng Wang	OS9-1 Pattern Recognition and Control I (5) Chair: Sun Haozhe
15:45-16:00	Coffee break				
16:00-17:15	OS5-2 Intelligent Life and Cybersecurity (2) Chair: I-Hsien Liu will be ended before 16:30	OS8-2 Intelligent Systems and Applications (3) will be ended before 16:45	OS17-2 Artificial Intelligence for Embedded Systems and Robotics (5) Chair: Hakaru Tamukoh	OS6-2 Intelligent Algorithm Application (2) Chair: Peng Wang will be ended before 16:30	OS9-2 Pattern Recognition and Control I (2) Chair: Sun Haozhe will be ended before 16:30

TIME TABLE (2/15)

2/15(Sat.)	Room 403	Room 404	Room 405	Room 406 Remote Session (ZOOM ID: 835 1207 4893)
9:40-	Registration (407)			
10:00-11:15	GS4 Robotics (4) Chair: will be ended before 11:00	OS18 Robotics in Biophilic-Designed Space Toward Co-existence of Humans, Robots, and Plants (4) Chair: Yuichiro Tanaka	GS5 Applications I (5) Chair: Tsutomu Ito	OS1-1 Human-Machine Interface (5) Chair: Norrima Mokhtar
11:15-11:30	Coffee break			
11:30-12:15	GS3 Image Processing II (3) Chair: Kuo-Hsien Hsia	OS16 Natural Computing (3) Chair: Marion Oswald	OS22 Navigation and Tracking (3) Chair: Chan Gook Park	OS1-2 Human-Machine Interface (3) Chair: Norrima Mokhtar
12:15-13:00	Lunch			
13:00-14:00	Chair: Eiji Hayshi (Room 302, 303) Plenary Speech PS3 Norrima Mokhtar (Universiti Malaya, Malysia)			OS26-1 Navigating the Digital Frontier: Innovations in the Age of Industry Revolution 4.0 (5) Chair: Wei Hong Lim will be ended before 14:15
14:00-14:20	Coffee break			
14:20-15:20	Chair: Amane Takei (Room 302, 303) Plenary Speech PS1 Shinobu Yoshimura (The University of Tokyo, Japan)			OS26-2 Navigating the Digital Frontier: Innovations in the Age of Industry Revolution 4.0 (4) Chair: Wei Hong Lim
15:20-15:40	Coffee break			
15:40-16:55	GS1 Machine Learning & Autonomous Driving (4) Chair: Obada Al aama will be ended before 16:40	OS11-1 Industrial Artificial Intelligence Robotics (5) Chair: Eiji Hayashi	OS15 Robotic Manipulation (5) Chair: Kensuke Harada	OS26-3 Navigating the Digital Frontier: Innovations in the Age of Industry Revolution 4.0 (3) Chair: Wei Hong Lim
16:55-17:05				
17:05-18:05	GS2 Image Processing I (4) Chair: Amane Takei	OS11-2 Industrial Artificial Intelligence Robotics (2) Chair: Eiji Hayashi will be ended before 17:35	OS24 New Media Interactions (3) Chair: R.P.C. Janaka Rajapakse Remote Session (ZOOM ID: 841 9902 1154)	OS2 Computer and Information Engineering (3) Chair: Norrima Mokhtar will be ended before 17:50
18:30-20:30	Banquet (Tokiwa Kaikan)			

TIME TABLE (2/16)

2/16(Sun.)	Room 403	Room 404	Room 405 Remote Session (ZOOM ID: 841 9902 1154)	Room 406 Remote Session (ZOOM ID: 835 1207 4893)	Room 407 Remote Session (ZOOM ID: 810 2039 1137)
9:40-	Registration (407)				
10:00-11:15	OS12-1 Advances in Field Robotics and Their Applications (5) Chair: Shinsuke Yasukawa	OS21 Applications of Information Technology for Better Living (5) Chair: Tetsuro Katayama	OS10-1 Pattern Recognition and Control II (5) Chair Zhai Hongshuo	OS13 Research Towards the Renewable Energy and the Sustainable Development Goals (SDG's) / Part A (5) Chair Ammar A.M. Al Talib	OS27 Industrial Revolution (4) Chair: Hazry Desa will be ended before 11:00
11:15-11:30	Coffee Break				
11:30-12:00	OS12-2 Advances in Field Robotics and Their Applications (2) Chair: Shinsuke Yasukawa	OS23-1 Mathematical Informatics (2) Chair: Takao Ito	OS10-2 Pattern Recognition and Control II (2) Chair Zhai Hongshuo	OS4 Robot Images and Motion (3) Chair: Miao Zhang Will be ended before 12:15	OS14 Research Towards Renewable Energy and the Sustainable Development Goals (SDG's)/ Part B (2) Chair: Firas Basim
12:00-13:00	Lunch				
13:00-15:00	OS12-3 Advances in Field Robotics and Their Applications (2) Chair: Shinsuke Yasukawa will be ended before 13:30	OS23-2 Mathematical Informatics (8) Chair: Takao Ito	OS20 Applications (8) Chair: Kasthuri Subaramaniam	OS3 Robot Path Planning (6) Chair: Shengwei Liu will be ended before 14:30	
15:00-16:00	Farewell Party				

The 2025 International Conference on ARTIFICIAL LIFE AND ROBOTICS (ICAROB2025)

February 13 (Thursday)

17:30-19:30 Welcome Party (Oita Century Hotel)

February 14 (Friday)

Room 302, 303

11:30-12:00 Opening Ceremony

Chair: Marion Oswald (Vienna University of Technology, Austria)

Welcome Addresses

1. General Chairman of ICAROB	Masanori Sugisaka (ALife Robotics Co., Ltd., Japan)
2. Co-General Chairman of ICAROB	Yingmin Jia (Beihang University, China)
3. Co-General Chairman of ICAROB	TaKao Ito (Hiroshima University, Japan)
4. Co General Chairman of ICAROB	Ju-Jang Lee (Honorary professor, KAIST)
5. Vice General Chair of ICAROB	Norrima Mokhtar (University of Malaya, Malaysia)
6. Vice General Chair of ICAROB	Katia Passerini (Seton Hall University, USA)

February 15 (Saturday)

Banquet: Tokiwa Kaikan

18:30-20:30

Chair: Takao Ito (Hiroshima University, Japan)

Welcome Addresses

Prof. Yingmin Jia (Beihang University, P.R. China)

Dr. Norrima Mokhtar (University of Malaya, Malaysia)

TECHNICAL PAPER INDEX

February 14 (Friday)

9:40-Registration

Room 302, 303
11:30-12:00 Opening Ceremony
Chair: Marion Oswald (Vienna University of Technology, Austria)

13:00-14:00
Plenary Speech PS2
Chair: Takao Ito (Hiroshima University, Japan)

PS2 *Toward 2035: Renewable Energy Innovations Transforming Our Future*
Firas Basim Ismail (Universiti Tenaga Nasional (UNITEN), Malaysia)

Room 403
10:00-11:30 OS7- Intelligent Robotics and Systems (6)
Chair: Kuo-Hsien Hsia (National Yunlin University of Science and Technology, Taiwan)
Co-Chair: Jia-Ming Hsiao (National Yunlin University of Science and Technology, Taiwan)

OS7-1 *Fuzzy-Controlled Multi-Valve Pneumatic Soprano Recorder Auto-Playing and Score Recognition System*
Chun-Chieh Wang*, Chung-Wen Hung, Kuo-Hsien Hsia, Chian C. Ho, Ying-Yuan Yao
(National Yunlin University of Science and Technology, Taiwan)

OS7-2 *Development of agricultural robots based on ROS*
Jr-Hung Guo*, Kuo-Hsien Hsia (National Yunlin University of Science and Technology, Taiwan)

OS7-3 *The ROS-based web information center of small manipulator*
Yue-Jie Wang, Jia-Ming Hsiao*, Shao-Yi Hsiao
(National Yunlin University of Science & Technology, Taiwan)

OS7-4 *Image-assisted Assembly and Disassembly Process Using TM Six-Axis Collaborative Robotic Arm*
Kuo-Hsien Hsia, Yi- Yan Liao, Ching-Yuan Pan
(National Yunlin University of Science & Technology, Taiwan)

OS7-5 *Quality Inspection of PVC Shoe Chopsticks: A Research Study*
Shu-Li Pai, Kuo-Hsien Hsia, Chian-Cheng Ho
(National Yunlin University of Science and Technology, Taiwan)

OS7-6 *Obstructing PLC Operations through Modbus Command Manipulation*
Nai-Yu Chen, Cheng-Ying He, Jung-Shain Li, Chu-Sing Yang, I-Hsien Liu
(National Cheng Kung University, Taiwan)

14:30-15:45 OS5-1 Intelligent Life and Cybersecurity (5)
Chair: I-Hsien Liu (National Cheng Kung University, Taiwan)
Co-Chair: Chu-Fen Li (National Formosa University, Taiwan)
Co-Chair: Cheng-Han Lin (Fooyin University, Taiwan)

OS5-1 *A Diamond Model Approach to Analyzing GhostSec's Intrusion Paths*
Cheng-Ying He, Nai-Yu Chen, Jung-Shain Li, I-Hsien Liu
(National Cheng Kung University, Taiwan)

OS5-2 *A Bandwidth-Aware Routing Mechanism to Control Hadoop Shuffle Traffic over Software-Defined Networking*
Ming-Syuan Wu (National Kaohsiung University of Science and Technology, Taiwan)
Cheng-Han Lin (Fooyin University, Taiwan)
Wen-Shyang Hwang (National Kaohsiung University of Science and Technology, Taiwan)
Ce-Kuen Shieh, Mao-Syun Lin (National Cheng Kung University, Taiwan)

OS5-3 *AI Sentencing System: Homicide Case Study in Taiwan*
Shih-Chin Lin, Cheng-Tsung Yeh, Chen-Yu Lai, Chih-Yun Chang, Chi-Ju Fu
(Ming Chuan University, Taiwan)

OS5-4 *Inferring ICS Topology and Behavior through Network Traffic Analysis*
Chien-Wen Tseng, Jung-Shain Li, I-Hsien Liu
(National Cheng Kung University, Taiwan)
Chu-Fen Li (National Formosa University, Taiwan)

OS5-5 *Using fuzzy control routing for dynamic load balancing over Software-Defined Networks*
Wen-Shyang Hwang, Ming-Syuan Wu, Sian-Fong Huang
(National Kaohsiung University of Science and Technology, Taiwan)
Cheng-Han Lin (Fooyin University, Taiwan),
Yan-Jing Wu (Shih Chien University, Taiwan)
Ming-Hua Cheng (Tzu-Hui Institute of Technology, Taiwan)

16:00-16:30 OS5-2 Intelligent Life and Cybersecurity (2)
Chair: I-Hsien Liu (National Cheng Kung University, Taiwan)
Co-Chair: Chu-Fen Li (National Formosa University, Taiwan)
Co-Chair: Cheng-Han Lin (Fooyin University, Taiwan)

OS5-6 *The Application of AI in the Real Estate Industry: Business Model Innovation Perspective*
Li-Min Chuang, Chih-Hung Chen (Chang Jung Christian University, Taiwan)

OS5-7 *The Impact of AI-Powered Health Monitoring on the Quality of Life and Social Participation of the Elderly: Technology Acceptance Model Perspective*
Li-Min Chuang, Zong-Sheng Li (Chang Jung Christian University, Taiwan)

Room 404
10:00-11:00 OS19 Intelligent Control (4)
Chair: Yingmin Jia (Beihang University, P.R.China)
Co-Chair: Weicun Zhang (University of Science and Technology Beijing, P.R.China)

OS19-1 *Practical Linearization Control of Nonholonomic Unicycles*
Lixia Yan, Yingmin Jia (Beihang University (BUAA), P.R.China)

OS19-2 *Task-Space Tracking Control for Dual-arm Free-floating Space Manipulators with Disturbances and Uncertainties*
Qian Sun, Yingmin Jia (Beihang University (BUAA), P.R.China)

OS19-3 *Manipulability Optimization for Redundant Dual-Arm Robots at the Acceleration Level*
Yang Zhang, Yingmin Jia (Beihang University (BUAA), P.R.China

OS19-4 *Flocking Control for Multiple Convex Polygonal Agents with Obstacle Avoidance*
Yaxin Li, Yingmin Jia (Beihang University (BUAA), Beijing P.R.China)

14:30-15:45 OS8-1 Intelligent Systems and Applications (5)
Chair: Chun-Liang Liu (National Yunlin University of Science and Technology, Taiwan)
Co-Chair: Yuting Hsiao (National Yunlin University of Science and Technology, Taiwan)

OS8-1 *MCU Ultrasonic oscillator driver with digital frequency sweep function*
Chung-Wen Hung, Chun-Chieh Wang, Yu-Hsing Su
(National Yunlin University of Science and Technology, Taiwan)

OS8-2 *Design and development of foot pressure sensing massage stick*
Yuting Hsiao, Dengchuan Cai, Chung-Wen Hung
(National Yunlin University of Science and Technology, Taiwan)

OS8-3 *Effects on physiological indicators of foot massage using a pressure sensing massage stick*
Dengchuan Cai, Yuting Hsiao, Chung-Wen Hung
(National Yunlin University of Science and Technology, Taiwan)

OS8-4 *Study on Maximum Power Point Tracking Technology for Solar Power Systems Using Power Variation to Adjust Step Response*
Chun-Liang Liu, Chung-Wen Hung, Yi-Feng Luo, Guan-Jhu Chen, Cheng-Sin Hu
(National Yunlin University of Science and Technology, Taiwan)

OS8-5 *Bearing faulty prediction based on knowledge distillation*
Chun-Liang Liu, Zheng-Jie Liao, Chung-Wen Hung
(National Yunlin University of Science and Technology, Taiwan)

16:00-16:45 OS8-2 Intelligent Systems and Applications (3)
Chair: Chun-Liang Liu (National Yunlin University of Science and Technology, Taiwan)
Co-Chair: Yuting Hsiao (National Yunlin University of Science and Technology, Taiwan)

OS8-6 *Accurate Brain Age Prediction Through Advanced Preprocessing and 3D DenseNet-50 Modeling*
Ting-An Chang, Chiang-Ming Yeh, Chun-Liang Liu
(National Yunlin University of Science and Technology, Yunlin, Taiwan)

OS8-7 *Intelligent agricultural landscape identification system*
Ching Ju Chen[1], Yu-Cheng Chen[1], Jing-Yao Lin[1], Rung-Tsung Chen[2], Candera Wijaya[3]
([1]National Yunlin University of Science and Technology, Taiwan) ([2]Taiwan Biodiversity Research Institute, Taiwan) ([3]Agricultural Engineering Research Center, Taiwan)

OS8-8 *Leveraging AIoT Visual Analytics for Optimizing Agricultural Sustainability and Efficiency*
Hsueh-Yen Shih[1], Xi-Wei Lin[2], Zhao-Sheng Chen[2], Ying-Cheng Chen[3], Ching-Ju Chen[2]
([1]Agricultural Engineering Research Center, Taiwan) ([2]National Yunlin University of Science and Technology, Taiwan) ([3]Tainan District Agricultural Research and Extension Station, Taiwan)

Room 405
10:00-11:15 GS6 Applications II (5)
Chair: Marion Oswald (TU Vienna, Austria)

GS6-1 *Signal Decomposition and Noise Reduction in Single-Channel EEG: A Morphological Component Analysis (MCA) Approach*
Kosei Shibata[1], Yide Yang[1], Rena Kato[1], (Kyushu Institute of Technology, Japan),
Hendry Ferreira Chame[2], Laurent Bougrain[2] (Université de Lorraine, CNRS, LORIA, France),
Tomohiro Shibata[1], Hiroaki Wagatsuma[1](Kyushu Institute of Technology, Japan)

GS6-2 *Variable Selection Methods for Multivariate Time Series Data Using Multivariate Granger Causality*
Keita Ohmori[1,2], Toshiki Saitoh[1], Akiko Fujimoto[1], Eiji Miyano[1]
([1]Kyushu Institute of Technology, [2]SUMCO, Japan

GS6-3 *A Support System for a Visually Impaired Person Finding Bus Route Numbers Employing MY VISION*
Daichi Nanaura, Seiji Ishikawa, Yui Tanjo (Kyushu Institute of Technology, Japan)

GS6-4 *Human Pose Estimation from Egocentric Videos*
Shunya Egashira, Yui Tanjo (Kyushu Institute of Technology, Japan)

GS6-5 *Analyzing Eye-Tracking Data to Detect Joint Attention in Hexgame Experiments*
Yide Yang[1], Rena Kato[1], Kosei Shibata[1] (Kyushu Institute of Technology, Japan),
Hendry Ferreira Chame[2], Laurent Bougrain[2] (Université de Lorraine, France),
Hiroaki Wagatsuma[1](Kyushu Institute of Technology, Japan)

14:30-15:45 OS17-1 Artificial Intelligence for Embedded Systems and Robotics (5)
Chair: Hakaru Tamukoh (Kyushu Institute of Technology, Japan)
Co-Chair: Yuma Yoshimoto (National Institute of Technology, Kitakyushu College, Japan)
Co-Chair: Dinda Pramanta (Kyushu Institute of Information Sciences, Japan)

OS17-1 *Optimizing Object Placement for Human Support Robots Using a Two-dimensional Irregular Packing Algorithm for Efficient Tray Storage*
Natee Buttawong, Kosei Isomoto, Kosei Yamao, Ninnart Fuengfusin, Hakaru Tamukoh (Kyushu Institute of Technology, Japan)

OS17-2 *Classification of Human Activity by Event-based Vision Sensors using Echo State Networks*
Rohan Saini, Aryan Rakheja, Ryuta Toyoda, Yuichiro Tanaka, Hakaru Tamukoh
(Kyushu Institute of Technology, Japan)

OS17-3 *Integrating Advanced Speech Recognition and Human Attribute Detection for Enhanced Receptionist Task in RoboCup@Home*
Koshun Arimura, Yuga Yano, Takuya Kawabata, Hakaru Tamukoh
(Kyushu Institute of Technology, Japan)

OS17-4 *Classification of Human Activity by Spiking Neural Networks using Event-based Vision Sensors*
Aryan Rakheja, Rohan Saini, Ryuta Toyoda, Yuichiro Tanaka, Hakaru Tamukoh
(Kyushu Institute of Technology, Japan)

OS17-5 *Robotic Grasping of Common Objects: Focusing on Edge Detection for Improved Handling*
Tomoya Shiba, Hakaru Tamukoh (Kyushu Institute of Technology, Japan)

16:00-17:15 OS17-2 Artificial Intelligence for Embedded Systems and Robotics (5)
Chair: Hakaru Tamukoh (Kyushu Institute of Technology, Japan)
Co-Chair: Yuma Yoshimoto (National Institute of Technology, Kitakyushu College, Japan),
Co-Chair: Dinda Pramanta (Kyushu Institute of Information Sciences, Japan)

OS17-6 *Proposal of a Grasp Verification Method Utilizing Background Subtraction and Depth Information*
Ryo Terashima, Yuga Yano, Koshun Arimura, Hakaru Tamukoh
(Kyushu Institute of Technology, Japan)

OS17-7 *Grasp Point Estimation Using Object Recognition Models with Simulator-Generated Datasets Including Pose Information*
Ryoga Maruno, Tomoya Shiba, Naoki Yamaguchi, Hakaru Tamukoh
(Kyushu Institute of Technology, Japan)

OS17-8 *A feasibility study of generative AI applications using EV-3 Robots at the Kyushu Institute of Information Sciences*
Dinda Pramanta[1], Hakaru Tamukoh[2]
([1]Kyushu Institute of Information Sciences, [2]Kyushu Institute of Technology, Japan)

OS17-9 *Development of a Collaborative System Between A Drone and A Home Service Robot for Enhanced Operational Efficiency*
Haruki Miura, Rion Yohu, Yuma Yoshimoto
(National Institute of Technology, Kitakyushu College, Japan)

OS17-10 *Efficient Object Detection with Color-Based Point Prompts for Densely Packed Scenarios in WRS FCSC 2024*
Naoki Yamaguchi[1], Tomoya Shiba[1], Hakaru Tamukoh[1]
([1]Kyushu Institute of Technology, Japan)

Room 406

10:00-11:00 OS14 Research Towards Renewable Energy and the Sustainable Development Goals (SDG's)/ Part B (4)

Chair: Firas Basim Ismail (University Tenaga National (UNITEN), Malaysia)

Co-Chair: Takao Ito (Hiroshima University, Japan)

OS14-1 *Empowering Decentralized Microgrids with A Blockchain-Based Peer-To-Peer Energy Trading Platform*
Firas Basim Ismail[1], Chetenraj Singh[1], Ammar A. Al-Talib[2], Nizar F.O. Al-Muhsen[1]
([1] UNITEN, Malaysia), ([2] UCSI University, Malaysia)

OS14-2 *Performance of Kenaf Fibre Reinforced Epoxy Biocomposite for High Voltage Insulator Applications*
Kang Rui Tan[1], Cik Suhana Bt. Hasan[1], Nor Fazilah Abdullah[1], Farah Adilah Jamaludin[1], Meng Choung Chiong[1], Eryana Hussin[1] ([1]UCSI University, Malaysia)

OS14-3 *Detection of Bullet Holes for Target Board in Malaysia Military (ATM) Shooting Exam Application*
Jilian.H.Wai Yin, Idayu M. Tahir, Ammar A.M. Al Talib, Osama Mohamed Magzoub
(UCSI University, Malaysia)

OS14-4 *Mobile App Development for Monitoring Goat Activities*
Samy M. Elmasri, Idayu M. Tahir, Ammar A.M. Al Talib
(UCSI University, Malaysia)

14:30-15:45 OS6-1 Intelligent Algorithm Application (5)

Chair: Peng Wang (Tianjin University of Science and Technology, China)

Co-Chair: Miao Zhang (Tianjin University of Science and Technology, China)

OS6-1 *Reliability Analysis and Optimization of Distribution Network with Distributed Generation*
Peng Wang, Mengyuan Hu, (Tianjin University of Science and Technology, China)

OS6-2 *Research on Improved PPLCNet Classification Network Based on CBAM Attention Model*
Peng Wang, Shengfeng Wang, Qikun Wang, Yuting Zhou
(Tianjin University of Science and Technology, China)

OS6-3 *Optimizing Microgrid Power Dispatch with Integrated Ground Source Heat Pumps Using Cellular Automata*
Peng Wang, Shunqi Yang (Tianjin University of Science and Technology, China)

OS6-4 *Indoor Personnel Thermal Comfort Monitoring System Based on Mobile Robots*
Peng Wang, Zihang Zhou (Tianjin University of Science and Technology, China)

OS6-5 *Market Trading Strategy of Integrated Energy Park from the Perspective of Non-cooperative Game*
Peng Wang, Siyi Wang, Liangyu Wang, Chengkai Miao
(Tianjin University of Science and Technology, China)

16:00-16:30 OS6-2 Intelligent Algorithm Application (2)
Chair: Peng Wang (Tianjin University of Science and Technology, China)
Co-Chair: Miao Zhang (Tianjin University of Science and Technology, China)

OS6-6 *Research on the Sensitivity of Thermal Comfort Using Sensitivity Algorithms Based on Variance and Stochastic Expansion*
Peng Wang, Yuting Zhou, Liangyu Wang, Chengkai Miao, Qikun Wang

OS6-7 *Deep Learning Based Infant and Child Monitoring System*
Peng Wang, Jiale Jia (Tianjin University of Science and Technology, China)

Room 407
11:30-12:00 OS25-1 Robotic and Communications (2)
Chair: Mastaneh Mokayef (UCSI University, Malaysia)
Co-Chair: Takao Ito (Hiroshima University, Japan)

OS25-1 *Exploring Techniques To Mitigate Interference In Drone Communication Systems*
Ahmed Alsaeed Rashad[1], Mastaneh Mokayef[1]*, M.K.A Ahamed Khan[1], MHD Amen Summakieh[1], Kim Soon Chong[1], Abdul Qayyum[2], Moona Mazher[3], Sanjoy Kumar Debnath[4], Chin Hong Wong[5, 6], Chua Huang Shen[7]
([1]UCSI University, Malaysia), ([2] Imperial College London, UK),
([3]University College London, UK), ([4]Chitkara University Institute of Engineering and Technology, India), ([5] Fuzhou University, China)
([6] Maynooth University, Ireland), ([7] UOW University Malaysia)

OS25-2 *An Automated Tracking System for Locating Impact Points on a Table Tennis Surface Using Ping Pong Balls*
Lee Wai Kit[1], Mastaneh Mokayef[1 *], MHD Amen Summakieh[1], M.K.A Ahamed Khan[1], Miad Mokayef[1], Sew Sun Tiang[1], Wei Hong Lim[1], Abdul Qayyum[2], Moona Mazher[3], Sanjoy Kumar Debnath[4]
([1] UCSI University, Malaysia), ([2] Imperial College London, UK), ([3] University College London, UK), ([4] Chitkara University, India)

13:00-14:00 OS25-2 Robotic and Communications (4)
Chair: Mastaneh Mokayef (UCSI University, Malaysia)
Co-Chair: Takao Ito (Hiroshima University, Japan)

OS25-3 *An Innovative Deep Learning Technique to Identify Potato Illness*
Abdul Majid Soomro[1 *], Muhammad Haseeb Asghar[2], Sanjoy Kumar Debnath[3], Susama Bagchi[3], and Awad Bin Naeem[2], M.K. A. Ahamed Khan[4], Mastaneh Mokayef[4]
([1] National University of Modern Languages, Pakistan), ([2] National College of Business Administration & Economics, Pakistan), ([3] Chitkara University, India), ([4]UCSI University, Malaysia)

OS25-4 *A Wearable Walking Support System Design And Simulation*
Omar Ayaman Yehiya[1], M. K. A. Ahamed Khan[1] *, Mastaneh Mokayef [1], Ridzuan A[1], Abdul Qayyum[2], Moona Mazher [3], Susama Bagchi[4], Sanjoy Kumar Debnath[4]
([1]UCSI University, Malaysia), ([2] Imperial College, London, UK), ([3]University College London, UK), ([4]Chitkara University, India)

OS25-5 *A Floor Tiling Robotic System*
Hue Chau Jieng[1], M. K. A. Ahamed Khan[1] *, Mastaneh Mokayef [1], Ridzuan A[1], Abdul Qayyum[2], Moona Mazher [3]. Susama Bagchi[4], Sanjoy Kumar Debnath[4]
([1]UCSI University, Malaysia), ([2] Imperial College, London, UK), ([3] University College London, UK), ([4]Chitkara University, India)

OS25-6 *Evaluation of Heart Disease Risk Using Deep Learning Technique with Image Enhancement*
Abdul Majid Soomro[1] *, Asad Abbas[2], Susama Bagchi[3], Sanjoy Kumar Debnath[3], Awad Bin Naeem[2], M. K. A. Ahamed Khan[4], Mastaneh Mokayef[4]
([1]National University of Modern Languages, Pakistan), ([2]National College of Business Administration & Economics, Pakistan), ([3]Chitkara University, India), ([4] UCSI University, Malaysia)

14:30-15:45 OS9-1 Pattern Recognition and Control I (5)

Chair: Sun Haozhe (Tianjin University of Science and Technology, China)

Co-Chair: Li Fangyan (Tianjin University of Science and Technology, China)

OS9-1 *A Study on Surface Defect Detection Algorithm of Strip Steel Based on YOLOv8n*
Haozhe Sun[1], Fengzhi Dai[1], Junjin Chen[2]
([1]Tianjin University of Science and Technology, [2]SMC (Beijing) Manufacturing Co., LTD., China)

OS9-2 *Prediction of Winter Wheat Growth Trends Based on NDVI Vegetation Index*
Lu Kang, Jiahao Xie, Chunli Li, Haoran Gong, Fengzhi Dai
(Tianjin University of Science and Technology, China)

OS9-3 *A Study on Artemia Culture System and Its Application*
Wanying Zhang, Yicheng Wu, Ziting Zhang,Yumei Huang
(Tianjin University of Science and Technology, China)

OS9-4 *Machine Vision-Based Chamfer Detection for Metal Parts*
Shangying Han[1], Kaili Guo[1], Yanzi Kong[1], Yanliang Gong[1], Junjin Chen[2], Ce Bian[3], Mengfan Zhang[3]
([1]Tianjin University of Science and Technology, [2]SMC (Beijing) Manufacturing Co., LTD., [3] Tianjin Tianke Intelligent Manufacture Technology Co., LTD., China)

OS9-5 *Deep Guard Dog - AI-Based Night Intrusion Detection Mobile Phone Software*
Keming Chen, Jiaxin Wang (Tianjin University of Science and Technology, China)

16:00-16:30 OS9-2 Pattern Recognition and Control I (2)

Chair: Sun Haozhe (Tianjin University of Science and Technology, China)

Co-Chair: Li Fangyan (Tianjin University of Science and Technology, China)

OS9-6 *Development of an Amphibious Surface Garbage Collection Robot and Its Applications*
Yu Su, Xin Wang, Long Shen, Zhenxing Liu, Xinrui Zhao, Xin Lin, Mengchen Huo, Yawen Qiao, Yan Zhang (Tianjin University of Science and Technology, China)

OS9-7 *Design of an Intelligent Orbital Inspection Robot Based on Machine Vision and Ultrasonic Guided Waves*
Xingwang Feng, Suqing Duan (Tianjin University of Science and Technology, China)

February 15 (Saturday)

9:40-Registration

Room 302, 303
13:00-14:00
Plenary Speech PS3
Chair: Eiji Hayashi (Kyusyu Institute of Technology, Japan)

PS3 *Integration of Human – Device Interface: Transforming the Future of Interaction*
Norrima Mokhtar (Universiti Malaya, Malysia)

14:20-15:20
Plenary Speech PS1
Chair: Amane Takei (University of Miyazaki, Japan)

PS1 *High-fidelity Multi-agent Simulations for Social Systems*
Shinobu Yoshimura (The University of Tokyo)

Room 403
10:00-11:00 GS4 Robotics (4)
Chair:

GS4-1 *Obstacle-Aware Autonomous Flipper Control Method Based on Terrain Geometry*
Kotaro Kanazawa, Noritaka Sato, Yoshifumi Morita
(Nagoya Institute of Technology, Japan)

GS4-2 *An Adaptive Control Method for a Knee-Joint Prosthetic Leg Toward Dynamic Stability and Gait Optimization*
Ge Yiqian[1], Purevdorj Choisuren[1], Shintaro Kasai[1], Hiroaki Wagatsuma[1]
([1]Kyushu Institute of Technology, Japan)

GS4-3 *A Gait Analysis with Multibody Dynamics Toward Energy-Efficient Active Knee Prostheses*
Purevdorj Choisuren[1], Ge Yiqian[1], Shintaro Kasai[1], Batbaatar Dondogjamts[2], Erdenesuren Naranbaatar[2] and Hiroaki Wagatsuma[1]
([1]Kyushu Institute of Technology, Japan; [2]Mongolian University of Science and Technology, Mongolia)

GS4-4 *Suppressing of Multi-Axial Vibration Caused in Carried Objects by Robot Using a Heuristic Algorithm Based on Evaluation of Actual Machine Information*
Yusuke Ueno[1], Hiroki Noguchi[1], Fumitoshi Shimono[1], Hiroshi Tachiya[2]
([1]Komatsu University, Japan) ([2]Kanazawa University, Japan)

11:30-12:15 GS3 Image Processing II (3)
Chair: Kuo-Hsien Hsia (National Yunlin University of Science and Technology, Taiwan)

GS3-1 *Shape-Preserving Embedding Technique for Binary Classification of Video Image of the Solar Surface*
Iori Tamura, Akiko Fujimoto, Soichiro Kondo, Reiri Noguchi
(Kyushu Institute of Technology, Japan)

GS3-2 *Seated Posture Estimation Based on Monocular Camera Images*
Hitoshi Shimomae, Tsubasa Esumi, Noriko Takemura (Kyushu Institute of Technology, Japan)

GS3-3 *Identification of lung nodules based on combining multi-slice CT images and clinical information*
Yuto Nishitaki, Tohru Kamiya (Kyushu Institute of Technology, Japan)
Shoji Kido (Osaka University, Japan)

15:40-16:40 GS1 Machine Learning & Autonomous Driving (4)
Chair: Obada Al aama (Kyushu Institute of Technology, Japan)

GS1-1 *Automatic classification of respiratory sounds by improving the loss function of ResNet*
Ryusei Oshima[1], Tohru Kamiya[1], Shoji Kido[2]
([1]Kyushu Institute of Technology, Japan), ([2]Osaka University, Japan)

GS1-2 *Classification of Heat Transfer Coefficient Using Deep Learning with Information from Boiling Images*
Fuga Mitsuyama, Ren Umeno, Tomohide Yabuki, Tohru Kamiya
(Kyushu Institute of Technology, Japan)

GS1-3 *A Data Format Integration of Open-Street-Map and Lanelet2 Toward the Ontology Framework for Safety Autonomous Driving systems*
Obada Al aama[1], Takahiro Koga[1], Tomoki Taniguchi[1], Davaanyam Jargal[1], Junya Oishi[2], Shigeru Nemoto[2], Wataru Mizushina[2], Kazuki Hirao[2], Hakaru Tamukoh[1], Hiroaki Wagatsuma[1]
([1]Kyushu Institute of Technology, Japan, [2]Aisan Technology Co., Ltd., Japan)

GS1-4 *Developing a Sound-Based Method to Synchronize Multiple Videos Recorded by Multiple Sound Sources*
Davaanyam Jargal, Rena Kato, Tomoki Taniguchi, Kosei Shibata, Takahiro Koga, Obada Al aama, Hakaru Tamukoh and Hiroaki Wagatsuma
(Kyushu Institute of Technology, Japan)

17:05-18:05 GS2 Image Processing I (4)
Chair: Amane Takei (University of Miyazaki, Japan)

GS2-1 *Recognition of Plastic Bottles Region Using Improved DeepLab v3+*
Yusuke Murata, Tohru Kamiya (Kyushu Institute of Technology, Japan)

GS2-2 *Non-Invasive Classification of EGFR Mutation from Thoracic CT Images Using Radiomics Features and LightGBM*
Reo Takahashi[1], Tohru Kamiya[1], Takashi Terasawa[2], Takatoshi Aoki[2]
([1]Kyushu Institute of Technology, Japan)
([2]University of Occupational and Environmental Health, Japan)

GS2-3 *Detection of Lung Nodules from Temporal Subtraction CT Image Using Elastic Net-Based Features Selection*
Natsuho Baba, Tohru Kamiya (Kyushu Institute of Technology, Japan)
Takashi Terasawa, Takatoshi Aoki (University of Occupational Health)

GS2-4 *Detection of Lung Nodules from CT Image Based on Ensemble Learning*
Natsuho Baba, Tohru Kamiya (Kyushu Institute of Technology, Japan)
Takashi Terasawa, Takatoshi Aoki (University of Occupational Health)

Room 404

10:00-11:00 OS18 Robotics in Biophilic-Designed Space Toward Co-existence of Humans, Robots, and Plants (4)

Chair: Yuichiro Tanaka (Kyushu Institute of Technology, Japan)
Co-Chair: Naoto Ishizuka (Kyushu Institute of Technology, Japan)
Co-Chair: Tomomi Sudo (Kyushu Institute of Technology, Japan)
Co-Chair: Hakaru Tamukoh (Kyushu Institute of Technology, Japan)

OS18-1 *Application of AI Robot Technology for Biophilic Design*
Kairi Manabe, Ryo Miyazono, Keitaro Ito, Tomomi Sudo, Naoto Ishizuka, Akinobu Mizutani, Yuki Anamizu, Etsushi Ueda, Honoka Tamai, Saya Nakano, Leon Furuya, Hakaru Tamukoh, Yuichiro Tanaka, Hirofumi Tanaka
(Kyushu Institute of Technology, Japan)

OS18-2 *Basic Research on the Development of Space Standards for the Use of Service Robots in Housing Using the Urban Renaissance Agency's Housing Complex*
Ren Matsuoka, Kanon Nonoshita, Naoto Ishizuka, Ryohei Kobayashi, Akinobu Mizutani, Hakaru Tamukoh, Hirofumi Tanaka (Kyushu Institute of Technology, Japan)

OS18-3 *An Exhibition Environment with 2D Markers for Guide Robot*
Akinobu Mizutani, Yui Hattori, Naoto Ishizuka, Yuichiro Tanaka, Hirofumi Tanaka, Hakaru Tamukoh
(Kyushu Institute of Technology, Japan)

OS18-4 *Prediction of Timing and Amount of Houseplants Watering by an Echo State Network on Jetson*
Wataru Yoshimura, Koshun Arimura, Ryohei Kobayashi, Akinobu Mizutani, Tomoaki Fujino, Yuichiro Tanaka, Tomomi Sudo, Naoto Ishizuka, Keitaro Ito, Hirofumi Tanaka, Hakaru Tamukoh
(Kyushu Institute of Technology, Japan)

11:30-12:15 OS16 Natural Computing (3)

Chair: Marion Oswald (TU Vienna, Austria)
Co-Chair: Yasuhiro Suzuki (Nagoya University, Japan)

OS16-1 *Modeling Yawning Contagion as a Reaction-Diffusion System: Emergence of Turing Patterns in Behavioral Contagion*
Yasuhiro Suzuki, (Nagoya University, Japan)

OS16-2 *Dominant Region Analysis: A Novel Framework for Quantifying Competitive Reactions Based on the Gillespie Algorithm*
Yasuhiro Suzuki, (Nagoya University, Japan)

OS16-3 *40 Hz sound exposure alters dissolved oxygen levels, gene expression, and colony formation in Saccharomyces cerevisiae BY4741*
Yasuhiro Suzuki, (Nagoya University, Japan)

15:40-16:55 OS11-1 Industrial Artificial Intelligence Robotics (5)

Chair Eiji Hayashi (Kyusyu Institute of Technology, Japan)

OS11-1 *Enhanced Deep Reinforcement Learning for Robotic Manipulation: Tackling Dynamic Weight in Noodle Grasping Task*
Gamolped Prem, Yon Pang Ja Sin, Vjosa Bytyqi, Eiji Hayashi
(Kyushu Institute of Technology, Japan)

OS11-2 *LiDAR-Enhanced Real-Time Tree Position Mapping for Forestry Robots*
M.A Munjer, Tan Chi Jie, Eiji Hayashi (Kyushu Institute of Technology, Japan)

OS11-3 *Kalman-YOLO Improving YOLO Tracking Performance through the Integration of a Kalman Filter for a Beach Cleaning Robot*
Rut Yatigul, Tan Chi Jie, Gamolped Prem, Eiji Hayashi
(Kyushu Institute of Technology, Japan)

OS11-4 *The research of AR System for introducing Industrial Robots*
Takuma Aiko, Gamolped Prem, Eiji Hayashi (Kyusyu Institute of Technology, Japan)

OS11-5 *Research on performance information editing support system for automatic piano - Development of a network model for improved dynamics accuracy -*
Taiyo Goto, Yoshiki Hori, Eiji Hayashi (Kyushu Institute of Technology, Japan)

17:05-17:35 OS11-2 Industrial Artificial Intelligence Robotics (2)

Chair: Eiji Hayashi (Kyusyu Institute of Technology, Japan)

OS11-6 *Research on Tactile-Gripping for Difficult-to-Grasp Objects*
Yoshitaka Sakata, Gamolped Prem, Eiji Hayashi (Kyushu Institute of Technology, Japan)

OS11-7 *Development of a drone obstacle avoidance system based on depth estimation*
Sora Takahashi, Eiji Hayashi (Kyushu Institute of Technology, Japan)

Room 405
10:00-11:15 GS5 Applications I (5)
Chair: Tsutomu Ito (Ube National College of Technology, Japan)

GS5-1 *A Study on Local Airports Contributions to Tourism Industry in Japan*
Tsutomu Ito[1], Seigo Matsuno[1], Makoto Sakamoto[2], Satoshi Ikeda[2], Takao Ito[3]
([1] Ube National College of Technology, Japan), ([2]University of Miyazaki, Japan)
([3]Hiroshima University, Japan)

GS5-2 *Analysis of Careless Mistakes Using Gaze Information*
Ryota Yabe, Noriko Takemura (Kyushu Institute of Technology, Japan)

GS5-3 *A Mathematical Framework for Logit Model in Transportation Mode Choice Analysis*
Ahmad Altaweel[1], Kazuhito MINE[1], Bo-Young Lee[2], Jang-Sok Yoon[2], Hiroaki Wagatsuma[1]
([1]Kyushu Institute of Technology, Japan; [2]Logistics Revolution Korea, Korea)

GS5-4 *A Computational Approach for Global Trade Analysis in Korea Contributing to the Forecasting of Future Efficacy in Global and Domestic Korean Transportations*
Bo-Young Lee[1], Ahmad Altaweel[2], Kazuhito Mine[2], Jang-Sok Yoon[1], Hiroaki Wagatsuma[2]
([1]Logistics Revolution Korea Co., Korea; [2]Kyushu Institute of Technology, Japan)

GS5-5 *Fundamental Research on Athlete Positions Estimation in Indoor Sports at Various View*
Iori Iwata, Yoshihiro Ueda, Kazuma Sakamoto, Riku Kaiba
(Komatsu University, Japan)

11:30-12:15 OS22 Navigation and Tracking (3)
Chair: Chan Gook Park (Seoul National University, Republic of Korea)

OS22-1 *Fine-registered Object LiDAR-inertial Odometry for a Solid-state LiDAR System*
Hanyeol Lee and Chan Gook Park (Seoul National University, Republic of Korea)

OS22-2 *A Fusion Method for Estimating the Walking Direction of Smartwatch Users*
Jae Hong Lee and Chan Gook Park (Seoul Nation University, Republic of Korea)

OS22-3 *Multi-Frame Track-Before-Detect with Adaptive Number of Frame as Noise Level*
Je Hwa Lee, Jae Hong Lee, and Chan Gook Park
(Seoul National University, Republic of Korea)

15:40-16:55 OS15 Robotic Manipulation (5)
Chair: Kensuke Harada (Osaka University, Japan)
Co-Chair: Tokuo Tsuji (Kanazawa University, Japan)
Co-Chair: Akira Nakamura (Saitama Institute of Technology, Japan)

OS15-1 *Surface Stiffness Estimation using Active Strobe Imager*
Taiki Yamaguchi[*1], Kensuke Harada[*1], Koji Mizoue[*2], Makoto Kaneko[*12]
([*1] Osaka University, Japan, [*2] Mizoue Project Japan Corp., Japan)

OS15-2 *Painting Task Planning for Large Structure using a Mobile Manipulator*
Hiroshi Tanaka[*1], Masato Tsuru[*1], Takuya Kiyokawa[*1], Kensuke Harada[*1]
([*1] Osaka University, Japan)

OS15-3 *Real-time Cable Tracking by Wire Segmentation and Coherent Point Drift*
Ryunosuke Yamada[1], Tokuo Tsuji[1], Takahiro Shimizu[2], Shota Ishikawa[1 2], Tomoaki Ozaki[2], Yusuke Sakamoto[1], Tatsuhiro Hiramitsu[1], Hiroaki Seki[1]
([1]Kanazawa University, [2]DENSO CORPORATION, Japan)

OS15-4 *Motion Prediction for Human-Robot Collaborative Tasks Using LSTM*
Kaihei Okada, Tokuo Tsuji, Tatsuhiro Hiramitsu, Hiroaki Seki, Toshihiro Nishimura, Yosuke Suzuki and Tetsuyou Watanabe (Kanazawa University, Japan)

OS15-5 *Individual Recognition of Food in Bulk by using 3D Model of Food*
Yuya Otsu, Tokuo Tsuji, Tatsuhiro Hiramitsu, Hiroaki Seki
(Kanazawa University, Japan)

17:05-17:50 OS24 New Media Interactions (3)
Chair: R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)

OS24-1 *Evaluation of Passive Interaction in XR Chakra Meditation Application Based on Behavioral Biometrics*
P. I. A. Gayathri Bimba (Japan Advanced Institute of Science and Technology, Japan)
Chien-Tung Lin (Tainan National University of the Arts, Taiwan)
R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)
Kazunori Miyata (Japan Advanced Institute of Science and Technology, Japan)

OS24-2 *ThoughtDiffusion: An Interactive Installation for Exploring Neuro-Art from EEG Data with Stable Diffusion Models*
R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)

OS24-3 *PassBy2: Passive Interaction through the Pedestrian Counts and Real-time Weather Information*
Chung Chien-Lin, R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)

Room 406
10:00-11:15 OS1-1 Human-Machine Interface (5)
Chair: Norrima Mokhtar (Universiti Malaya, Malaysia)
Co-Chair: Heshalini Rajagopal (Mila University, Malaysia)

OS1-1 *Non-Invasive Glucose Monitoring Based on Mid-Infrared Spectroscopy*
Puteri Nur Sofea Mohd Zakki, Nani Fadzlina Naim, Hasnida Saad, Wan Norsyafizan Wan Muhamad, Suzi Seroja Sarnin, Norsuzila Ya'acob, Noor Fitrah Abu Bakar
([1]Universiti Teknologi MARA (UITM), Malaysia)

OS1-2 *Investigation of Electromagnetic Radiation (EMR) Before and After Super Brain Yoga Exercise Comparing with Short-Term Memory*
Ros Shilawani S Abdul Kadir, Suzi Seroja Sarnin, Muhammad Afiq Kamil Arif, Faizul Hafizzi Ahmad, Wan Norsyafizan Wan Muhamad, Aziati Husna Awang
(Universiti Teknologi MARA (UITM), Malaysia)

OS1-3 *Preliminary Investigation of Electromagnetic Radiation (EMR) Between Adults and Children*
Ros Shilawani S Abdul Kadir, Aziati Husna Awang, Muhamad Azizularif Mohamad Azizan, Suzi Seroja Sarnin, Suhaila Subahir, Roshakimah Mohd Isa
(Universiti Teknologi MARA (UITM), Malaysia)

OS1-4 *Solar Powered Smart Parcel Box System: Energy Efficient Solution for Modern Deliveries*
Wan Norsyafizan W. Muhammad, AmerulAshraf Zulkepli, Nani Fazlina Naim, Suzi Seroja Sarnin, Ros Shilawani S Abdul Kadir, Md Nor Mat Tan
(Universiti Teknologi MARA (UITM), Malaysia)

OS1-5 *Desktop-Based Expiry Date Application for Retailers Inventory Management*
Nur Hazwani Ahmad Halil[1], Suzi Seroja Sarnin[1], Nani Fazlina Naim[1], Azlina Idris[1], Wan Norsyafizan W. Muhammad[1], Ros Shilawani S Abdul Kadir[1], Md Nor Mat Tan[1], Raudah Abu Bakar[1], Zarina Baharudin Zamani[2]
([1]Universiti Teknologi MARA (UITM), Malaysia), ([2]UTEM, Malaysia)

11:30-12:15 OS1-2 Human-Machine Interface (3)
Chair Norrima Mokhtar (Universiti Malaya, Malaysia)
Co-Chair Heshalini Rajagopal (Mila University, Malaysia)

OS1-6 *Automatic Metal Debris Collection Robot for Laboratory Safety: A Review*
Sophia Fahima Hapizan[1], Ja'aris Samsudin[1], Heshalini Rajagopal[2], Takao Ito[3], Muhammad Amirul Aiman Asri[4], Anees Ul Husnain[5]
([1]Mara-Japan Industrial Insititute, Malaysia), ([2]Mila University, Malaysia)
([3]Hiroshima University, Japan), ([4]Universiti Malaya, Malaysia)
([5]The Islamia University of Bahawalpur, Pakistan)

OS1-7 *Smart Solar LED Street Light with ESP32 Camera Modulee*
Suzi Seroja Sarnin[1], Ros Shilawani S Abdul Kadir[1], Nurul Farhana Zailani[1], Nani Fazlina Naim[1], Md Nor Mat Tan[1], Raudah Abu Bakar[1], Zarina Baharudin Zamani[2]
([1]Universiti Teknologi MARA (UITM), Malysia), ([2]UTEM, Malaysia)

GS5-6 *Overview of the development of low earth orbit satellite navigation enhancement technology*
Dingcheng Tang*, Jinliang Wang, Jianfeng Shan, and Guoji Zou
(Space star technology co, LTD, Beijing, China)

13:00-14:15 OS26-1 Navigating the Digital Frontier: Innovations in the Age of Industry Revolution 4.0 (5)
Chair: Wei Hong Lim (UCSI University, Malaysia)
Co-Chair: Takao Ito (Hiroshima University, Japan)

OS26-1 *Driver State Monitoring Using Pose Estimation: Detecting Fatigue, Stress, and Emotional States for Safer Roads*
Hao Feng Chan[1], Dexter Sing Fong Leong[1], Shakir Hussain Naushad Mohamed[1], Wui Chung Alton Chau[1], Takao Ito[2], Zheng Cai[1], Xinjie Deng[1], Yit Hong Choo[1*]
([1]Deakin University, Australia) ([2]Hiroshima University, Japan)

OS26-2 *Optimizing Face Embedding Sizes and Accuracy in Facial Recognition Systems*
Wui Chung Alton Chau[1], Dexter Sing Fong Leong[1], Shakir Hussain Naushad Mohamed[1], Hao Feng Chan[1], Takao Ito[2], Zheng Cai[1], Xinjie Deng[1], Yit Hong Choo[1*]
([1]Deakin University, Australia) ([2]Hiroshima University, Japan)

OS26-3 *Suspicious Behavior Detection Using Computer Vision*
Dexter Sing Fong Leong[1], Hao Feng Chan[1], Shakir Hussain Naushad Mohamed[1], Wui Chung Alton Chau[1], Takao Ito[2], Zheng Cai[1], Xinjie Deng[1], Andi Prademon Yunus[2], Yit Hong Choo[1*]
([1]Deakin University, Australia) ([2]Hiroshima University, Japan)

OS26-4 *Sign Language Recognition Algorithms Using Hybrid Techniques*
Shakir Hussain Naushad Mohamed[1], Hao Feng Chan[1], Dexter Sing Fong Leong[1], Wui Chung Alton Chau[1], Takao Ito[2], Zheng Cai[1], Xinjie Deng[1], Yit Hong Choo[1*]
([1]Deakin University, Australia) ([2]Hiroshima University, Japan)

OS26-5 *Geographic and Risk Factor Analysis of Non-Communicable Cardiovascular Diseases in Central Java using Machine Learning*
Nurhasanah, Andi Prademon Yunus*(Telkom University, Indonesia)

14:20-15:20 OS26-2 Navigating the Digital Frontier: Innovations in the Age of Industry Revolution 4.0 (4)
Chair: Wei Hong Lim (UCSI University, Malaysia)
Co-Chair: Takao Ito (Hiroshima University, Japan)

OS26-6 *Analysis of Geographical Characteristics and Risk Factors for Non-Communicable Diseases: Diabetes in Central Java Using Random Forest and SHAP*
Ambar Arum Prameswari, Andi Prademon Yunus*(Telkom University, Indonesia)

OS26-7 *Geographic Analysis of Risk Factors for Chronic Respiratory Non-Communicable Diseases Using Machine Learning*
Ayu Susilowati, Andi Prademon Yunus*(Telkom University, Indonesia)

OS26-8 *Role-Play Prediction Using Ontology-Based Graph Convolutional Network Model*
Asyafa Ditra Al Hauna[1], Andi Prademon Yunus[1*], Siti Khomsah[1], Fukui Masanori[2]
([1]Telkom University, Indonesia, [2]Iwate Prefectural University, Japan)

OS26-9 *Experimental Exploration of Neural Style Transfer: Hyperparameter Impact and VGG Feature Dynamics in Batik Motif Generation*
Happy Gery Pangestu[1], Andi Prademon Yunus[1]*, Ratih Alifah Putri[1], Takao Ito[2]
([1]Telkom University, Indonesia) ([2]Hiroshima University)

15:40-16:40 OS26-3 Navigating the Digital Frontier: Innovations in the Age of Industry Revolution 4.0 (4)
Chair: Wei Hong Lim (UCSI University, Malaysia)
Co-Chair: Takao Ito (Hiroshima University, Japan)

OS26-10 *Exploring Non-Communicable Disease Risk Factors on Cancer Rates in Central Java Using Random Forest and SHAP*
Novi Ramadani, Andi Prademon Yunus* (Telkom University, Indonesia)

OS26-11 *Addressing Noise Challenges in CNN-based Pneumonia Detection: A Study Using Indonesian Thoracic Imagery*
Wahyu Andi Saputra*, Andi Prademon Yunus
(Telkom University, Indonesia)

GS4-5 *Developing Low-Cost BCI-Based Brain-Limb Interaction Device with Prosthetic Hand*
Nethika Jayith Rajapakse
(International Bilingual International Science Park, Taiwan)

17:05-17:50 OS2 Computer and Information Engineering (3)
Chair: Norrima Mokhtar (University of Malaya, Malaysia)
Co-Chair: Heshalini Rajagopal (MILA University, Malaysia)

OS2-1 *Efficient Weed Detection in Agricultural Landscapes using DeepLabV3+ and MobileNetV3*
Renuka Devi Rajagopal[1], Manthena Rishit Varma[1], Heshalini Rajagopal[2]
([1]Vellore Institute of Technology, India), ([2]Mila University, Malaysia)

OS2-2 *AI-Based Weed Detection Algorithm using YOLOv8*
Renuka Devi Rajagopal[1], Rethvik Menon C[1], T S PradeepKumar[1], Heshalini Rajagopal[2]
([1]Vellore Institute of Technology, India), ([2]Mila University, Malaysia)

OS2-3 *Novel Gender and Age- Based Detection Technique for Facial Recognition System*
Pratham Gupta[1], Amutha S[1], Dhanush R[1], Heshalini Rajagopal[2]
([1]Vellore Institute of Technology, India), ([2]Mila University, Malaysia)

February 16 (Sunday)

9:40-Registration

Room 403
10:00-11:15 OS12-1 Advances in Field Robotics and Their Applications (5)
Chair: Shinsuke Yasukawa (Kyushu Institute of Technology, Japan)
Co-Chair: Kazuo Ishii (Kyushu Institute of Technology, Japan)

OS12-1 *Practical Exercise on An Autonomous Driving System Using Mobile Devices and IoT Devices for An Agricultural Tractor*
Daigo Katayama, Yuto Nakazuru, Hikaru Sato, Shoun Masuda, Yuya Nishida, Shinsuke Yasukawa, Kazuo Ishii
(Kyushu Institute of Technology, Japan)

OS12-2 *Estimation of Image-Based End-Effector Approach Angles for Tomato Harvesting Robots*
Kizuna Yoshinaga, Hikaru Sato, Kazuo Ishii, Shinsuke Yasukawa
(Kyushu Institute of Technology, Japan)

OS12-3 *Visual-Based System for Fish Detection and Velocity Estimation in Marine Aquaculture*
Raji Alahmad, Dominic Solpico, Shoun Masuda, Takahito Ishizuzuka, Kenta Naramura, Zhangchi Dong, Zongru Li, Kazuo Ishii
(Kyushu Institute of Technology, Japan)

OS12-4 *Evaluating of Tree Branch Recognition Algorithm in Pruning Robots under Augmented Environmental Conditions*
Mohammad Albaroudi, Raji Alahmad, Abdullah Alraee, Kazuo Ishii
(Kyushu Institute of Technology, J Japan)

OS12-5 *Trajectory Analysis for a Mobile Robot Adapted Three Omni Rollers in Constant Roller's Speed*
[1]Kenji Kimura, [1]Kazuki Nakayama, [2]Katsuaki Suzuki, [3]Kazuo Ishii,
([1]National Institute of Technology, Matsue College, [2]Kumamoto Industrial Research Institute,
[3]Kyushu Institute of Technology, Japan)

11:30-12:00 OS12-2 Advances in Field Robotics and Their Applications (2)
Chair: Shinsuke Yasukawa (Kyushu Institute of Technology, Japan)
Co-Chair: Kazuo Ishii (Kyushu Institute of Technology, Japan)

OS12-6 *Cross-Disciplinary Learning Through Manufacturing: Toward Student-Centered STEAM Education*
Kenji Kimura (National Institute of Technology, Matsue College, Japan)

OS12-7 *Development of a Rotary Actuator Capable of Multidirectional Rapid Motion and Variable Stiffness*
[1]Katsuaki Suzuki, [2]Yuya Nishida, [3]Kenji Kjmura, [2] Kazuo Ishii
([1]Kumamoto Industrial Research Institute, [2]Kyushu Institute of Technology,
[3]National Institute of Technology, Matsue College, Japan)

13:00-13:30 OS12-3 Advances in Field Robotics and Their Applications (2)
Chair: Shinsuke Yasukawa (Kyushu Institute of Technology, Japan)
Co-Chair: Kazuo Ishii (Kyushu Institute of Technology, Japan)

OS12-8 *Study of Evaluation Operation Log Analysis Using 2^3- ERC on Matsue National College of Technology*
Takumi Ueda, So Takei, Akira Nakano (National Institute of Technology, Kurume College, Japan)
Kenji Kimura (National Institute of Technology, Matsue College, Japan)
Kazutaka Matsuzaki (Nishinippon Institute of Technology, Japan)

OS12-9 *Efficient Ball Position Estimation for Tennis Court Robot Assistants using Dual-Camera System*
Abdullah Alraee, Raji Alahmad, Hussam Alraie, Mohammad Albaroudi, Kazuo Ishii
(Kyushu Institute of Technology, Japan)

Room 404
10:00-11:15 OS21 Applications of Information Technology for Better Living (5)
Chair: Tetsuro Katayama (University of Miyazaki, Japan)
Co-Chair: Hiroki Tamura (University of Miyazaki, Japan)

OS21-1 *Prototype of MixVRT Which Is a Visual Regression Testing Tool That Highlights Layout Defects in Web Pages*
Naoki Aridome[1], Tetsuro Katayama[1], Yoshihiro Kita[2], Hisaaki Yamaba[1], Kentaro Aburada[1], Naonobu Okazaki[1]
([1]University of Miyazaki, Japan), ([2]University of Nagasaki, Japan)

OS21-2 *Proposal of a Method for Automatic Fill-in Fields Detection and for Labels Assignment to Generate Electronic Forms*
Yuya Kimura[1], Tetsuro Katayama[1], Yoshihiro Kita[2], Hisaaki Yamaba[1], Kentaro Aburada[1], Naonobu Okazaki[1]
([1]University of Miyazaki, Japan), ([2]University of Nagasaki, Japan)

OS21-3 *A Study on Methodology of Measurement for the Physical Burden on Preschool Children*
Sachiko Kido, Hiroki Tamura (University of Miyazaki, Japan)

OS21-4 *Evaluation of Ankle Joint Movements in Frontal Plane for a Normal Coordinated Gait*
Praveen Nuwantha Gunaratne, Hiroki Tamura (University of Miyazaki, Japan)

OS21-5 *Development of a Real-Time Multi-Person 3D Keypoint Detection System Using Stereoscopic Cameras and RTMPose*
Taufik Hidayat Soesilo, Praveen Nuwantha Gunaratne, Hiroki Tamura
(University of Miyazaki, Japan)

11:30-12:00 OS23-1 Mathematical Informatics (2)
Chair: Takao Ito (Hiroshima University, Japan)
Co-Chair: Amane Takei (University of Miyazaki, Japan)

OS23-1 *Simplification of Rip Current Detection by Image Averaging Based on the Number of Wave Breaks*
Ota Hamasuna[1], Leona Kimura[1], Satoshi Ikeda[1], Kaoru Ohe[1], Kenji Aoki[1], Amane Takei[1], Akihiro Kudo[2], Makoto Sakamoto[1*]
([1]University of Miyazaki, Japan), ([2]National Institute of Technology, Tomakomai College, Japan)

OS23-2 *Automated Classification of High-Grade Dried Shiitake Mushrooms Using Machine Learning*
Leona Kimura[1], Ota Hamasuna[1], Kaoru Ohe[1], Satoshi Ikeda[1], Kenji Aoki[1], Amane Takei[1], Akihiro Kudo[2], Kazuhide Sugimoto[3], Makoto Sakamoto[1*]
([1]University of Miyazaki, Japan)
([2]National Institute of Technology, Tomakomai Collage, Japan)
([3]SUGIMOTO Co., Ltd., Japan)

13:00-15:00 OS23-2 Mathematical Informatics (8)
Chair: Takao Ito (Hiroshima University, Japan)
Co-Chair: Amane Takei (University of Miyazaki, Japan)

OS23-3 *Development of a Plant Growing Experience Application for Physically Challenged Children Using VR*
Masatoshi Beppu[1], Masatomo Ide[1], Kaoru Ohe[1], Satoshi Ikeda[1], Kenji Aoki[1], Amane Takei[1], Akihiro Kudo[2], Makoto Sakamoto[1*]
([1]University of Miyazaki, Japan), ([2] National Institute of Technology, Tomakomai Collage, Japan)

OS23-4 *Exploring Social Media's Role in Predicting Stock Market Trends*
Masatoshi Beppu[1], Masatomo Ide[1], Seita Nagashima[2], Satoshi Ikeda[1]*, Amane Takei[1], Makoto Sakamoto[1], Tsutomu Ito[3], Takao Ito[4]
([1]University of Miyazaki, Japan), ([2]MEITEC CORPORATION, Japan), ([3]National Institute of Technology, Ube College, Japan), ([4]Hiroshima University, Japan)

OS23-5 *Development of a Shrine Festival Support Application with Non-Technical Management Features: Functional Evaluation and Sustainability for Future Generations*
Masatomo Ide[1], Masatoshi Beppu[1], Satoshi Ikeda[1], Kaoru Ohe[1], Kenji Aoki[1], Amane Takei[1], Akihiro Kudo[2], Makoto Sakamoto[1*]
([1]University of Miyazaki, Japan), ([2]National Institute of Technology, Tomakomai Collage, Japan)

OS23-6 *A Novel Approach to Reducing Ranking Discrepancies in Tennis Based on Tournament Choices*
Masatomo Ide[1], Masatoshi Beppu[1], Kousei Yano[2], Satoshi Ikeda[1]*, Amane Takei[1], Makoto Sakamoto[1], Tsutomu Ito[3], Takao Ito[4]
([1]University of Miyazaki, Japan), ([2]GENBASUPPORT Co., Japan), ([3]National Institute of Technology, Ube College, Japan), ([4]Hiroshima University, Japan)

OS23-7 *Sound Field Evaluation on Acoustical Experiment with Several Loudspeaker Locations*
Akihiro Kudo[1*], Shun Kubota[1], Amane Takei[2], Makoto Sakamoto[2]
([1]National Institute of Technology, Tomakomai College, Japan), ([2]University of Miyazaki, Japan)

OS23-8 *Parallel High-Frequency Electromagnetic Field Analysis Based on Hierarchical Domain Decomposition Method*
Amane Takei[1*], Akihiro Kudo[2], Makoto Sakamoto[1]
([1]University of Miyazaki, Japan), ([2]National Institute of Technology, Tomakomai College, Japan)

OS23-9 *Adsorption Equilibrium of Selenium Oxyanions Using FeY Mixed Oxides*
Kaoru Ohe*, Amu Wakamatsu, Tatsuya Oshima (University of Miyazaki, Japan)

OS23-10 *Influence of CNN Layer Depth on Spiral Visual Illusions*
Kenji Aoki*, Makoto Sakamoto (University of Miyazaki, Japan)

Room 405
10:00-11:15 OS10-1 Pattern Recognition and Control - 2 (5)
Chair: Zhai Hongshuo (Tianjin University of Science and Technology, China)
Co-Chair: Li Huahao (Tianjin University of Science and Technology, China)

OS10-1 *Intelligent Temperature Control System for Chip Soldering Station*
Huahao Li [1], Junjin Chen [2], Ce Bian [3], Mengfan Zhang [3]
([1]Tianjin University of Science and Technology, China; [2] SMC (Beijing) Manufacturing Co., LTD., China; [3] Tianjin Tianke Intelligent Manufacture Technology Co., LTD., China)

OS10-2 *A Review of Object Detection Techniques Applied to Pest Images*
Hongshuo Zhai[1], Fengzhi Dai[1], Lijiang Zhang[2], Qiang Wang[3]
([1]Tianjin University of Science and Technology, China; [2]Xinjiang Shenhua Biotechnology Co., Ltd, Xinjiang, China; [3]Easy Control Intelligent Technology(Tianjin) Co., Ltd., China)

OS10-3 *Design of an Intelligent Pet Feeding System Based on STM32*
Shuhuan Peng[1], Qiang Wang[2] ([1] Tianjin University of Science and Technology, China,
[2] Easy Control Intelligent Technology (Tianjin) Co., Ltd., China)

OS10-4 *Design of Teaching Attendance System Based on Image Processing*
Shengyu Wang[1], Ce Bian[2] ([1] Tianjin University of Science and Technology, China,
[2] Tianjin Tianke Intelligent Manufacture Technology Co., LTD., China)

OS10-5 *Semi-automatic Leek Harvester Based on Multi-angle Adjustment*
Hongpi Zhao, Xuefeng Jia, Wenqi Fu, Yizhun Peng
(Tianjin University of Science and Technology, China)

11:30-12:00 OS10-2 Pattern Recognition and Control - 2 (2)
Chair: Zhai Hongshuo (Tianjin University of Science and Technology, China)
Co-Chair: Li Huahao (Tianjin University of Science and Technology, China)

OS10-6 *Smart Inspection Guard - Inspection Robot for Unattended Plants*
Liangyu Wang, Yanhong Yu, Yizhun Peng
(Tianjin University of Science and Technology, China)

OS10-7 *Deep Learning Based Integrated Removable Smart Waste Sorting Device*
Yanhong Yu, Liangyu Wang, Yizhun Peng
(Tianjin University of Science and Technology, China)

13:00-15:00 OS20 Applications (8)
Chair: Kasthuri Subaramaniam (University of Malaya, Malaysia)
Co-Chair: Abdul Samad Bin Shibghatullah (Universiti Tenaga Nasional, Malaysia)

OS20-1 *Low-light Image Enhancement with Color Space (Cielab)*
Lee Kok Xiong[1], Kasthuri Subaramaniam[2], Umm E Mariya Shah[1], Abdul Samad Bin Shibghatullah[3], Oras Baker[4]
([1]UCSI University, Malaysia, [2]University of Malaya, Malaysia, [3]Universiti Tenaga Nasional, Malaysia, [4]University of Ravensbourne, England)

OS20-2 *Integrated AI Voice Assistant News Website for Enhancing User Experience - AI-ReadSmart*
Mohammed Mohi Uddin[1], Ghassan Saleh Hussein Al-Dharhani[1], Keoy Kay Hooi[1], Chit Su Mon[2], Kasthuri Subaramaniam[3]
([1]UCSI University, Malaysia, [2]Heriot-Watt University Malaysia Campus, Malaysia, [3]University of Malaya, Malaysia)

OS20-3 *Developing a Mobile Healthcare Application - MyHealth*
Abdulrahman Salmo Alhamada[1], Ghassan Saleh Hussein Al-Dharhani[1], Kasthuri Subaramaniam[2], Raenu Kolandaisamy[1]
([1]UCSI University, Malaysia [2]University of Malaya, Malaysia)

OS20-4 *Developing a Body Posture Detection for Fitness*
Kai Xuan Chong[1], Abdul Samad Bin Shibghatullah[2], Kasthuri Subaramaniam[3], Chit Su Mon[4]
([1]UCSI University, Malaysia, [2]Universiti Tenaga Nasional, Malaysia, [3]University of Malaya, Malaysia, [4]Heriot-Watt University Malaysia Campus, Malaysia)

OS20-5 *Medical Mate: Healthcare and Medical Chat Bot*
Harris Hue Chee Kin[1], Javid Thirupattur[2], Kasthuri Subaramaniam[3], Shabana Anjum Shaik[4]
([1]UCSI University, [2]Sunway University, [3]University of Malaya, [4]Taylor's University, Malaysia)

OS20-6 *Crimes Identification System for Campus Safety and The Threat of Suspicious Student Conduct*
Wong Zhen Bang[1], Kay Hooi Keoy[1], Kasthuri Subaramaniam[2], Sellappan Palaniappan[3], Oras Baker[4]
([1]UCSI University, Malaysia, [2]University of Malaya, Malaysia, [3]Help University, Malaysia, [4]University of Ravensbourne, England)

GS1-5 *Intelligent Path Planning for Robots and Practical Implementation of Programmable Headlights for Autonomous Vehicles*
Farkad Adnan, Abdul Samad Bin Shibghatullah, Mohd Radzi Bin Aridi
(Universiti Tenaga Nasional (Uniten), Malaysia)

GS3-4 *Graphical User Interface (GUI) Design for Mobile Commerce Site for Women Seller in Rural Area*
Kho Irene[1], Dr. Shayla Islam[1], Abdul Samad Shibghatullah[2]
([1]UCSI University, Malaysia), ([2]University Tenaga Nasional, Malaysia)

Room 406

10:00-11:15 OS13 Research Towards the Renewable Energy and the Sustainable Development Goals (SDG's) / Part A (5)

Chair: Ammar A.M. Al Talib (UCSI University, Malaysia)

Co-Chair: Takao Ito (Hiroshima University, Japan)

OS13-1 *Solar-Powered IoT-Based Smart Aquaponic System for Sustainable Agriculture*
Alvi Khan Chowdhury[1], Sarah 'Atifah Saruchi[2], Ammar A.M. Al-Talib[1], Abdirisak Mubarik Muhumed[1], Teh Boon Hong[3], Pavindran A/L Shanmugavel[3], Annanurov Kerim[3], Ng Weng Kent[3]
([1]Monash University Malaysia, Malaysia), ([2]UCSI University, Malysia), ([3]UMPSA, Malaysia)

OS13-2 *AI-Powered Detection of Forgotten Children in Vehicles Using YOLOv11 for Enhanced Safety*
Nur Atikah Jefri[1], Sarah 'Atifah Saruchi[1], Radhiyah Abd Aziz [1], Aqil Hafizzan Nordin[1], Ammar A.M. Al-Talib[2], Zulhaidi Mohd Jawi[3]
([1]UMPSA, Malaysia) ([2]UCSI University, Malaysia) ([3]MIROS, Malaysia)

OS13-3 *Exploring the Performance of YOLOv11: Detecting Compostable and Non-Compostable Kitchen Waste in Real-Time Applications*
Ain Atiqa Mustapha[1], Sarah 'Atifah Saruchi[1], Mahmud Iwan Solihin[2], Fatima Karam Aldeen[2], Ammar A.M. Al-Talib[2] ([1]UMPSA, Malaysia), ([2]UCSI University, Malaysia)

OS13-4 *Comparative Analysis of Machine Learning Algorithms for Rainfall Prediction in Kuantan, Pahang, Malaysia*
Seri Liyana Ezamzuri[1], Sarah 'Atifah Saruchi[1], Ammar A.M. Al-Talib[2]
([1]UMPSA, Malaysia), ([2]UCSI University, Malaysia)

OS13-5 *Autonomous Vehicle Navigation in Highway with Deep Q-Network (DQN) using Reinforcement Learning Approach*
Sumiya Tamanna Fujita[1], Sarah 'Atifah Binti Saruchi[1], Ammar A.M. Al-Talib[2], Nurbaiti Wahid[3]. Siti Nurhafizza Maidin[3], Alvi Khan Chowdhury[4] ([1]UMPSA, Malaysia) ([2]UCSI University, Malaysia) ([3]UiTM Dungun, Malaysia) ([4]Monash Universiti, Malaysia)

11:30-12:15 OS4 Robot Images and Motion (3)

Chair: Miao Zhang (Tianjin University of Science and Technology, China)

Co-Chair: Peng Wang (Tianjin University of Science and Technology, China)

OS4-1 *An improved Laser SLAM algorithm based on Cartographer*
Lei Jiang, Miao Zhang (Tianjin University of Science and Technology, Tianjin, China)

OS4-2 *Semi-Global Stereo Matching Algorithm Based on Optimized Image Preprocessing*
Huaijiao Sha,Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

OS4-3 *Simulation of a 3-DOF Robotic Arm Pick and Place Task Based on Inverse Kinematics*
Songyang Mei, Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

13:00-14:30 OS3 Robot Path Planning (6)
Chair: Shengwei Liu (Tianjin University of Science and Technology, China)
Co-Chair: Yiming Wang (Tianjin University of Science and Technology, China)

OS3-1 *A Novel Path Planning Scheme Based on Improved Bi-RRT* Algorithm*
Shengwei Liu, Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

OS3-2 *Rapidly Exploring Random Tree-back (RRT-Bcak)*
Junsheng Gao, Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

OS3-3 *Path Planning for Mobile Robots Based on Improved A-star Algorithm*
Yuanyuan Zhang, Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

OS3-4 *Improvement of the APF-RRT*-Connect Algorithm for Efficient Path Planning in 3D Environments*
Yiming Wang, Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

OS3-5 *Improved RRT*-Connect based on MATLAB*
Ruofan Zhang, Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

OS3-6 *Rapidly Exploring Gmapping*
Congchuang Han, Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

Room 407
10:00-11:00 OS27 Industrial Revolution (4)
Chair: Hazry Desa (UniMAP, Malaysia)

OS27-1 *Digital Guardians: The Role of AI and Robotics in Protecting Construction Heritage*
Muhammad Azizi Azizan, Nurfadzillah Ishak, Hazry Desa (UniMAP, Malaysia)

OS27-2 *Architectural Memories: AI Redefines Dilapidation Analysis and Conservation*
Muhammad Azizi Azizan, Nurfadzillah Ishak, Hazry Desa, (UniMAP, Malaysia)

OS27-3 *The Future of Robotics in Contract Management*
Muhammad Firdzaus Mat Ros, Muhammad Azizi Azizan, Hazry Desa (UniMAP, Malaysia)

OS27-4 *Leveraging AI to Enhance Extended Producer Responsibility Compliance in Construction Waste Management*
Mohamed Fuad Shahariman, Muhammad Azizi Azizan, Hazry Desa, Nur Amierah Harun (UniMAP, Malaysia)

Farewell Party

Abstract
PS Abstract (3)
PS1 High-fidelity Multi-agent Simulations for Social Systems

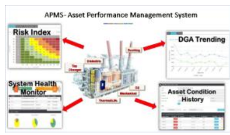
Shinobu Yoshimura, Hideki Fujii (The University of Tokyo)

For realizing the rational and quantitative design of social systems, we have been developing intelligent multi-agent based simulations together with high-fidelity models of social systems. The one is MATES (Multi-Agents based Traffic and Environmental Simulator), and the other is a virtual nursing care process simulator based on a multi-agent model. In this talk, I first describe the objectives of the research and some key technologies, and then introduce their practical applications with verification and validation, i.e. Tram line extension problem in an actual middle sized-city, Okayama, Japan, and an excretion care process in an actual day-care facility for elderly persons.

PS2 Toward 2035: Renewable Energy Innovations Transforming Our Future

Firas Basim Ismail (Universiti Tenaga Nasional (UNITEN), Malaysia)

This paper explores the transformative role of renewable energy innovations in addressing global challenges such as climate change and energy security. By 2035, significant milestones are expected to be achieved through targeted policies and technological advancements. Key innovations in solar technology, wind energy, bioenergy, and energy storage are discussed, along with the role of digitalization and AI in optimizing renewable energy systems. The paper also highlights global and regional efforts, including Malaysia's National Energy Transition Roadmap (NETR), and addresses challenges such as intermittency, high costs, and material safety. The conclusion emphasizes the importance of strategic investments and collaboration to ensure a sustainable energy future.



PS3 Integration of Human – Device Interface: Transforming the Future of Interaction

Norrima Mokhtar, Takao Ito
(Universiti Malaya, Malysia) (Hiroshima University, Japan)

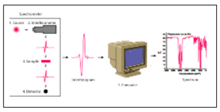
This paper explores advancement in Human-Device Interface technology, focusing on emerging trends like gesture control, brain computer interfaces, speech, eye movement via EOG, eye movement via camera and many more effortless inputs to enable user experience and action. Artificial intelligent techniques, whether supervise, semi- supervise and autonomous, play a major role in processing the input to classification categories for action and data monitoring. Potential applications include automation, healthcare, education, entertainment and the increasingly popular gaming industry. These technologies are redefining the Human-Device collaboration and significantly enhancing user experiences.

OS Abstract
OS1 Human-Machine Interface (7)
Chair Norrima Mokhtar (Universiti Malaya, Malaysia)
Co-Chair Heshalini Rajagopal (Mila University, Malaysia)

OS1-1 Non-Invasive Glucose Monitoring Based on Mid-Infrared Spectroscopy

Puteri Nur Sofea Mohd Zakki, Nani Fadzlina Naim, Hasnida Saad, Wan Norsyafizan Wan Muhamad, Suzi Seroja Sarnin, Norsuzila Ya'acob, Noor Fitrah Abu Bakar
(Universiti Teknologi MARA (UITM), Malaysia)

This paper presents on non-invasive method of glucose monitoring using Mid-Infrared (M-IR) spectroscopy. Glucose samples are prepared and analyzed using M-IR spectroscopy. Using Fourier-transform of the M-IR spectroscopy, we experimentally track variations in the mid-infrared glucose absorption peak. The glucose samples and the relation with diabetic people is also presented in this paper. It is found that as the glucose concentration increases, the wavelengths at which absorbance peaks occur also increase particularly for wavelength range 1400-1470nm.



OS1-2 Investigation of Electromagnetic Radiation (EMR) Before and After Super Brain Yoga Exercise Comparing with Short-Term Memory

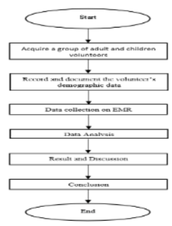
Ros Shilawani S Abdul Kadir, Suzi Seroja Sarnin, Muhammad Afiq Kamil Arif, Faizul Hafizzi Ahmad, Wan Norsyafizan Wan Muhamad, Aziati Husna Awang
(Universiti Teknologi MARA (UITM), Malaysia)

This research is concerned with the investigation of the Electromagnetic Radiation (EMR) on a human body before and after performing Super Brain Yoga (SBY) exercise compared with short-term memory. 20 participants were involved in this research and asked to perform SBY exercise for about two weeks, 20 times in the morning and 20 in the evening at their own comfortable place. From the analysis, it shows that the right side improved by 76% and the left side by 62%. Also, there are increments of 80% of the Digit Span test by the participant after performing SBY. Other than that, 3% of the participants decreased while 17% of them remained the same. In conclusion, this research finding shows that performing SBY exercise gives some benefits to an individual; the EMR of the human body is improving significantly with a better short-term memory of a participant.

OS1-3 Preliminary Investigation of Electromagnetic Radiation (EMR) Between Adults and Children

Ros Shilawani S Abdul Kadir, Aziati Husna Awang, Muhamad Azizularif Mohamad Azizan, Suzi Seroja Sarnin, Suhaila Subahir, Roshakimah Mohd Isa
(Universiti Teknologi MARA (UITM), Malaysia)

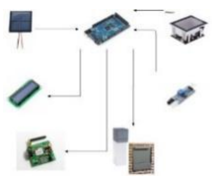
This research is concerned with the initial investigation of electromagnetic radiation (EMR) between adults and children which involved 30 participants from the adult group and 32 participants from the children group. A frequency detector was used to measure the EMR frequencies (in MHz) around 16 points of the human body. The obtained data were assessed by examining the pattern and behavior of captured frequencies, as well as comparing the frequencies of adults and children. From the data analysis, adults have higher frequencies of reading in all body points as compared to children. The analysis also showed that children have better health scores as compared to adults. In conclusion, the EMR emitted from adults and children are significantly different in frequency and have their own characterized frequency patterns and children have better health scores compared to adults.



OS1-4 Solar Powered Smart Parcel Box System: Energy Efficient Solution for Modern Deliveries

Wan Norsyafizan W. Muhammad, AmerulAshraf Zulkepli, Nani Fazlina Naim, Suzi Seroja Sarnin,
Ros Shilawani S Abdul Kadir, Md Nor Mat Tan
(Universiti Teknologi MARA (UITM), Malaysia)

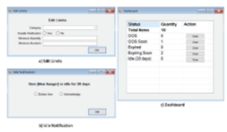
This work introduces an innovative solution: the "Energy Efficiency Smart Parcel Box System Using Solar Energy." The proposed system consists of a network of smart parcel lockers strategically placed in residential and commercial areas. Each locker unit is equipped with solar panels to power its operations, QR code authentication, an infrared sensor for detection and automated door closure, and a user-friendly mobile application. The system's ability to operate independently using solar power makes it a cost-effective and sustainable solution for modern parcel delivery needs. This innovation aligns with the ongoing global efforts to create a greener and more sustainable future, providing a promising solution for the challenges faced in modern logistics and parcel management.



OS1-5 Desktop-Based Expiry Date Application for Retailers Inventory Management

Nur Hazwani Ahmad Halil[1], Suzi Seroja Sarnin[1], Nani Fazlina Naim[1], Azlina Idris[1], Wan Norsyafizan W. Muhammad[1], Ros Shilawani S Abdul Kadir[1], Md Nor Mat Tan[1], Raudah Abu Bakar[1],
Zarina Baharudin Zamani[2]
([1]Universiti Teknologi MARA (UITM), Malaysia), ([2]UTEM, Malaysia)

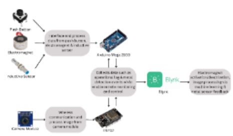
This project is proposed to automate inventory management process by developing a real-time inventory monitoring system and notify user on low stock items and items close to expiration. The desktop-based inventory management application is equipped with a PC and a handheld scanner and involve three main inventory processes; scan incoming inventory, scan outgoing inventory during customer checkout and allow users to manage stock counts and update data if necessary. With this system, issues related to inventory can be resolved quickly based on real-time data monitoring, less time spent for repetitive works while providing a better customer experience.



OS1-6 Automatic Metal Debris Collection Robot for Laboratory Safety: A Review

Sophia Fahima Hapizan[1], Ja'aris Samsudin[1], Heshalini Rajagopal[2], Takao Ito[3], Muhammad Amirul Aiman Asri[4], Anees Ul Husnain[5]
([1]Mara-Japan Industrial Insititute, Malaysia), ([2]Mila University, Malaysia)
([3]Hiroshima University, Japan), ([4]Universiti Malaya, Malaysia)
([5]The Islamia University of Bahawalpur, Pakistan)

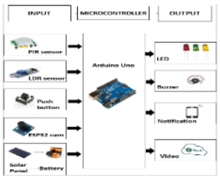
Laboratory safety is paramount, especially when managing hazardous metal debris. This review explores advancements in robotics for waste management, emphasizing autonomous sorting and collection systems driven by AI and machine learning. These technologies enhance the precision and adaptability of robots, enabling accurate detection, classification, and handling of metal debris. By integrating advanced sensors and real-time decision-making, such systems improve resource management and safety in laboratory environments. Challenges such as power efficiency and scalability are also discussed, highlighting future opportunities for optimizing robotic solutions in critical waste management applications.



OS1-7 Smart Solar LED Street Light with ESP32 Camera Modulee

Suzi Seroja Sarnin[1], Ros Shilawani S Abdul Kadir[1], Nurul Farhana Zailani[1], Nani Fazlina Naim[1], Md Nor Mat Tan[1], Raudah Abu Bakar[1], Zarina Baharudin Zamani[2]
([1]Universiti Teknologi MARA (UITM), Malysia), ([2]UTEM, Malaysia)

Streetlights are crucial for the operation of smart cities. The internet of things introduces the idea of smart lighting with solar technology. This study was to determine how smart streetlights can adapt to the environment, which helps people today with their ability to observe their surroundings clearly while avoiding the existence of places where criminals can hide. This project is designed based on LDR sensor to turn ON the LED depends on light intensity while PIR sensor to control the brightness and microcontroller used is Arduino Uno. Automation is intended to reduce manpower with the help of intelligent systems, since the supply of electricity is limited for various of reasons, power consumption is always a top priority.



OS2 Computer and Information Engineering (3)

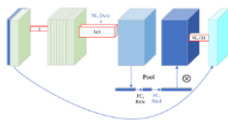
Chair Norrima Mokhtar (University of Malaya, Malaysia)

Co-Chair Heshalini Rajagopal (MILA University, Malaysia)

OS2-1 Efficient Weed Detection in Agricultural Landscapes using DeepLabV3+ and MobileNetV3

Renuka Devi Rajagopal[1], Manthena Rishit Varma[1], Heshalini Rajagopal[2]
([1]Vellore Institute of Technology, India), ([2]Mila University, Malaysia)

Weed detection is a crucial task in precision agriculture, significantly impacting crop yields and reducing the dependency on herbicides. Effective weed management enhances agricultural productivity by ensuring that crops receive adequate nutrients, water, and sunlight, which weeds would otherwise consume. Traditional weed control methods are labor-intensive and often rely heavily on chemical herbicides, which can have detrimental environmental effects. This paper presents a deep learning approach for weed detection, utilizing the DeepLabv3+ model with a MobileNetv3 backbone. This study underscores the potential of integrating advanced deep learning techniques into agricultural practices, paving the way for more sustainable and efficient weed management strategies.



OS2-2 AI-Based Weed Detection Algorithm using YOLOv8

Renuka Devi Rajagopal[1], Rethvik Menon C[1], T S PradeepKumar[1], Heshalini Rajagopal[2]
([1]Vellore Institute of Technology, India), ([2]Mila University, Malaysia)

The development of a country relies heavily on agricultural produce and its related sectors. However, farmers face significant challenges due to the uncontrolled growth of weeds, which reduces their yield. Weed detection is a key step in the removal process, and advances in technology, such as the YOLOv8 model, have simplified this task. YOLOv8 offers improved weed and crop detection, with enhancements of 1.3% and 1.17% in mAP50 and mAP50-95, respectively, over the previous YOLOv5 model. This allows farmers to efficiently identify and eliminate weeds, leading to higher productivity and better crop yields, ultimately supporting the agricultural growth of the country.

OS2-3 Novel Gender and Age- Based Detection Technique for Facial Recognition System

Pratham Gupta[1], Amutha S[1], Dhanush R[1], Heshalini Rajagopal[2]
([1]Vellore Institute of Technology, India), ([2]Mila University, Malaysia)

This paper presents gender and age-based classification methods in facial recognition, addressing challenges from demographic diversity. It employs unsupervised detection using autoencoders to learn facial features, enhancing robustness across populations. Ethical concerns, such as fairness and bias mitigation, are emphasized, ensuring more responsible use of facial recognition. The method effectively reduces biases found in traditional supervised approaches, improving system reliability in diverse real-world applications. The research is significant for developers, policymakers, and researchers focused on ethical AI, offering a novel approach that promotes inclusivity and fairness in facial recognition technologies.

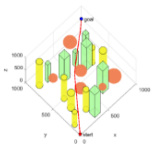
OS3 Robot Path Planning (6)

Chair Shengwei Liu (Tianjin University of Science and Technology, China)
Co-Chair Yiming Wang (Tianjin University of Science and Technology, China)

OS3-1 A Novel Path Planning Scheme Based on Improved Bi-RRT* Algorithm

Shengwei Liu, Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

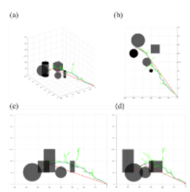
The Bi-RRT* algorithm is a path planning algorithm for industrial robots. In this paper, Bi-RRT* algorithm is studied and improved. The improved Bi-RRT* algorithm reduces the iteration time by introducing artificial potential field method. And, the path cost is reduced and the path smoothness is improved by introducing greedy algorithm. Finally, the improved Bi-RRT* algorithm was simulated in three dimensional environment, and the superiority of the improved Bi-RRT* algorithm was demonstrated by comparative experiments.



OS3-2 Rapidly Exploring Random Tree-back (RRT-Bcak)

Junsheng Gao, Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

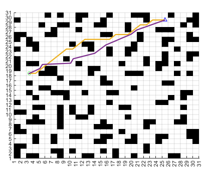
This paper proposes a backtracking Rapidly Exploring Random Tree (RRT-Back) algorithm to reduce the length of the generated path. The proposed algorithm enhances path optimization by employing path backtracking to eliminate redundant nodes and utilizing direct linear connections between discontinuous nodes to shorten path length. To minimize computational expense, the method incorporates cost-effective connections within the already generated path, following the principles of the RRT algorithm. The experimental results demonstrate that the RRT-Back algorithm significantly enhances the feasibility and efficiency of paths in complex environments.



OS3-3 Path Planning for Mobile Robots Based on Improved A-star Algorithm

Yuanyuan Zhang, Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

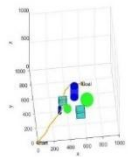
In this paper, an improved A* algorithm is presented, which decreases the search time and path cost by introducing a bidirectional search strategy, enhancing the evaluation function, and eliminating redundant nodes. Through the addition of a corner optimization algorithm, the path smoothness is augmented, and the running speed and reliability of A* are enhanced. Eventually, the superiority of the improved A* algorithm is verified through comparative experiments.



OS3-4 Improvement of the APF-RRT*-Connect Algorithm for Efficient Path Planning in 3D Environments

Yiming Wang, Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

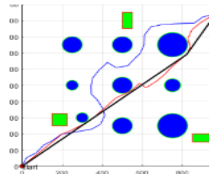
This paper presents an improved APF-RRT*-Connect algorithm for three-dimensional environments. The improved method integrates the APF during path generation to optimize each newly generated node in real-time. This approach reduces the number of node optimization processes, collision detections and optimizes the target nodes in the attractive potential field. Through Matlab simulation, the paper compares the path length and planning time of the traditional APF-RRT*-Connect algorithm and the improved algorithm. The results indicate that the improved algorithm can find shorter paths in less time and enhances the smoothness of the path.



OS3-5 Improved RRT*-Connect based on MATLAB

Ruofan Zhang, Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

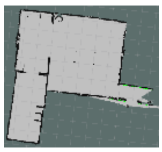
The article analyzes an improved RRT*-Connect path planning algorithm based on MATLAB software. First, the sampling domain of the algorithm is changed to an elliptical sampling domain. Second, adaptive compensation expansion is introduced to accelerate the search speed of the path planning algorithm. Finally, the algorithm is combined with a greedy algorithm to optimize the path. Simulation results show that this algorithm significantly improves both the path length and planning time compared to the traditional RRT*-connect algorithm.



OS3-6 Rapidly Exploring Gmapping

Congchuang Han, Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

This paper presents an improved algorithm for the proposal distribution. The proposed algorithm enhances the range of filter values, the precision of re-sampling, and the accuracy of the map building. Additionally, the computational overhead is reduced, thereby optimizing the issue of the particle degeneration. The comparative experiments demonstrate that the proposed algorithm enhances the mapping precision and speed.



OS4 Robot Images and Motion (3)

Chair Miao Zhang (Tianjin University of Science and Technology, China)
Co-Chair Peng Wang (Tianjin University of Science and Technology, China)

OS4-1 An improved Laser SLAM algorithm based on Cartographer

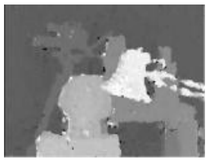
Lei Jiang, Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

SLAM is one of the core technologies in the field of robotics. At present, Laser SLAM has become the mainstream mapping solution for general mobile robots. Cartographer algorithm is one of the mainstream Laser SLAM algorithms, which has attracted much attention because of its high accuracy and suitable for large scenes. However, the effect of sensor data fusion using Unscented Kalman Filter (UKF) is not ideal. Therefore, an improved Cartographer algorithm is proposed in this paper, which uses Adaptive Unscented Kalman Filter (AUKF) to fuse information of sensors, aiming to improve the accuracy of localization and mapping.

OS4-2 Semi-Global Stereo Matching Algorithm Based on Optimized Image Preprocessing

Huaijiao Sha, Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

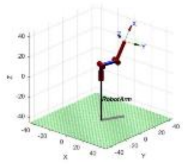
This paper proposes an improved Semi-Global Matching (SGM) algorithm that enhances the quality of disparity images by optimizing image preprocessing. The method applies denoising, contrast enhancement, and Sobel filtering to compute pixel gradients and highlight image edges, thereby improving image quality and reducing noise interference to enhance the clarity of feature boundary. The preprocessed images are used for the SGM algorithm, which improves matching accuracy and adaptability through adaptive sliding windows and dynamic aggregation strategies for cost calculation and aggregation. Experimental results indicate that the improved algorithm enhances the accuracy and robustness of disparity images.



OS4-3 Simulation of a 3-DOF Robotic Arm Pick and Place Task Based on Inverse Kinematics

Songyang Mei, Miao Zhang
(Tianjin University of Science and Technology, Tianjin, China)

This paper proposes a simulation method for grasping and placing a 3-DOF robotic arm based on inverse kinematics. Through the MATLAB GUI, the user enter target coordinates, computes the joint angles by a geometric approach and simulates the motion path to achieve the task operation. In this approach. the robotic arm can precisely reach specified positions, reducing the complexity and error. The position error is calculated by comparing the target with the actual position. The simulation results demonstrate that the robot arm shows a small average error in the five groups of experiments, and the maximum error is maintained within a reasonable range, thereby verifying the accuracy of the method in grasping and placing tasks.



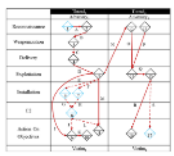
OS5 Intelligent Life and Cybersecurity (7)

Chair I-Hsien Liu (National Cheng Kung University, Taiwan)
Co-Chair Chu-Fen Li (National Formosa University, Taiwan)
Co-Chair Cheng-Han Lin (Fooyin University, Taiwan)

OS5-1 A Diamond Model Approach to Analyzing GhostSec's Intrusion Paths

Cheng-Ying He, Nai-Yu Chen, Jung-Shain Li, I-Hsien Liu (National Cheng Kung University, Taiwan)

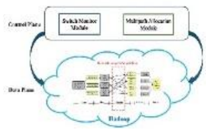
The convergence of Operational Technology (OT) and Information Technology (IT) has heightened risks for critical infrastructure (CI) and industrial control systems (ICS), leading to a surge in diverse and sophisticated OT attacks with severe consequences. Thus, this study combines the Diamond Model with the Cyber Kill Chain to analyze potential attack paths and methods in the GhostSec case, where attackers compromised a Berghof PLC to demonstrate their access capabilities. Understanding these attack paths offers valuable insights into adversary strategies, aiding in the development of defense measures to prevent similar attacks.



OS5-2 A Bandwidth-Aware Routing Mechanism to Control Hadoop Shuffle Traffic over Software-Defined Networking

Ming-Syuan Wu (National Kaohsiung University of Science and Technology, Taiwan)
Cheng-Han Lin (Fooyin University, Taiwan)
Wen-Shyang Hwang (National Kaohsiung University of Science and Technology, Taiwan)
Ce-Kuen Shieh, Mao-Syun Lin (National Cheng Kung University, Taiwan)

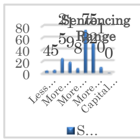
In the Hadoop computing architecture, MapReduce is the main operation program. Between the process of Map and Reduce, servers are necessary to exchange large amounts of data with each other. The shuffle stage exchanges data between servers which will cause the problem of insufficient network bandwidth. In this paper, we proposed an algorithm to distribute Hadoop shuffle traffic and allocate it to all the possible paths. By using the central control network architecture, Software Defined Network (SDN), the proposed mechanism collects the parameters of the network status for allocating shuffle traffic. In distributing the traffic of the Hadoop shuffle stage, the network operation was simulated by the Mininet simulator to build the network topology. The Ryu controller used to be the SDN controller in the simulation. As shown in the simulation results, the proposed mechanism is superior to the Spanning tree protocol and bandwidth-aware algorithm in Hadoop completion time by distributing Hadoop shuffle traffic.



OS5-3 AI Sentencing System: Homicide Case Study in Taiwan

Shih-Chin Lin, Cheng-Tsung Yeh, Chen-Yu Lai, Chih-Yun Chang, Chi-Ju Fu
(Ming Chuan University, Taiwan)

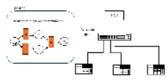
This study examines 198 homicide verdicts from district courts between 2016 and 2023, focusing on sentencing factors based on Article 57 of Taiwan's Criminal Code. It analyzes the most frequently considered factors by judges and the variations in sentencing severity that result from these considerations. In conclusion, this paper addresses potential challenges that may arise in the practical implementation of the sentencing information system and considers how judges might use it as a supplementary tool in sentencing, or whether it could fully guide sentencing decisions.



OS5-4 Inferring ICS Topology and Behavior through Network Traffic Analysis

Chien-Wen Tseng, Jung-Shain Li, I-Hsien Liu (National Cheng Kung University, Taiwan)
Chu-Fen Li (National Formosa University, Taiwan)

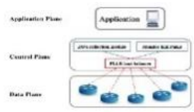
Ensuring the proper operation of industrial control systems (ICS) is a critical issue. Previous studies have focused on observing the controllers directly; however, this research proposes that analyzing network traffic could reduce system interference. This study utilizes a cybersecurity testing platform for dam systems to investigate network traffic as a means to infer the composition of network nodes and the communication behavior between them. By doing so, it provides an alternative perspective for monitoring ICS operations. This model aids in understanding system topology, tracking potential deviations, and ensuring operational stability.



OS5-5 Using fuzzy control routing for dynamic load balancing over Software-Defined Networks

Wen-Shyang Hwang, Ming-Syuan Wu, Sian-Fong Huang
(National Kaohsiung University of Science and Technology, Taiwan)
Cheng-Han Lin (Fooyin University, Taiwan), Yan-Jing Wu (Shih Chien University, Taiwan)
Ming-Hua Cheng (Tzu-Hui Institute of Technology, Taiwan)

Traditional Software-Defined Networks (SDNs) load-balancing and forwarding mechanisms often rely on static path selection. The conventional mechanisms lead to uneven resource utilization and bottlenecks under high traffic. Therefore, existing methods need more flexibility in decision-making. In order to solve the difficulty of setting multi-variable input thresholds, this study introduces a load-balancing algorithm based on fuzzy logic. The algorithm uses fuzzy logic to combine network parameters (queue length, link utilization, link delay, and packet loss rate) as fuzzy inputs. Converting fuzzy input into a score is the key to achieving the optimal traffic allocation in the mechanism. The proposed mechanism improves resource utilization, reduces bottlenecks and reliability in dynamic network load situation. The results indicate that the proposed method achieves higher throughput under high-load conditions. Moreover, the result also maintains low packet loss and delay rates. Based on the fuzzy logic, load balancing thus provides an effective solution for SDN environments.



OS5-6 The Application of AI in the Real Estate Industry: Business Model Innovation Perspective

Li-Min Chuang, Chih-Hung Chen (Chang Jung Christian University, Taiwan)

This paper mainly discusses how Taiwan's real estate agency system uses artificial intelligence (AI) to reshape the traditional business model and create business model innovation. This study uses the Fuzzy Analytical Hierarchy Process (FAHP) and the literature analysis method to construct an analytical framework of four major dimensions based on nine key elements of the business model. Through empirical analysis of relative weights, important propositions are established. The research object is targeted at Tainan, Taiwan. A survey of real estate agency companies and marketing businesses was conducted to show that AI technology drives real estate agency business model innovation, and based on consumer business behavior, it accurately predicts market trends and improves marketing performance, thereby enhancing competitive advantages.

OS5-7 The Impact of AI-Powered Health Monitoring on the Quality of Life and Social Participation of the Elderly: Technology Acceptance Model Perspective

Li-Min Chuang, Zong-Sheng Li (Chang Jung Christian University, Taiwan)

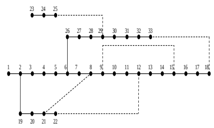
This study investigates the acceptance of AI-powered health monitoring systems among the elderly population in Taiwan. Utilizing the Technology Acceptance Model (TAM) as a theoretical framework, this research employed a combination of literature review and the Fuzzy Analytic Hierarchy Process (FAHP) to analyze questionnaire data collected from the elderly in Tainan, Taiwan. By examining perceived ease of use, perceived usefulness, attitude, and behavioral intention, and calculating the relative weights of these constructs, the study found that the introduction of AI-powered health monitoring systems enabled the elderly to obtain medical advice, thereby reducing the frequency of medical visits and enhancing their ability to live independently.

OS6 Intelligent Algorithm Application (7)
Chair Peng Wang (Tianjin University of Science and Technology, China)
Co-Chair Miao Zhang (Tianjin University of Science and Technology, China)

OS6-1 Reliability Analysis and Optimization of Distribution Network with Distributed Generation

Peng Wang, Mengyuan Hu, (Tianjin University of Science and Technology, China)

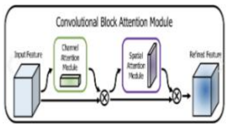
With the access of distributed power sources, the distribution network is facing the problems of voltage quality decline and cost increase. In this paper, the Markov Monte Carlo simulation method(MCMC) is used to verify the impact of the distributed power supply on the distribution network. Secondly, a bi-level joint programming model based on particle swarm optimization (PSO) and K-means clustering algorithm is used to minimize the cost. The results are applied to the upper planning level to improve energy efficiency and economy. Finally, the improved IEEE-33 node system is used to verify the example, and the optimal location of the distributed power supply are analyzed.



OS6-2 Research on Improved PPLCNet Classification Network Based on CBAM Attention Model

Peng Wang, Shengfeng Wang, Qikun Wang, Yuting Zhou
(Tianjin University of Science and Technology, China)

This paper studies pedestrian attribute recognition based on the pplcnet network because it is of great significance in the field of traffic security. Firstly, the research status of pedestrian attribute recognition and common deep learning models are introduced. Secondly, considering that CBAM contains both spatial attention module and channel attention module, we add this attention model to pplcnet to improve performance. Finally, this paper verifies the model through the pa100k dataset and obtains good results.



OS6-3 Optimizing Microgrid Power Dispatch with Integrated Ground Source Heat Pumps Using Cellular Automata

Peng Wang, Shunqi Yang (Tianjin University of Science and Technology, China)

As a new type of energy supply and management system, this paper improves a simulation method based on cellular automata (CA) to optimize power dispatching in microgrids, especially for the ground source heat pump (GHP) system. Firstly, this paper simulates the dynamic behavior and interaction of each unit in the microgrid using cellular automata, addressing uncertainties on both the supply and demand sides. Secondly, this paper uses a two-level optimization method to enhance the maximization of self-consumption and optimize energy flow in the grid. The results show that this method can effectively improve microgrid energy utilization efficiency, reduce operating costs, and enhance system reliability and resilience.

OS6-4 Indoor Personnel Thermal Comfort Monitoring System Based on Mobile Robots

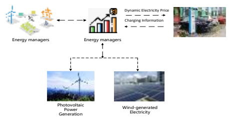
Peng Wang, Zihang Zhou (Tianjin University of Science and Technology, China)

With the advancement of smart building technology, monitoring thermal comfort in indoor environments has become increasingly important. Research has shown that non-invasive infraredography (IRT) technology can effectively predict thermal comfort. This paper explores the use of mobile robots to monitor the thermal comfort of indoor occupants. Mobile robots can collect information of occupants from multiple perspectives, locate and estimate their thermal comfort in real time. This paper first studies how humans perceive temperature and how environmental factors affect comfort., experiments are conducted to evaluate the accuracy and reliability of the data collected by the robot. Finally, the paper analyzes how this data can be translated into thermal comfort.

OS6-5 Market Trading Strategy of Integrated Energy Park from the Perspective of Non-cooperative Game

Peng Wang, Siyi Wang, Liangyu Wang, Chengkai Miao
(Tianjin University of Science and Technology, China)

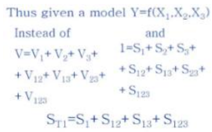
This paper introduces a park trading framework including energy managers, distributed photovoltaic and wind power users and electric vehicle charging service providers, and establishes a non-cooperative game model in which three subjects pursue maximum benefits. Taking a typical winter day in a park as an example, the simulation results show that: in the game equilibrium, energy managers profit from energy supply, distributed photovoltaic and wind power users improve resource utilization and reduce costs through margin online sales, and electric vehicle charging service providers choose low-bid charging to reduce costs and assist users to absorb excess resources and reduce the load of distribution network.



OS6-6 Research on the Sensitivity of Thermal Comfort Using Sensitivity Algorithms Based on Variance and Stochastic Expansion

Peng Wang, Yuting Zhou, Liangyu Wang, Chengkai Miao, Qikun Wang

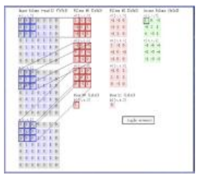
In the research field of modern architectural environment, the sensitivity research of human thermal comfort factors is of crucial significance. In this paper, first of all, common algorithms in the field of sensitivity analysis and data sets for thermal comfort research are elaborated in detail. Secondly. The variance method is taken into consideration. It has the capacity to reflect the fluctuation of the influence that different factors exert on the results. Meanwhile, the stochastic expansion method is also regarded. It is capable of handling complex non-linear relationships. A decision is made to combine these two methods. And the combined methods will be applied to conduct the sensitivity analysis of thermal comfort factors. Finally, the most critical factors for thermal comfort are successfully identified, providing an important basis for the construction and optimization of the thermal comfort prediction model.



OS6-7 Deep Learning Based Infant and Child Monitoring System

Peng Wang, Jiale Jia (Tianjin University of Science and Technology, China)

This study primarily investigates an infant monitoring system based on computer vision and a multi-branch convolutional neural network. Initially, the collected photos are processed, followed by training algorithms using the OpenCV library to obtain a model for detecting infant faces. Subsequently, leveraging the OpenCV and YOLOv8 algorithm technologies to track and analyze infant behavior trajectories and detect faces. Ultimately, functionalities such as target tracking, night vision enhancement, image segmentation, and infant facial detection are achieved with successful detection outcomes.



OS7 Intelligent Robotics and Systems (6)
Chair Kuo-Hsien Hsia (National Yunlin University of Science and Technology, Taiwan)
Co-Chair Jia-Ming Hsiao (National Yunlin University of Science and Technology, Taiwan)

OS7-1 Fuzzy-Controlled Multi-Valve Pneumatic Soprano Recorder Auto-Playing and Score Recognition System

Chun-Chieh Wang*, Chung-Wen Hung, Kuo-Hsien Hsia, Chian C. Ho, Ying-Yuan Yao (Nation Yunlin University of Science and Technology, Taiwan)

This study presents advanced enhancements in score recognition and soprano recorder performance. An improved method for removing staff lines enhances musical symbol distinction, while refined note classification ensures accurate pitch assignment. The playback system replaces motor-driven actuators with a solenoid-based 9-valve configuration, optimizing speed, reducing noise, and ensuring precise air pressure control across an extended pitch range. Automated tuning using fuzzy control significantly improves efficiency and accuracy. Experimental results demonstrate substantial improvements in tonal precision and overall performance quality for automated soprano recorder systems.

OS7-2 Development of agricultural robots based on ROS

Jr-Hung Guo*, Kuo-Hsien Hsia (National Yunlin University of Science and Technology, Taiwan)

The agricultural industry has become an important issue in many countries due to drastic changes in weather and a reduction in the number of manpower willing to engage in agriculture. Therefore, this study attempts to develop an agricultural robot platform so that robots can assist people in agricultural work. This research uses ROS as the software foundation. Through this convenient software foundation, different agricultural robots can be quickly developed. This research uses this architecture to develop lawn mowers and leaf sweepers, which can be easily converted to different agricultural applications in the future.

OS7-3 The ROS-based web information center of small manipulator

Yue-Jie Wang, Jia-Ming Hsiao*, Shao-Yi Hsiao
(National Yunlin University of Science & Technology, Taiwan)

A ROS-based web information center with respect to small manipulator is developed to show the feasibility of the one for factories. Small ROS-based 6-axis manipulators for teaching and research are utilized to simulate the applications of robot arms in the factories. The task of robot arm is designed to recognize the object color and then to grip the object to the desired storage area. Production operation data and status information from the robot arm are transmitted through the ROS rosbridge_suite to the web interface for display and saved in a database. A start-stop function is also implemented to remotely start and stop the operation of robot arm.

OS7-4 Image-assisted Assembly and Disassembly Process Using TM Six-Axis Collaborative Robotic Arm

Kuo-Hsien Hsia, Yi- Yan Liao, Ching-Yuan Pan
(National Yunlin University of Science & Technology, Taiwan)

This paper explores the development of the TM collaborative robot arm in industrial applications. With its in-house developed TMflow software, TM robot streamlines the intricate human-machine interface of industrial robots and modularizes various tool functions, allowing operators to quickly familiarize themselves. The focus of this paper is on programming the TM collaborative robot arm using TMflow to achieve automatic image-assisted localization for assembly and disassembly. Collaborative arms improve the accuracy and efficiency of assembly and disassembly, reducing manual errors and wasted time. In terms of safety, compared with general robotic arms, collaborative arms are safer, allowing them to cooperate safely with human workers and reducing the incidence of workplace accidents.

OS7-5 Quality Inspection of PVC Shoe Chopsticks: A Research Study

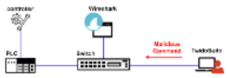
Shu-Li Pai, Kuo-Hsien Hsia, Chian-Cheng Ho
(National Yunlin University of Science and Technology, Taiwan)

In the past decade, the advancement of artificial intelligence technology has led to substantial growth in industrial automation. However, most manufacturing industries are still at the stage of physical labor automation, especially in the area of product inspection, where manual inspection is predominantly used. Taking a PVC shoe chopstick factory as an example, the production line operates 24 hours a day, but quality inspection requires employees to conduct checks on the next working day. This approach makes it difficult to promptly address any defects that occur during production. By incorporating computer vision technology into the PVC production line to track and measure products, it is possible to reduce the defect rate in production and decrease the personnel costs and time delays associated with manual inspection.

OS7-6 Obstructing PLC Operations through Modbus Command Manipulation

Nai-Yu Chen, Cheng-Ying He, Jung-Shain Li, Chu-Sing Yang, I-Hsien Liu
(National Cheng Kung University, Taiwan)

Security vulnerabilities in Programmable Logic Controllers (PLCs) within Industrial Control Systems (ICS) using the Modbus/TCP protocol pose significant risks, particularly through stop-and-start command injection attacks that impact PLC operations and cause severe industrial consequences. Supported by Taiwan's National Science and Technology Council (NSTC) and the Water Resources Agency, this research establishes a cybersecurity testbed for water resource systems to investigate these threats. Unauthorized or forged commands are shown to manipulate PLC configurations and ladder logic diagrams, revealing critical weaknesses. Flowchart analyses and Modbus packet examinations highlight the risks and offer actionable insights into effective defense mechanisms for enhancing ICS security.



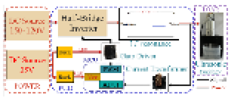
OS8 Intelligent Systems and Applications (8)

Chair Chun-Liang Liu (National Yunlin University of Science and Technology, Taiwan)
Co-Chair Yuting Hsiao (National Yunlin University of Science and Technology, Taiwan)

OS8-1 MCU Ultrasonic oscillator driver with digital frequency sweep function

Chung-Wen Hung, Chun-Chieh Wang, Yu-Hsing Su
(National Yunlin University of Science and Technology, Taiwan)

An Ultrasonic oscillator driver with digital frequency sweep function is proposed in this paper. Ultrasonic transducers are widely utilized in various applications, such as liquid atomization. The drive circuit causes the ultrasonic transducer to vibrate, while the attached atomization component converts the liquid into fine particles. In this paper, the LC resonant circuit is adopted to drive the ultrasonic transducer. Due to small variations in the resonant frequency of each transducer, the optimal operating frequency also varies and may change as physical conditions change. Then, microcontroller units (MCUs) are used to control circuit switches to achieve frequency adjustment, scanning and tracking, so that the ultrasonic oscillator works in the best state.



OS8-2 Design and development of foot pressure sensing massage stick

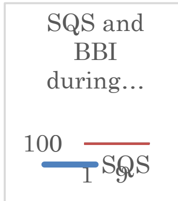
Yuting Hsiao, Dengchuan Cai, Chung-Wen Hung
(Nation Yunlin University of Science and Technology, Taiwan)

There are three key points in the operation of the Foot Massage Stick: sequence, direction and force. However, the force is not easy to be described and not easy to be learned. In order to provide the operator with visualization and information about the force during the operation, a pressure sensing massage stick was developed in this study. The features developed in this study are as follows. 1) The function of the assistive device is in line with the precision level of commercially available instruments. 2) The force and time duration of the operation can be displayed at any time during the execution process. 3) At the end of the execution, the maximum, average, standard deviation, and time duration of the force of the operation can be presented. This visualizes the force of the operation and makes it easy for the learner to check and meet the requirements for use.

OS8-3 Effects on physiological indicators of foot massage using a pressure sensing massage stick

Dengchuan Cai, Yuting Hsiao, Chung-Wen Hung
(Nation Yunlin University of Science and Technology, Taiwan)

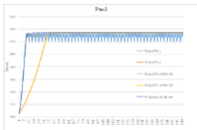
Foot massage is often used as a complementary and alternative therapy. This study uses a special foot pressure-sensing massage stick to massage the soles of the feet, and tests the effects on the physiological indicators of the massaged person after foot massage. After 3 sessions of reflexology, the results show: 1) Sleep quality and body energy scores improved. Women showed greater improvement than men. 2) Decrease in stress index and body age. For women, the decline was lower than that for men. 3) Blood oxygen concentration, respiratory rate, and heart rate amplitude do not change much and tend to decrease. The degree of decline was the same for both sexes. The planning and results of this study can provide a reference for the design of foot massage aids.



OS8-4 Study on Maximum Power Point Tracking Technology for Solar Power Systems Using Power Variation to Adjust Step Response

Chun-Liang Liu, Chung-Wen Hung, Yi-Feng Luo, Guan-Jhu Chen, Cheng-Sin Hu
(Nation Yunlin University of Science and Technology, Taiwan)

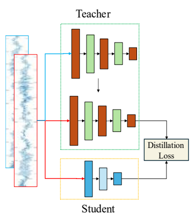
The Perturb and Observe (P&O) method is a popular MPPT algorithm for photovoltaic power generation systems. However, it has a trade-off between step size and transient response. This paper proposes an adjustable step size factor to address this issue. The method starts with a larger initial step size, which is then multiplied by a factor to reduce the perturbation step size. The method retains the transient advantage of a larger step size while reducing power loss by minimizing steady-state oscillations. It achieves faster perturbation convergence and up to 99.98% accuracy in steady-state tracking.



OS8-5 Bearing faulty prediction based on knowledge distillation

Chun-Liang Liu, Zheng-Jie Liao, Chung-Wen Hung
(National Yunlin University of Science and Technology, Taiwan)

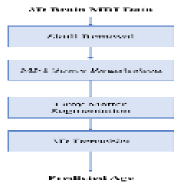
This paper employs knowledge distillation to train teacher and student models using different motor bearing vibration datasets. The signal is transformed from the time domain to the frequency domain using Fast Fourier Transform (FFT), and a Convolutional Neural Network (CNN) model is used to recognize the bearing conditions. The teacher model is a deeper model trained with a larger dataset, while the student model is a shallower model trained with less data. The student model is guided by the soft labels provided by the teacher model. The results demonstrate that knowledge distillation improves the student model's recognition performance and enables knowledge transfer, allowing the student model to achieve good recognition accuracy even with limited training data.



OS8-6 Accurate Brain Age Prediction Through Advanced Preprocessing and 3D DenseNet-50 Modeling

Ting-An Chang, Chiang-Ming Yeh, Chun-Liang Liu
(National Yunlin University of Science and Technology, Yunlin, Taiwan)

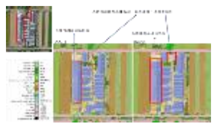
An innovative method for predicting brain age is proposed in this study. The approach consists of three key stages. First, during the data collection phase, high-resolution T1-weighted Magnetic Resonance Imaging (T1W-MRI) scans were gathered to ensure a sample with broad age distribution and diversity. Secondly, in the data preprocessing stage, several critical steps were implemented: skull removal was performed to eliminate interference from non-brain tissue, spatial standardization to the Montreal Neurological Institute (MNI) space was conducted to ensure comparability across subjects, and brain tissue segmentation was applied to extract gray matter, white matter, and cerebrospinal fluid. Finally, for the model architecture, a 3D DenseNet121 network was selected based on its proven efficacy in processing medical imaging data and capturing complex spatial features. This multi-step methodology was designed to develop an accurate and robust brain age prediction model.



OS8-7 Intelligent agricultural landscape identification system

Ching Ju Chen[1], Yu-Cheng Chen[1], Jing-Yao Lin[1], Rung-Tsung Chen[2], Candera Wijaya[3]
([1]National Yunlin University of Science and Technology, Taiwan) ([2]Taiwan Biodiversity Research Institute, Taiwan) ([3]Agricultural Engineering Research Center, Taiwan)

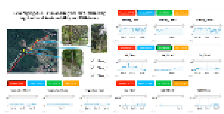
This paper addresses Taiwan's agricultural land decline and its impact on food supply and biodiversity. It proposes a semantic segmentation-based recognition system using drone technology to classify agricultural landscapes, watersheds, and habitats. Two models, U-Net with VGG16 and U-Net with ResNet50, are tested for semantic segmentation of farmland images. Results show that while these models effectively classify landscape categories, misclassification occurs for similar features like grassland, fallowland, and dry land. The paper suggests improving model accuracy by increasing dataset size and sample diversity.



OS8-8 Leveraging AIoT Visual Analytics for Optimizing Agricultural Sustainability and Efficiency

Hsueh-Yen Shih[1], Xi-Wei Lin[2], Zhao-Sheng Chen[2], Ying-Cheng Chen[3], Ching-Ju Chen[2]
([1]Agricultural Engineering Research Center, Taiwan) ([2]National Yunlin University of Science and Technology, Taiwan) ([3]Tainan District Agricultural Research and Extension Station, Taiwan)

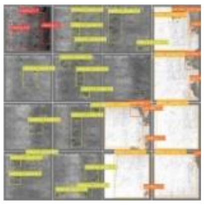
This study integrates artificial intelligence (AI) technology and Internet of Things (IoT) sensors to enhance precision management and real-time monitoring of pests and diseases in agriculture, promote transparency of agricultural data and visual decision-making, and achieve optimized resource management, real-time pest control to reduce the use of pesticides and improve land reusability. By deploying sensors in the farmland and using AI to analyze data, the company will establish an intelligent decision-making platform to provide farmers with real-time, forward-looking management advice and collaborate to realize the sustainable development of smart agriculture and the environment.



OS9 Pattern Recognition and Control 1 (7)
Chair Sun Haozhe (Tianjin University of Science and Technology, China)
Co-Chair Li Fangyan (Tianjin University of Science and Technology, China)

OS9-1 A Study on Surface Defect Detection Algorithm of Strip Steel Based on YOLOv8n
Haozhe Sun[1], Fengzhi Dai[1], Junjin Chen[2]
([1]Tianjin University of Science and Technology, [2]SMC (Beijing) Manufacturing Co., LTD., China)

Hotrolled steel strip has been extensively applied in industrial production and processing due to its outstanding properties. Nevertheless, during the production procedure, as a result of technological constraints, defects will inevitably occur on the surface of the steel strip, and they significantly influencing the performance and safety of the steel strip. Hence, how to detect the surface defects of steel strips has turned into the key point. In this paper, an enhanced YOLOv8n network model is proposed to make it applicable for the surface defect detection tasks of hot rolled steel strips. The mAP50 of the enhanced model is superior to that of the original YOLOv8n model, particularly for small target defects.



OS9-2 Prediction of Winter Wheat Growth Trends Based on NDVI Vegetation Index
Lu Kang, Jiahao Xie, Chunli Li, Haoran Gong, Fengzhi Dai
(Tianjin University of Science and Technology, China)

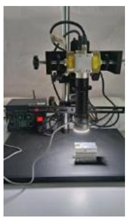
As modern technology emerged, the level of agricultural remote sensing has been further improved. This paper takes winter wheat as the research object, studying on the area of Liangshan in Shandong Province, where the planting coverage of winter wheat is high. Image preprocessing is carried out using ArcGIS, combined with ENVI to invoke satellite data in the near-infrared and infrared bands to calculate the NDVI index from the regreening stage to the maturity stage of winter wheat in this area. During the maturity stage of winter wheat, NDRE is used instead of NDVI to solve the problem of inaccurate NDVI measurement in high-density vegetation coverage. The simulation results show that the test data matches the actual winter wheat output value.

OS9-3 A Study on Artemia Culture System and Its Application
Wanying Zhang, Yicheng Wu, Ziting Zhang,Yumei Huang
(Tianjin University of Science and Technology, China)

Aimed at the shortcomings of the low efficiency and high cost of Artemia culture, this paper proposes a high-density Artemia culture system based on the Internet of Things technology. The system detects and controls the breeding environment through sensors and actuators, and uses a cloud platform to analyze and process the collected data. Automation control and remote monitoring of the system reduce the cost of breeding and human resource. The system's Internet of Things technology provides scientific basis and decision support for Artemia culture.

OS9-4 Machine Vision-Based Chamfer Detection for Metal Parts
Shangying Han[1], Kaili Guo[1], Yanzi Kong[1], Yanliang Gong[1], Junjin Chen[2], Ce Bian[3], Mengfan Zhang[3]
([1]Tianjin University of Science and Technology, [2]SMC (Beijing) Manufacturing Co., LTD.,
[3] Tianjin Tianke Intelligent Manufacture Technology Co., LTD., China)

This paper introduces a detection system specifically designed for chamfering in metal holes, aimed at achieving precise detection of the chamfers. Chamfering, as a process of beveling the edges or corners of metal parts, plays a crucial role in the subsequent machining and assembly stages. Through multiple experimental validations, this paper employs an industrial camera with a telecentric lens to capture images of the metal chamfers, achieving optimal results. This paper utilizes computer vision techniques to accurately identify the location of the chamfers and delineate their dimensions. A comprehensive analysis of the chamfer radius effectively determines the presence of defects.



OS9-5 Deep Guard Dog - AI-Based Night Intrusion Detection Mobile Phone Software

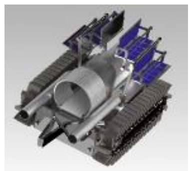
Keming Chen, Jiaxin Wang (Tianjin University of Science and Technology, China)

This article introduces an Android mobile app called "Super Electronic Watchdog", which aims to solve the problem of home security. The application utilizes Android Studio, NCNN framework and Opt2Ada night vision algorithm to realize humanoid object detection and night image enhancement. Users can switch the camera, select the humanoid detection model and CPU/GPU operation mode, and activate night vision through the app. The application has vibration and voice alarm functions to alert the user that someone has entered the monitored area. The software is divided into Native layer and Java layer, using C++ and Java development, the overall design structure is clear, efficient and practical.

OS9-6 Development of an Amphibious Surface Garbage Collection Robot and Its Applications

Yu Su, Xin Wang, Long Shen, Zhenxing Liu, Xinrui Zhao, Xin Lin, Mengchen Huo, Yawen Qiao, Yan Zhang
(Tianjin University of Science and Technology, China)

This paper presents an amphibious water - surface garbage - collecting robot. It incorporates innovative technologies such as efficient garbage collection, accurate identification and classification, stable amphibious operation, and sustainable energy utilization. The double - four - bar linkage and "three - pipe" collection device ensure effective collection and classification. The amphibious crawler provides buoyancy and land - moving ability. Visual recognition technology has high accuracy. GPS automatic cruise and solar charging system are also included. The physical model meets design requirements, aiming to provide an efficient and intelligent solution for water - surface garbage disposal.



OS9-7 Design of an Intelligent Orbital Inspection Robot Based on Machine Vision and Ultrasonic Guided Waves

Xingwang Feng, Suqing Duan (Tianjin University of Science and Technology, China)

This paper introduces a track inspection robot based on machine vision and ultrasonic guided wave, integrating BeiDou positioning and autonomous driving system, which can efficiently detect defects such as track cracks and settlements. The innovative wheel-foot switching structure and electro-hydraulic leveling platform enhance the multi-terrain adaptability, and the combination of particle swarm optimization and support vector machine algorithm realizes defect recognition. Tests show that the system has high detection accuracy and stability, providing a technical reference for intelligent track inspection.

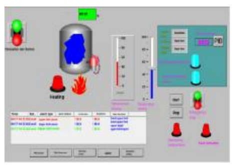
OS10 Pattern Recognition and Control - 2 (7)

Chair Zhai Hongshuo (Tianjin University of Science and Technology, China)
Co-Chair Li Huahao (Tianjin University of Science and Technology, China)

OS10-1 Intelligent Temperature Control System for Chip Soldering Station

Huahao Li [1], Junjin Chen [2], Ce Bian [3], Mengfan Zhang [3]
([1] Tianjin University of Science and Technology, China; [2] SMC (Beijing) Manufacturing Co., LTD., China;
[3] Tianjin Tianke Intelligent Manufacture Technology Co., LTD., China)

Chip is the general name of semiconductor component products, mainly by the semiconductor material, solid state electronic devices, silicon wafers and other materials processed through a number of responsible processes. According to the functional requirements, the intelligent temperature control system of the chip welding bench is designed with PLC as the control core. In the PLC system design using PID instructions, when the temperature is close to the specified temperature using low-power heating, when the temperature difference is large high-power heating. In this paper, we design the temperature control system of chip soldering bench with Siemens S7-200 PLC as the control core.



OS10-2 A Review of Object Detection Techniques Applied to Pest Images

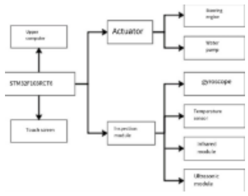
Hongshuo Zhai [1], Fengzhi Dai [1], Lijiang Zhang [2], Qiang Wang [3]
([1] Tianjin University of Science and Technology, China; [2] Xinjiang Shenhua Biotechnology Co., Ltd, Xinjiang, China; [3] Easy Control Intelligent Technology (Tianjin) Co., Ltd., China)

In agricultural information management, crop pest control has always been an important topic, and the image detection technology of small target pests is particularly critical in this process. At present, the technology faces challenges such as difficult data collection and insufficient robustness. This paper first introduces the development of object detection technology and its application in the field of agriculture, then analyzes the challenges of information-based pest control, discusses the research progress of pest dataset construction, image data augmentation technology and object detection algorithm, and finally points out the future research direction in this field.

OS10-3 Design of an Intelligent Pet Feeding System Based on STM32

Shuhuan Peng [1], Qiang Wang [2] ([1] Tianjin University of Science and Technology, China; [2] Easy Control Intelligent Technology (Tianjin) Co., Ltd., China)

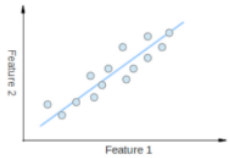
This design leverages the STM32F103RCT6 microcontroller to develop a smart pet feeding system, enabling automated food and water dispensing with a user-friendly interface and remote monitoring capabilities. The system uses infrared and ultrasonic sensors to control feeding with precision, preventing overfeeding or accidental dispensing. A gyroscope and temperature sensor ensure operational safety by monitoring device stability and environmental conditions. Users can easily set feeding schedules through a touchscreen interface, while Bluetooth connectivity allows real-time notifications to smart devices, keeping pet owners informed.



OS10-4 Design of Teaching Attendance System Based on Image Processing

Shengyu Wang [1], Ce Bian [2] ([1] Tianjin University of Science and Technology, China; [2] Tianjin Tianke Intelligent Manufacture Technology Co., LTD., China)

Traditional classrooms often require teachers to roll call one by one in class, which not only affects the length of the class but also affects the quality of the class, and there is no timely data feedback, resulting in the lag of the work of many college students. This topic mainly uses computer simulation software for algorithm research, which is mainly divided into four parts, the first is the initial establishment of the face library, the second is the use of PCA algorithm for face image dimensionality reduction in face recognition, the Euclidean distance is used again for face closest matching, and finally the function is realized by GUI interface.



OS10-5 Semi-automatic Leek Harvester Based on Multi-angle Adjustment

Hongpi Zhao, Xuefeng Jia, Wenqi Fu,Yizhun Peng
(Tianjin University of Science and Technology, China)

The team is committed to solving the domestic leek harvesting process of excessive human input, high cost, intelligence, low level of mechanization, to provide users with a diversified range of high-performance semi-automatic leek harvester equipment and solutions, which can make the leek production safer, time-saving, labor-saving. The team independently researched and developed leek harvester which can walk independently and harvest automatically, adopting new adjustable mechanical structure, artificial intelligence algorithm and human-computer interaction software application, which makes the domestic automatic harvesting gradually become possible.

OS10-6 Smart Inspection Guard - Inspection Robot for Unattended Plants

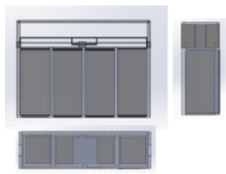
Liangyu Wang, Yanhong Yu, Yizhun Peng (Tianjin University of Science and Technology, China)

This document describes an intelligent inspection robot, based on Arduino and Raspberry Pi, with autonomous navigation and video surveillance. The robot uses infrared sensors for tracing, RFID for localization, and uploads the captured video to the server. The innovative integration of wireless charging technology realizes the unattended function, as well as the self-designed camera clamping mechanism. Key technologies cover differential control, wireless charging, data communication and server design. Tests show that the robot can improve inspection efficiency and quality, save labor costs, and comply with the trend of intelligent manufacturing.

OS10-7 Deep Learning Based Integrated Removable Smart Waste Sorting Device

Yanhong Yu, Liangyu Wang, Yizhun Peng (Tianjin University of Science and Technology, China)

In this study, an intelligent waste sorting device based on Inception v3 and migration learning is developed to achieve fast and accurate waste recognition and sorting through deep learning and sensor fusion techniques. The device is designed to be detachable and adaptable to existing bins, with the ability to continuously learn new waste types. Through real-time data transmission, the device supports remote monitoring and management, which effectively improves the efficiency of waste classification and is important for urban environmental protection.



OS11 Industrial Artificial Intelligence Robotics (7)

Chair Eiji Hayashi (Kyusyu Institute of Technology, Japan)

OS11-1 Enhanced Deep Reinforcement Learning for Robotic Manipulation: Tackling Dynamic Weight in Noodle Grasping Task

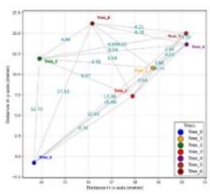
Gamolped Prem, Yon Pang Ja Sin, Vjosa Bytyqi, Eiji Hayashi (Kyushu Institute of Technology, Japan)

Handling food items with dynamic weight changes over time, which alter physical properties such as shape, size, and weight, poses significant challenges, particularly when precise output weight is required. This study introduces an enhanced deep reinforcement learning framework for robotic manipulation, focusing on the task of spaghetti grasping. Building on prior research, we propose a data augmentation strategy that simulates diverse environmental conditions, including variations in image observations and the physical properties of spaghetti, to improve models. The model is validated using metrics such as grasp success rate, average grasp time, and generalization score under varying environmental conditions. This work advances the robustness of robotic models in previously unseen environments.

OS11-2 LiDAR-Enhanced Real-Time Tree Position Mapping for Forestry Robots

M.A Munjer, Tan Chi Jie, Eiji Hayashi (Kyushu Institute of Technology, Japan)

This article evaluates the effectiveness of an autonomous robot in creating a real time tree pose map in both simulated and experimental environments. This paper also demonstrates the implementation of FastSLAM on a four-wheeled differential-drive robot, integrating real-time tree detection and tracking through LiDAR-based point cloud data. An algorithm is proposed to generate a map showing both the robot's path and detected tree positions during movement. Performance Metrics Analysis revealed a high True Acceptance (TA) rate, confirming accurate tree position estimation. Experimental results validated the algorithm's reliability, showcasing strong distance accuracy with minimal discrepancies between actual and estimated positions. These findings highlight the system's potential for advancing forestry management through precise robotic navigation and mapping.



OS11-3 Kalman-YOLO Improving YOLO Tracking Performance through the Integration of a Kalman Filter for a Beach Cleaning Robot

Rut Yatigul, Tan Chi Jie, Gamolped Prem, Eiji Hayashi (Kyushu Institute of Technology, Japan)

Ocean waste poses a significant threat to both human and marine life as industries and individuals continue to dump garbage into the ocean. Sea creatures are poisoned by materials such as plastics and chemicals, which in turn contaminate humans who consume them. This paper introduces an innovative approach using Image Instance Segmentation with YOLOv8 to segment and track beach garbage. However, YOLOv8's object tracking struggles in dynamic environments with challenges like occlusion, shadows, and perspective changes in RGB frames. To address this, the author presents Kalman-YOLO, combining the Kalman Filter with YOLO for improved performance. Results show notable performance improvement, especially in tracking garbage for the Beach Cleaning Robot.

OS11-4 The research of AR System for introducing Industrial Robots

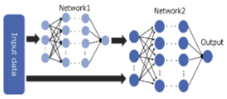
Takuma Aiko, Gamolped Prem, Eiji Hayashi (Kyusyu Institute of Technology, Japan)

In recent years, Japan has been suffering from a labor shortage in all industries. By introducing robots, it is possible to reduce manpower, and it is expected to contribute to resolving labor shortages. However, the introduction of industrial robots is not easy due to the high cost of equipment and system integration. Therefore, we are developing an Augmented Reality (AR) application for the purpose of introducing robots. In this study, we developed a mobile AR system that can check the movement path of a robot when it is introduced without using the actual robot and confirmed its operation.

OS11-5 Research on performance information editing support system for automatic piano - Development of a network model for improved dynamics accuracy-

Taiyo Goto, Yoshiki Hori, Eiji Hayashi (Kyushu Institute of Technology, Japan)

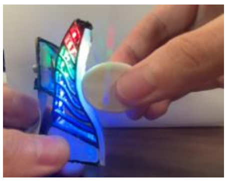
The automatic piano player, which was previously developed in this laboratory, is attached to the keys and pedals of a grand piano, and enables accurate keystrokes and pedal operation with appropriate control from a computer. To control the device, music data is required, but if music score data is simply input into the device, the performance will be flat, and will not sound like a human being, which is the goal. This is because pianists play with their own intonation when they play. Previous research has developed a system that uses deep learning to predict performance information, but the accuracy of predicting sound volume (Velo) was not good. This research aims to enhance the accuracy of Velo in automatic piano performance. A new deep learning system combining two networks was developed to address limitations in existing methods.



OS11-6 Research on Tactile-Gripping for Difficult-to-Grasp Objects

Yoshitaka Sakata, Gamolped Prem, Eiji Hayashi (Kyushu Institute of Technology, Japan)

This study focuses on the automation of food preparation and boxing in the food manufacturing industry. An important point of food grasping by robots is that the shape of the food should not be damaged. However, it is difficult for a conventional robot hand to perform this task perfectly. Therefore, an end-effector equipped with a camera-based tactile sensor has been developed to perform this task in previous studies. However, the performance of this end-effector depends on the reflectance of the target object, since it estimates contact based on the reflectance of light. We have developed a camera-based tactile sensor and contact estimation system to solve this problem. In addition, we have developed a pickup motion combined with object detection.



OS11-7 Development of a drone obstacle avoidance system based on depth estimation

Sora Takahashi, Eiji Hayashi (Kyushu Institute of Technology, Japan)

This study developed an obstacle avoidance system for drones using depth estimation from RGB cameras, aiming to reduce reliance on expensive sensors like RGB-D cameras or LiDAR. The system employs the deep learning model ZoeDepth for depth estimation and integrates it with ROS and Gazebo for simulation. Two autonomous systems were evaluated: one using RGB-D cameras and the other using depth estimation with RGB cameras. Experimental results show that while the RGB-D camera system outperformed in accuracy, the depth estimation-based system provided cost-effective and reasonable performance, especially in complex environments. The research concludes with plans to improve the system for denser obstacle environments and conduct real-world experiments.

OS12 Advances in Field Robotics and Their Applications (9)

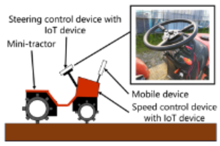
Chair Shinsuke Yasukawa (Kyushu Institute of Technology, Japan)

Co-Chair Kazuo Ishii (Kyushu Institute of Technology, Japan)

OS12-1 Practical Exercise on An Autonomous Driving System Using Mobile Devices and IoT Devices for An Agricultural Tractor

Daigo Katayama, Yuto Nakazuru, Hikaru Sato, Shoun Masuda, Yuya Nishida,
Shinsuke Yasukawa, Kazuo Ishii (Kyushu Institute of Technology, Japan)

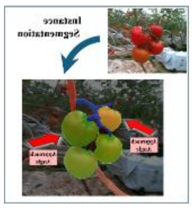
The current method of agriculture is expected to make sustainable production difficult due to the effects of a declining and aging workforce. To solve these issues, research and development of smart agriculture technologies, including automated tractor operation, have been underway. We have developed an automated driving system for a commercially mini-tractor using mobile and IoT devices to more facilitate the introduction of automated driving technology for tractors. In addition, the exercise was conducted for students to implement and operate this system with an aim of education for robotics engineers. This exercise consists of lectures and development exercises for the system. This paper describes the developed autonomous driving system, the overview of the exercise, and the results of the exercise.



OS12-2 Estimation of Image-Based End-Effector Approach Angles for Tomato Harvesting Robots

Kizuna Yoshinaga, Hikaru Sato, Kazuo Ishii, Shinsuke Yasukawa
(Kyushu Institute of Technology, Japan)

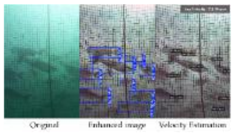
We propose a method to estimate a suitable approach angle for the end-effector of a tomato harvesting robot based on image data. Agricultural harvesting robots often face obstacles such as other fruits or stems around the target crop. Additionally, it is important to approach the target from a direction appropriate for harvesting, considering the shape of the end-effector. The proposed method uses a deep learning-based instance segmentation model to extract regions of fruits and stems, and estimates the suitable approach angle based on their positional relationships. We demonstrated the usefulness of the proposed method using an image dataset acquired in an actual tomato greenhouse.



OS12-3 Visual-Based System for Fish Detection and Velocity Estimation in Marine Aquaculture

Raji Alahmad, Dominic Solpico, Shoun Masuda, Takahito Ishizuzuka, Kenta Naramura, Zhangchi Dong, Zongru Li, Kazuo Ishii (Kyushu Institute of Technology, Japan)

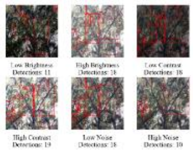
As global aquaculture continues to expand to meet the rising seafood demand, optimization of feeding remains a crucial issue for the industry to address to achieve sustainable development. This study proposed a visual-based system for estimating fish velocity, which is to be integrated into a farmer's feeding operation to determine the optimal feed amount. The YOLOv8 algorithm was utilized to detect fish in underwater videos, enabling precise monitoring of fish behavior. The results indicate a successful fish detection with an accuracy of 85%. The fish velocity estimation approach demonstrated the difference between the hungry fish and the normal fish behavior.



OS12-4 Evaluating of Tree Branch Recognition Algorithm in Pruning Robots under Augmented Environmental Conditions

Mohammad Albaroudi, Raji Alahmad, Abdullah Alraee, Kazuo Ishii
(Kyushu Institute of Technology, J Japan)

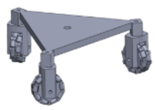
The integration of service robots has transformed various sectors by improving accuracy, efficiency, and scalability. In automating tasks like tree pruning, where precise branch detection is essential, this research examines YOLOv8's ability to recognize branches as a step toward full automation. To tackle challenges in diverse real-world conditions, video sequences are augmented with simulated variations in lighting, saturation, and noise. Metrics such as precision, true detections, and false detections reveal YOLOv8's robust performance in branch perception. These findings underscore its potential to enhance pruning systems, enabling efficient and scalable robotic solutions for tree maintenance and similar tasks.



OS12-5 Trajectory Analysis for a Mobile Robot Adapted Three Omni Rollers in Constant Roller's Speed

[1]Kenji Kimura, [1]Kazuki Nakayama, [2]Katsuaki Suzuki, [3]Kazuo Ishii,
([1]National Institute of Technology, Matsue College, [2]Kumamoto Industrial Research Institute,
[3]Kyushu Institute of Technology, Japan)

In recent years, mobile robots have been developed for the logistics industry. It is an omni-directional mobile mechanism with multiple omni rollers, and its kinematics have been proposed. In this study, a theoretical equation for the trajectory of the robot is derived when the roller speed is constant. In order to reduce the cost and time of validation experiments, the robot is validated in a simulation environment as a preliminary step.



OS12-6 Cross-Disciplinary Learning Through Manufacturing: Toward Student-Centered STEAM Education

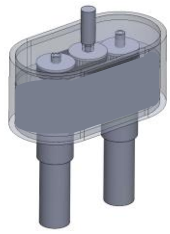
Kenji Kimura (National Institute of Technology, Matsue College, Japan)

In recent years, with the importance of cross-disciplinary educational programs such as STEAM education, it has become necessary to provide mathematical education at the early stages of elementary and junior high school to prevent the increase in the number of students who have dropped out of science. As a result, educational institutions are also becoming more active in efforts such as robot-themed education as part of their contribution to the local community. In this study, we propose a method for students to decide their own theme about mechanics, obtain a production budget, and engage in cross-disciplinary learning through extracurricular activities with support outside and inside the school.

OS12-7 Development of a Rotary Actuator Capable of Multidirectional Rapid Motion and Variable Stiffness

[1]Katsuaki Suzuki, [2]Yuya Nishuda, [3]Kenji Kjmura, [2] Kazuo Ishii
([1]Kumamoto Industrial Research Institute, [2]Kyushu Institute of Technology,
[3]National Institute of Technology, Matsue College, Japan)

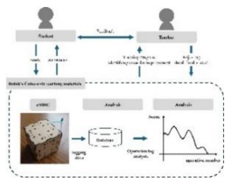
With the advancement of automation and digital transformation in the manufacturing industry, it is expected that industrial machines will be required to perform new tasks. Enhancing the multifunctionality of actuators is one approach to achieving these tasks. This paper proposes a new mechanism that combines two types of cams with different contour shapes, springs, two motors, and other mechanical components, and introduces an electric actuator incorporating this mechanism. The key feature of this actuator is its ability to achieve three functions (normal motion, rapid motion, and variable stiffness) while maintaining the same output characteristics, even when the initial posture of the output shaft is changed by switching the driving patterns of the two motors.



OS12-8 Study of Evaluation Operation Log Analysis Using 2^3- ERC on Matsue National College of Technology

Takumi Ueda, So Takei, Akira Nakano (National Institute of Technology, Kurume College, Japan),
Kenji Kimura (National Institute of Technology, Matsue College, Japan),
Kazutaka Matsuzaki (Nishinippon Institute of Technology, Japan)

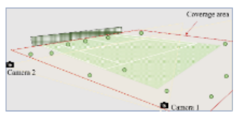
In response to the demand for educational proposals that address advancements in science and technology in Japanese school education, cross-disciplinary education and STEAM education are receiving increased attention. Given this context, a new unit called "Mathematics and Human Activities" was introduced. However, the need for innovative mathematics teaching materials is necessary. We developed the 2^3 Electric Rubik's Cube (2^3-ERC) with two LEDs for each cube edge for easy tractable operation logging data. By utilizing the operation log data as feedback, we evaluated log data analysis through the experiment on Matsue National Institute of Technology's students. The results from the operation log show the level of understanding of 2^3-ERC from the number of operations and consideration time.



OS12-9 Efficient Ball Position Estimation for Tennis Court Robot Assistants using Dual-Camera System

Abdullah Alraee, Raji Alahmad, Hussam Alraie, Mohammad Albaroudi, Kazuo Ishii
(Kyushu Institute of Technology, Japan)

During tennis training, professional players use many balls distributed randomly around the court. Collecting the balls manually is inefficient due to the effort and time required. A mobile robot for ball collection has been introduced to save energy and training time. The robot's tasks include ball detection, estimating positions, and finding the best path for efficient collection. In previous work, we addressed ball detection using a YOLOv8 neural network algorithm. This study focuses on the next step: ball position estimation using two cameras to cover the court. Results show successful position estimation on the x- and y-axes, with 94.48% accuracy.



Dual-Camera System for tennis ball detection

OS13 Research Towards the Renewable Energy and the Sustainable Development Goals (SDG's) / Part A (5)

Chair Ammar A.M. Al Talib (UCSI University, Malaysia)
Co-Chair Takao Ito (Hiroshima University, Japan)

OS13-1 Solar-Powered IoT-Based Smart Aquaponic System for Sustainable Agriculture

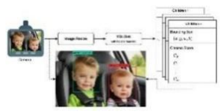
Alvi Khan Chowdhury[1], Sarah 'Atifah Saruchi[2], Ammar A.M. Al-Talib[3], Abdirisak Mubarik Muhumed[3,] Teh Boon Hong[3], Pavindran A/L Shanmugavel[3], Annanurov Kerim[3], Ng Weng Kent[3]
([1]Monash University Malaysia, Malaysia) ([2]UMPSA, Malaysia) ([3]UCSI University, Malysia),

This paper introduces a groundbreaking smart aquaponics system designed to address the limitations of conventional setups. The system leverages IoT technology, renewable energy, and automation to achieve real-time monitoring and environmental control. Key innovations include slidable grow beds for optimal sunlight exposure, solar tracking mechanisms for efficient energy utilization, and automated fish feeding using Real Time Clock (RTC) modules. The system demonstrated 90% water efficiency, significant energy savings, and streamlined resource management in prototype testing. With its modular and scalable design, this solution is ideal for urban farming and sustainable agriculture.

OS13-2 AI-Powered Detection of Forgotten Children in Vehicles Using YOLOv11 for Enhanced Safety

Nur Atikah Jefri[1], Sarah 'Atifah Saruchi[1], Radhiyah Abd Aziz[1], Aqil Hafizzan Nordin[1], Ammar A.M. Al-Talib[2], Zulhaidi Mohd Jawi[3],
([1]UMPSA, Malaysia) ([2]UCSI University, Malaysia) ([3]MIROS, Malaysia)

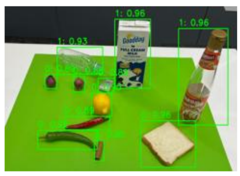
This study proposes a child presence detection system in vehicles, focusing on evaluating the performance of YOLOv11 for accurate detection and identification. To train the system, images simulating a child's presence in vehicles were collected using a doll, and these annotated images were labeled with the Computer Vision Annotation Tool (CVAT). The study emphasizes the potential of YOLOv11 as an effective and reliable solution for unattended child detection in vehicles. By leveraging advanced deep learning techniques, this research highlights the importance of addressing critical safety issues.



OS13-3 Exploring the Performance of YOLOv11: Detecting Compostable and Non-Compostable Kitchen Waste in Real-Time Applications

Ain Atiqa Mustapha[1], Sarah 'Atifah Saruchi[1], Mahmud Iwan Solihin[2], Fatima Karam Aldeen[2], Ammar A.M. Al-Talib[2] ([1]UMPSA, Malaysia) ([2]UCSI University, Malaysia)

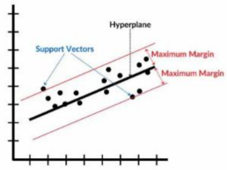
This paper investigates the advancements of YOLOv11, the latest model in the YOLO series in real-time object detection tasks on small datasets of compostable and non-compostable kitchen waste. Using a custom compostable and non-compostable kitchen waste dataset, YOLOv11 achieves an accuracy of 90.7% and a mean Average Precision (mAP) of 0.91, with a reduced inference time of 10.5 milliseconds. The study highlights YOLOv11's architectural enhancements, training methodology, and potential applications in waste management. While YOLOv11 sets a new benchmark in object detection, challenges like high computational demands, paving the way for future research on optimization for edge devices



OS13-4 Comparative Analysis of Machine Learning Algorithms for Rainfall Prediction in Kuantan, Pahang, Malaysia

Seri Liyana Ezamzuri[1], Sarah ‘Atifah Saruchi[1], Ammar A.M. Al-Talib[2]
([1]UMPSA, Malaysia), ([2]UCSI University, Malaysia)

This study compares the performance and accuracy of four ML algorithms which are Support Vector Regressor (SVR), Artificial Neural Network (ANN), Random Forest Regressor (RFR), and Linear Regression (LR) in the rainfall prediction application. All four methods employ the same input parameters which are temperature (°c), dew point (°c), humidity (%), wind speed (Kph) and pressure (Hg). Meanwhile the output parameter is set to be the rainfall (mm) which indicates the precipitation in Kuantan, Pahang, Malaysia. The analysis shows that the SVR consistently outperforms the other machine learning algorithms, achieving the lowest Mean Absolute Error (MAE) and Mean Squared Error (MSE).



OS13-5 Autonomous Vehicle Navigation in Highway with Deep Q-Network (DQN) using Reinforcement Learning Approach

Sumiya Tamanna Fujita[1], Sarah ‘Atifah Binti Saruchi[1], Ammar A.M. Al-Talib[2], Nurbaiti Wahid[3], Siti Nurhafizza Maidin[3], Alvi Khan Chowdhury[4], ([1]UMPSA, Malaysia) ([2]UCSI University, Malaysia) ([3]UiTM Dungun, Malaysia) ([4]Monash Universiti Malaysia, Malaysia)

This study addresses the collision avoidance problem in autonomous vehicles under dynamic and unpredictable environments, such as rain and dust storms. A Proximal Policy Optimization (PPO)-based reinforcement learning approach is proposed to develop an autonomous driving agent capable of navigating safely in adverse conditions. The agent is trained using the CARLA simulator, specifically in the Town04 environment. The methodology involves dynamically adjusting the vehicle's steering angle and speed based on environmental feedback to prevent collisions. Experimental results demonstrate the agent's ability to learn effective driving strategies, showcasing the potential of reinforcement learning for enhancing the reliability of autonomous vehicles in challenging weather scenarios.

OS14 Research Towards Renewable Energy and the Sustainable Development Goals (SDG’s)/ Part B (4)

Chair Firas Basim Ismail (University Tenaga National (UNITEN), Malaysia)
Co-Chair Takao Ito (Hiroshima University, Japan)

OS14-1 Empowering Decentralized Microgrids with A Blockchain-Based Peer-To-Peer Energy Trading Platform

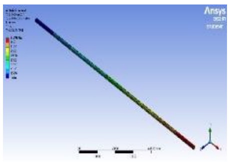
Firas Basim Ismail[1,] Chetenraj Singh[1], Ammar A. Al-Talib[2], Nizar F.O. Al-Muhsen[3]
([1] UNITEN, Malaysia), ([2] UCSI University, Malaysia), ([3]Middle Technical University, Iraq)

This study explores the concept of a Peer-to-Peer Energy Trading Platform for Decentralized Microgrids through Blockchain Technology. The presented work focuses on the creation of a decentralized peer-to-peer energy trading platform using blockchain technology, which is designed for microgrid ecosystems. The study delves into blockchain’s ability to build trust and ensure the integrity of energy transactions. The study concludes with recommendations for future enhancements, including expanding platform capabilities, integrating diverse energy sources, and navigating regulatory challenges for widespread adoption and impactful change.

OS14-2 Performance of Kenaf Fibre Reinforced Epoxy Biocomposite for High Voltage Insulator Applications

Kang Rui Tan[1], Cik Suhana Bt. Hasan[1], Nor Fazilah Abdullah[1], Farah Adilah Jamaludin[1], Meng Choung Chiong[1], Eryana Hussin[1] ([1]UCSI University, Malaysia)

The aim of this research is to evaluate the performance of kenaf fibre-reinforced epoxy composite as the core of high voltage insulator subjected to identified wind load conditions by using finite element analysis which is then further validated using the theory of mechanics of materials. The performance of kenaf FRP is almost comparable to the conventional material from glass FRP and the results indicate that the kenaf FRP is able to sustain different types of loads when it is under operating.



OS14-3 Detection of Bullet Holes for Target Board in Malaysia Military (ATM) Shooting Exam Application

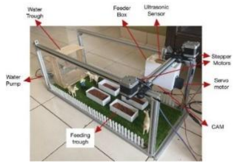
Jilian. H. Wai Yin, Idayu M. Tahir, Ammar A.M. Al Talib, Osama Mohamed Magzoub
(UCSI University, Malaysia)

This study focuses on designing and developing a bullet hole detection system for target boards in the Malaysia Army (ATM) shooting exercise environment. The deep learning algorithm used is based on YOLO models, utilizing Raspberry Pi and IoT via Blynk for remote monitoring.. The prototype includes a Raspberry Pi 4b, HQ Camera Module Lens, 35mm Telephoto Lens, and tripod stand, all at an affordable cost. The study demonstrates that the bullet hole detection system is accurate and effective for ATM shooting exams, meeting SDG 3, SDG 9, SDG 11, and SDG 12 goals.

OS14-4 Mobile App Development for Monitoring Goat Activities

Samy M. Elmasri, Idayu M. Tahir, Ammar A.M. Al Talib
(UCSI University, Malaysia)

"Mobile App Development for Monitoring Goat Activities" aims to create an automated and efficient system for managing goat feeding and water consumption. The prototype, constructed using a Raspberry Pi 4 equipped with a camera module and ultrasonic sensors, collects real-time data on the status of feeding troughs, detecting whether they are empty, partially empty, or full. By demonstrating the potential of integrating AI and IoT technologies in agriculture, this project highlights the benefits of continuous monitoring and timely alerts in maintaining a well-managed and sustainable farming operation.



OS15 Robotic Manipulation (5)

Chair Kensuke Harada (Osaka University, Japan)
Co-Chair Tokuo Tsuji (Kanazawa University, Japan)
Co-Chair Akira Nakamura (Saitama Institute of Technology, Japan)

OS15-1 Surface Stiffness Estimation using Active Strobe Imager

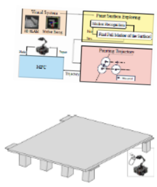
Taiki Yamaguchi[*1], Kensuke Harada[*1], Koji Mizoue[*2], Makoto Kaneko[*12]
([*1] Osaka University, Japan, [*2] Mizoue Project Japan Corp., Japan)

In this paper, we propose a method for estimating mechanical impedance of surface like skin using an Active Strobe Imager (ASI). ASI has the capability to non-contactly excite the target surface using an air jet flow, and to qualitatively observe traveling waves through strobe illumination. On the other hand, this paper shows that the surface impedance parameter can be estimated without contacting the target surface along with the visualization through ASI. We solve two problems: one is the inverse problem and the other is the forward problem. In the inverse problem, we estimate the spring constant of the target using information obtained from the measurement data. In the forward problem, we determine the surface displacement from the applied force, we compared the obtained displacement from the measurement data, demonstrating that the spring constants were correctly estimated.

OS15-2 Painting Task Planning for Large Structure using a Mobile Manipulator

Hiroshi Tanaka[*1], Masato Tsuru[*1], Takuya Kiyokawa[*1], Kensuke Harada[*1] ([*1] Osaka University, Japan)

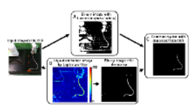
Painting a large structure with a robotic arm that is fixed to the ground is difficult due to its limited reachable range. To plan the robotic painting motion of such a large structure, we develop a ROS-based robotic software system assuming a mobile manipulator to explore the environment using SLAM. Our software system includes both detection of AR markers and construction of the envitonmental map to determine the painting location. It can measure the error in self-position estimation that occurs during the movement. It can also generate spray trajectories for the recognized painting location and control the whole body using Model Predictive Control (MPC) to perform painting over a wide area.



OS15-3 Real-time Cable Tracking by Wire Segmentation and Coherent Point Drift

Ryunosuke Yamada[1], Tokuo Tsuji[1], Takahiro Shimizu[2], Shota Ishikawa[1 2], Tomoaki Ozaki[2],
Yusuke Sakamoto[1], Tatsuhiro Hiramitsu[1], Hiroaki Seki[1]
([1]Kanazawa University, [2]DENSO CORPORATION, Japan)

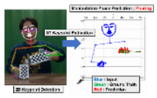
In this paper, a real-time cable tracking system by fast segmentation method and Coherent Point Drift (CPD) is proposed. Fast cable segmentation based on color space is inaccurate because of background contrast. Therefore, this technique uses edge information from the image to address this problem. The method consists of three processes: threshold processing in the Luv color space, edge processing using a Laplacian filter, and processing for extracting the common region of the binary images generated by each process. In the experiments, the accuracy of the segmentation region and the processing time required for each process of the tracking system are shown.



OS15-4 Motion Prediction for Human-Robot Collaborative Tasks Using LSTM

Kaihei Okada, Tokuo Tsuji, Tatsuhiro Hiramitsu, Hiroaki Seki,
Toshihiro Nishimura, Yosuke Suzuki and Tetsuyou Watanabe (Kanazawa University, Japan)

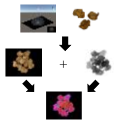
This study proposes an assistive robot system to reduce caregiving burdens in an aging society by supporting impaired body movements. The system focuses on bimanual tasks, such as pouring a drink from a bottle into a cup. Using 3D skeletal data excluding the impaired left hand, a deep learning model (LSTM) predicts the motion stages and 3D positions of the left hand, and the robot performs the substitute motions. The system uses data from multiple users to show its potential for improving patient independence and reducing caregiver workload.



OS15-5 Individual Recognition of Food in Bulk by using 3D Model of Food

Yuya Otsu, Tokuo Tsuji, Tatsuhiro Hiramitsu, Hiroaki Seki (Kanazawa University, Japan)

In this paper, we propose a method of individual recognition of food in bulk by using 3D model of food. First, color images and depth images of them are generated by using 3D model of food and physics engine of simulator. Then, color and depth composite images are created by converting two channels from color images and one channel from depth images. In the experiments, the accuracy of individual recognition of food in bulk with color and depth composite images are shown to compare the accuracy with only color images.



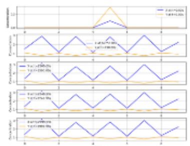
OS16 Natural Computing (3)

Chair Marion Oswald (TU Vienna, Austria)
Co-Chair Yasuhiro Suzuki (Nagoya University, Japan**)**

OS16-1 Modeling Yawning Contagion as a Reaction-Diffusion System: Emergence of Turing Patterns in Behavioral Contagion

Yasuhiro Suzuki (Nagoya University, Japan)

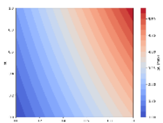
When we see someone yawning, we often feel compelled to yawn ourselves - a phenomenon known as behavioral contagion in psychology. While one person's yawn acts as an activator that triggers yawns in others, we sometimes suppress the urge to yawn in situations like meetings, representing an inhibitor of this behavior. We formulated this yawning contagion as a reaction-diffusion phenomenon in an activator-inhibitor system and confirmed the emergence of Turing patterns. Our findings provide a theoretical framework for understanding and potentially controlling the spread of social behaviors in human populations.



OS16-2 Dominant Region Analysis: A Novel Framework for Quantifying Competitive Reactions Based on the Gillespie Algorithm

Yasuhiro Suzuki (Nagoya University, Japan)

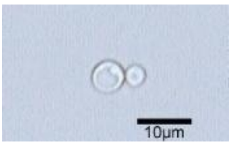
Understanding the quantitative relationships between competing reactions is crucial for analyzing chemical reaction systems. While conventional approaches often focus on static analysis, we propose a novel concept called "dominant region" to capture the dynamic nature of reaction competition. The dominant region concept can be viewed as an extension of the traditional rate-determining step in reaction kinetics. This enables quantitative prediction of how dominant reactions dynamically change with variations in reactant concentrations.



OS16-3 40 Hz sound exposure alters dissolved oxygen levels, gene expression, and colony formation in Saccharomyces cerevisiae BY4741

Yasuhiro Suzuki (Nagoya University, Japan)

We found that 40 Hz exposure significantly increased dissolved oxygen levels in yeast culture medium, but not in purified water. RNA-seq and DNA microarray analyses revealed that 40 Hz exposure significantly altered the expression of genes involved in cell adhesion, cell wall organization, and stress response. Notably, the expression of FLO11 and several PAU genes, which are important for yeast biofilm formation, was upregulated by 40 Hz exposure. Our results suggest that 40 Hz sound exposure can enhance dissolved oxygen levels and biofilm formation in S. cerevisiae, potentially through the upregulation of adhesion-related genes.



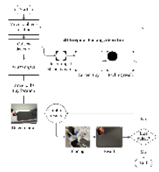
OS17 Artificial Intelligence for Embedded Systems and Robotics (10)

Chair Hakaru Tamukoh (Kyushu Institute of Technology, Japan)
Co-Chair Yuma Yoshimoto (National Institute of Technology, Kitakyushu College, Japan),
Co-Chair Dinda Pramanta (Kyushu Institute of Information Sciences, Japan)

OS17-1 Optimizing Object Placement for Human Support Robots Using a Two-dimensional Irregular Packing Algorithm for Efficient Tray Storage.

Natee Buttawong, Kosei Isomoto, Kosei Yamao, Ninnart Fuengfusin, Hakaru Tamukoh
(Kyushu Institute of Technology, Japan)

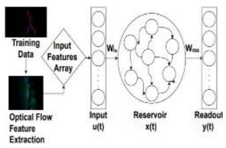
Human support robots (HSR) are robots that assist humans in their daily tasks. Their main application is tidying up, which involves detecting objects, determining appropriate placement locations, and organizing them. This study focuses on tidying up tray storage. Determining suitable storage positions is essential for storing objects in the tray. If the robot store objects in fixed predetermined locations, it can lead to inefficient use of storage space, and in the worst case, objects might collide and overflow from the tray. To address this limitation, we propose a 2-dimensional irregular packing algorithm utilizing an object mask method to calculate the best placement location. This study evaluates the proposed packing algorithm against the standard method to determine which approach is more effective in HSR applications.



OS17-2 Classification of Human Activity by Event-based Vision Sensors using Echo State Networks

Rohan Saini, Aryan Rakheja, Ryuta Toyoda, Yuichiro Tanaka, Hakaru Tamukoh
(Kyushu Institute of Technology, Japan)

We propose a system for human activity recognition using an event-based vision sensor (EVS) with echo state networks (ESNs). Conventional cameras are susceptible to motion blur and require computationally intensive methods, whereas EVS provides no motion blur and low latency. Our research aims to enable accurate recognition of human activities by using energy-efficient methods. Therefore, we adopt ESNs, which require low computational costs, for the classifier. Additionally, we use feature extraction algorithms such as optical flow and histogram of gradients to improve accuracy. We used an EVS activity recognition dataset created by us containing six human activities and a total of 600 videos. The results showed that our hybrid approach outperformed several techniques. We achieved 89% accuracy when trained with ridge regression.



OS17-3 Integrating Advanced Speech Recognition and Human Attribute Detection for Enhanced Receptionist Task in RoboCup@Home

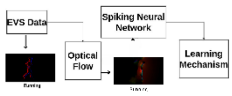
Koshun Arimura, Yuga Yano, Takuya Kawabata, Hakaru Tamukoh
(Kyushu Institute of Technology, Japan)

RoboCup@Home is held to integrate service robots into society. It includes a task called "Receptionist" that evaluates Human-Robot Interaction. In this task, a robot must ask guests for their names and favorite drinks and guide them to available seats. Additionally, the robot must introduce the guest's features such as their clothing to others. We developed a system integrating speech recognition and human attribute detection to achieve these functions. The robot can determine which seat a person is sitting in by detecting the person's skeletal coordinates. Additionally, the robot can identify individuals by recognizing human attributes. To verify the effectiveness of the developed system, we participated in the Receptionist task at RoboCup@Home 2024. We won first place in our league and demonstrated the effectiveness of our system.

OS17-4 Classification of Human Activity by Spiking Neural Networks using Event-based Vision Sensors

Aryan Rakheja, Rohan Saini, Ryuta Toyoda, Yuichiro Tanaka, Hakaru Tamukoh
(Kyushu Institute of Technology, Japan)

We propose a human action classification system that integrates spiking neural networks (SNNs) with event-based vision sensors (EVS) to address the limitations of conventional camera methods. Conventional approaches require significant computational resources and suffer from motion blur and limited dynamic range. EVS provides asynchronous data, enabling efficient and low-latency motion analysis with high temporal resolution and a wide dynamic range. SNNs further enhance this by processing data in an event-driven manner, reducing energy consumption and improving scalability. We created a dataset of 600 clips using EVS, with optical flow for feature extraction, achieving 93% classification accuracy. This approach offers an efficient solution for real-time action recognition in dynamic environments.



OS17-5 Robotic Grasping of Common Objects: Focusing on Edge Detection for Improved Handling

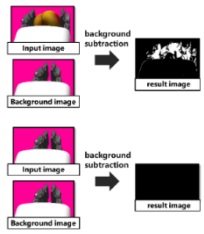
Tomoya Shiba, Hakaru Tamukoh (Kyushu Institute of Technology, Japan)

Grasping objects like plates and cups poses unique challenges for robots because of their irregular shapes and the difficulty of finding reliable grasp points. Traditional approaches often attempt to grasp the object at its center, but this strategy tends to fail for items like plates or cups, whose shapes deviate from simple forms like cubes or spheres. To address this issue, we propose a new method that utilizes AI-powered image analysis to identify the best edges for grasping. Through experiments conducted with a home service robot and a set of YCB objects, we evaluated the effectiveness of our approach compared to conventional methods. The results revealed a significant improvement in the success rate, particularly for objects with prominent edges, such as cups.

OS17-6 Proposal of a Grasp Verification Method Utilizing Background Subtraction and Depth Information

Ryo Terashima, Yuga Yano, Koshun Arimura, Hakaru Tamukoh (Kyushu Institute of Technology, Japan)

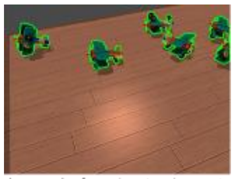
Commonly grasp verification approach involves using the opening width of the robot's gripper. However, methods based on the opening width of the gripper may not apply to slender objects. In this study proposes a grasp verification method using background subtraction. Our proposed method uses depth information to mask the background, isolating only the images of the gripper and the grasped object. Subsequently, a difference image is created by comparing the current image with the pre-grasp state, and the grasp state is detected based on the magnitude of the observed changes. The method minimizes environmental influences by masking the background, enabling highly accurate grasp verification even for complex objects. Through experiments, we validate the effectiveness of the proposed method.



OS17-7 Grasp Point Estimation Using Object Recognition Models with Simulator-Generated Datasets Including Pose Information

Ryoga Maruno, Tomoya Shiba, Naoki Yamaguchi, Hakaru Tamukoh (Kyushu Institute of Technology, Japan)

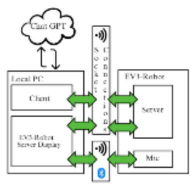
We have developed a system that automatically generates training datasets for object recognition models using a simulator. In this study, we have successfully incorporated pose information into the dataset. The figure shows part of the dataset generated by the simulator. We used this information to develop a system for estimating grasp points for objects that are difficult to grasp by robots. We chose a toy airplane as the target object. As shown in the figure, three specific points were assigned to the object: the front, center, and back. In the grasp point estimation process, the center point was designated as the grasp point. The appropriate grasp was achieved by moving the robot's arm perpendicularly to the line connecting the front and back points. This system calculates both the coordinates of the grasp point and the required arm angle.


An example of experiment environments
RED point : visible, BLUE point : not visible

OS17-8 A feasibility study of generative AI applications using EV-3 Robots at the Kyushu Institute of Information Sciences

Dinda Pramanta[1], Hakaru Tamukoh[2]
([1]Kyushu Institute of Information Sciences, [2]Kyushu Institute of Technology, Japan)

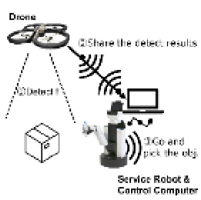
The advent of generative artificial intelligence (Gen AI) using large language models (LLMs) has brought about a transformation in the fields of education and home robotics. This study examines how students at the Kyushu Institute of Information Sciences perceive and utilize ChatGPT, with a particular on their familiarity, ethical considerations, and trust (FET) concerns. To further investigate this, we developed an educational EV-3 robot powered by ChatGPT and simulate the voice command using socket connections. By combining surveys with hands-on experiments, we uncovered the strengths and limitations of both the educational and home robotics roles. Our findings highlight the importance of FET, and 88% of respondents have no issues with implementing such technologies in the future.



OS17-9 Development of a Collaborative System Between A Drone and A Home Service Robot for Enhanced Operational Efficiency

Haruki Miura, Rion Yohu, Yuma Yoshimoto (National Institute of Technology, Kitakyushu College, Japan)

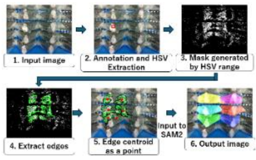
When considering the operation of home service robots, there are problems such as the impossibility to obtain information on objects hidden behind obstacles or in distant areas of a large room. There is also the potential issue of not being able to detect people behind the robot, leading to collisions. Therefore, this research proposes a system that coordinates home service robots with drones to improve task efficiency. As an experiment, we conduct a search and pick-up task that integrates the home service robot and the drone. The drone's bottom camera and YOLOv8 are used to detect objects and send the information to the robot. The robot then moves to the room where the object is located and grasps it. The time required to complete the task is evaluated.



OS17-10 Efficient Object Detection with Color-Based Point Prompts for Densely Packed Scenarios in WRS FCSC 2024

Naoki Yamaguchi[1], Tomoya Shiba[1], Hakaru Tamukoh[1]
([1]Kyushu Institute of Technology, Japan)

We propose the use of color-based point prompts for efficient object detection in densely packed scenarios, specifically targeting the World Robot Summit (WRS) Future Convenience Store Contest (FCSC) 2024. Our system leverages color information to generate point prompts, which are processed by Segment Anything Model 2 (SAM 2) to create object masks and estimate grasping points. SAM 2 is a foundation model for image segmentation. It takes an image and prompts about object positions as input, then outputs the segments of the specified objects. We applied the WRS FCSC 2024 Stock Sub Task to evaluate the system. Despite achieving 7th place in the competition, the effectiveness of our recognition system was confirmed.



OS18 Robotics in Biophilic-Designed Space Toward Co-existence of Humans, Robots, and Plants (4)

Chair Yuichiro Tanaka (Kyushu Institute of Technology, Japan)
Co-Chair Naoto Ishizuka (Kyushu Institute of Technology, Japan)
Co-Chair Tomomi Sudo (Kyushu Institute of Technology, Japan)
Co-Chair Hakaru Tamukoh (Kyushu Institute of Technology, Japan)

OS18-1 Application of AI Robot Technology for Biophilic Design

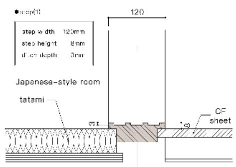
Kairi Manabe, Ryo Miyazono, Keitaro Ito, Tomomi Sudo, Naoto Ishizuka, Akinobu Mizutani, Yuki Anamizu, Etsushi Ueda, Honoka Tamai, Saya Nakano, Leon Furuya, Hakaru Tamukoh, Yuichiro Tanaka, Hirofumi Tanaka (Kyushu Institute of Technology, Japan)

To enhance human health and well-being, Biophilic design has been increasingly recognized in recent years. This design is characterized by the integration of natural elements such as plants, nature light, and water into spaces. However, challenges are posed by the maintenance of live plants, as their decline can be caused by insufficient environmental conditions. The aim of this study is to propose a robotic system of autonomously relocate houseplants to environments optimized based on sensor data, including light, temperature, and humidity. Through the integration of AI robotics with ecological design principles, it is aimed to enhance sustainability and redefine the relationship between humans, nature, and technology, fostering a harmonious interaction among "robots, nature, and humans."

OS18-2 Basic Research on the Development of Space Standards for the Use of Service Robots in Housing Using the Urban Renaissance Agency's Housing Complex

Ren Matsuoka, Kanon Nonoshita, Naoto Ishizuka, Ryohei Kobayashi, Akinobu Mizutani, Hakaru Tamukoh, Hirofumi Tanaka (Kyushu Institute of Technology, Japan)

This study conducts basic research on developing space standards for integrating home service robots in residences, using a room in an apartment complex from the Urban Renaissance Agency as a model. The study compares these spaces with existing robot mobility standards to identify issues. The robot-friendly level of the room was evaluated using the RFA standard. Results showed that most rooms fell into Level C, the lowest of three levels, in areas such as "fixture width" and "steps." Furthermore, when operating two types of home service robots, the study uncovered unique robot behavior issues in the housing that were not anticipated by the RFA standard.



OS18-3 An Exhibition Environment with 2D Markers for Guide Robot

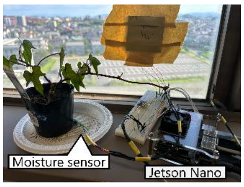
Akinobu Mizutani, Yui Hattori, Naoto Ishizuka, Yuichiro Tanaka, Hirofumi Tanaka, Hakaru Tamukoh
(Kyushu Institute of Technology, Japan)

The guide robot in an exhibition environment is expected to entertain visitors and reduce the maintenance cost of updating the robot database according to the changes in exhibition contents. The exhibition comprises printed panels for humans and a 2D marker printed with ultraviolet ink for robots. 2D markers are attached to the bottom of the exhibition furniture to make the markers invisible to visitors. The exhibition point may change depending on the contents of the exhibition. By searching for 2D markers autonomously in an exhibition space, the robot can update its internal database. This enables us to keep both the printed panel and the robot database without changing the robot database by users. The experiment is conducted in an exhibition environment, and the success rate in finding the exhibition point is evaluated.

OS18-4 Prediction of Timing and Amount of Houseplants Watering by an Echo State Network on Jetson

Wataru Yoshimura, Koshun Arimura, Ryohei Kobayashi, Akinobu Mizutani, Tomoaki Fujino, Yuichiro Tanaka, Tomomi Sudo, Naoto Ishizuka, Keitaro Ito, Hirofumi Tanaka, Hakaru Tamukoh
(Kyushu Institute of Technology, Japan)

Cultivating of houseplants in biophilic-designed spaces requires appropriate timing and amount of watering. However, determining them is challenging, as fluctuations in ambient temperature can influence these factors. We develop a system capable of predicting ambient temperature changes and determining the appropriate timing and amount of watering. The system acquires ambient data using sensors connected to a Jetson Nano and processes the data using a neural network for the prediction and determination. We adopt an echo state network, a lightweight neural network, enabling a power-efficient system capable of running on edge devices. Additionally, we implement a function to notify the user of the timing and amount of watering via LINE whenever the soil moisture content drops below a predefined threshold.



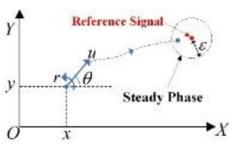
OS19 Intelligent Control (4)

Chair Yingmin Jia (Beihang University, P.R.China)
Co-Chair Weicun Zhang (University of Science and Technology Beijing, P.R.China)

OS19-1 Practical Linearization Control of Nonholonomic Unicycles

Lixia Yan, Yingmin Jia (Beihang University (BUAA), P.R.China)

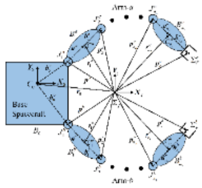
Due to underactuation, the states of nonholonomic systems cannot be steered toward arbitrary direction of the state space. This note takes unicycles as an example and demonstrates a new idea of control design for nonholonomic systems. More precisely, we apply state transformation technique and external dynamic oscillator, and convert an underactuated nonholonomic unicycle into a fully-actuated and linearizable one. A control law, capable of tracking and stabilization uses, is then constructed. The tradeoff therein is that the tracking/stabilization errors can only be steered into the neighborhood of the origin rather than converging to zero. Numerical simulations are carried out to validate the proposed control scheme.



OS19-2 Task-Space Tracking Control for Dual-arm Free-floating Space Manipulators with Disturbances and Uncertainties

Qian Sun, Yingmin Jia (Beihang University (BUAA), P.R.China)

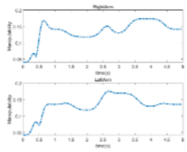
This paper investigates trajectory tracking control for dual-arm free-floating space manipulators (DFFSM) in task space subject to unknown disturbances, kinematic and dynamic uncertainties. First, we design an adaptive sliding mode disturbance observer to compensate for the unknow disturbances. Then, a backstepping tracking control algorithm is proposed, and two adaptive laws are developed to estimate the kinematic and dynamic uncertainties. It is validated through Lyapunov analysis that the tracking errors of the end-effectors are uniformly ultimately bounded with the proposed control scheme. Numerical simulations validate the effectiveness of the proposed control scheme.



OS19-3 Manipulability Optimization for Redundant Dual-Arm Robots at the Acceleration Level

Yang Zhang, Yingmin Jia (Beihang University (BUAA), P.R.China)

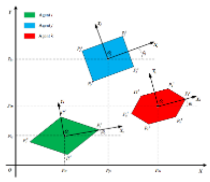
Existing manipulability optimization schemes typically solve at the velocity level, which cannot consider joint acceleration limits and are unsuitable for torque control of robotic arms. Therefore, this paper constructs a cost function that considers both joint torque constraints and manipulability optimization of the manipulator, and equivalently transforms it into a convex quadratic function. The proposed scheme addresses the non-convexity issue of manipulability with respect to the robotic arm joint acceleration and the inversion problem of the generalized Jacobian matrix. Simulation results show that the proposed method can maximize the manipulability of redundant dual-arm robots at the acceleration level, verifying the effectiveness of the scheme.



OS19-4 Flocking Control for Multiple Convex Polygonal Agents with Obstacle Avoidance

Yaxin Li, Yingmin Jia (Beihang University (BUAA), Beijing P.R.China)

This paper addresses the flocking control for second-order convex polygonal multiagent systems with obstacle avoidance. Typically, existing research reduces agent shapes to points or circles, which can lead to suboptimal use of spatial resources. To rectify this, the paper introduces an approach to compute the relative distance between agents. A potential function is designed based on these calculated distances. A flocking trajectory steers movement, while an obstacle avoidance path is triggered when an agent approaches an obstacle. The proposed control strategy integrates the potential function, reference trajectory, and obstacle avoidance trajectory to achieve flocking behavior and obstacle avoidance. Stability analysis proves the effectiveness of the algorithm.



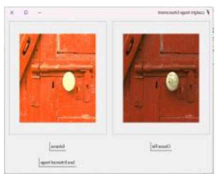
OS20 Applications (6)

Chair Kasthuri Subaramaniam (University of Malaya, Malaysia)
Co-Chair Abdul Samad Bin Shibghatullah (Universiti Tenaga Nasional, Malaysia)

OS20-1 Low-light Image Enhancement with Color Space (Cielab)

Lee Kok Xiong[1], Kasthuri Subaramaniam[2], Umm E Mariya Shah[1], Abdul Samad Bin Shibghatullah[3], Oras Baker[4] ([1]UCSI University, Malaysia, [2]University of Malaya, Malaysia
[3]Universiti Tenaga Nasional, Malaysia, [4]University of Ravensbourne, England)

In this project, we are implementing a color transformation from RGB to CIELAB to enhance low-light images. This transformation separates color information from brightness information, which improves contrast and overall quality. We are using a standard color conversion formula and combining it with other techniques, such as histogram equalization and neural networks, for better results. The project will have a user-friendly interface that allows users to upload and download images and compare the original and enhanced versions. The programming language used and the specific details of the implementation process are not mentioned.



OS20-2 Integrated AI Voice Assistant News Website for Enhancing User Experience – AI-ReadSmart

Mohammed Mohi Uddin[1], Ghassan Saleh Hussein Al-Dharhani[1], Keoy Kay Hooi[1], Chit Su Mon[2], Kasthuri Subaramaniam[3]
([1]UCSI University, Malaysia, [2]Heriot-Watt University Malaysia Campus, Malaysia
[3]University of Malaya, Malaysia)

The offline newspaper sector has been declining for years, and following the epidemic there were further decreases. To assist users with the complexity issue of the current online news sector, this study aims to develop a website that is beneficial to both users and the news sector. To ensure a seamless transition to online reading, research was conducted on the existing state of the online news sector. The survey used in this study allowed the researcher to understand how individuals feel about the state of the current online news sector as well as how they feel about voice integration on news websites.

OS20-3 Developing a Mobile Healthcare Application – MyHealth

Abdulrahman Salmo Alhamada[1], Ghassan Saleh Hussein Al-Dharhani[1], Kasthuri Subaramaniam[2], Raenu Kolandaisamy[1]
([1]UCSI University, Malaysia [2]University of Malaya, Malaysia)

MyHealth targets common mobile healthcare problems like medication forgetfulness and basic health knowledge gaps. This mobile application gives medication reminders, health lessons and appointment scheduling. It employs user centric design principles in combination with modern technology for user engagement & health management. System architecture, user interface design and development process are discussed with regard to healthcare accessibility and patient compliance impact.

OS20-4 Developing a Body Posture Detection for Fitness

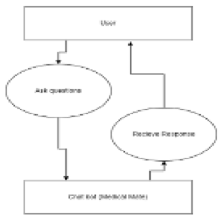
Kai Xuan Chong[1], Abdul Samad Bin Shibghatullah[2], Kasthuri Subaramaniam[3], Chit Su Mon[4]
([1]UCSI University, Malaysia, [2]Universiti Tenaga Nasional, Malaysia
[3]University of Malaya, Malaysia, [4]Heriot-Watt University Malaysia Campus, Malaysia)

The Body Posture Detection System for Fitness is an innovative technology that aims to enhance exercise technique and movement patterns by providing real-time monitoring and feedback. It utilizes computer vision and machine learning algorithms to track and analyze body movements during fitness. The system's ability to provide immediate feedback and correction significantly improves exercise effectiveness and user safety. These efforts are aimed at enhancing the functionality and usability of the Body Posture Detection System for Fitness while addressing user needs and optimizing fitness training experiences.

OS20-5 Medical Mate: Healthcare and Medical Chat Bot

Harris Hue Chee Kin[1], Javid Thirupattur[2], Kasthuri Subaramaniam[3], Shabana Anjum Shaik[4]
([1]UCSI University, [2]Sunway University, [3]University of Malaya, [4]Taylor's University, Malaysia)

The proposed idea is to develop a web-based medical chat bot called "Medical Mate" that will be placed by every patient's bedside, serving as a companion, and providing necessary care and support. It aims to offer convenience and accessibility for patients while easing the workload of hospital staff. Usability tests, employing mixed methods research, were conducted to ensure the chat bot's design and functionalities meet user requirements. Medical Mate pivoted to a web-based application using HTML, CSS, JS, jQuery, and MySQL. Ultimately, the goal of Medical Mate is to be a helpful and reliable companion for patients during their hospital stay, offering care and reducing the burden on medical personnel.



OS20-6 Crimes Identification System for Campus Safety and The Threat of Suspicious Student Conduct

Wong Zhen Bang[1], Kay Hooi Keoy[1], Kasthuri Subaramaniam[2], Sellappan Palaniappan[3], Oras Baker[4]
([1]UCSI University, Malaysia, [2]University of Malaya, Malaysia, [3]Help University, Malaysia
[4]University of Ravensbourne, England)

Ensuring campus safety is of paramount importance for educational institutions. With the increasing prevalence of crime and the potential threat of suspicious student conduct, there is a need for an effective crimes identification system. This paper aims to propose a comprehensive system that utilizes advanced technologies, such as video surveillance, data analytics, and behavioral monitoring, to detect and mitigate potential threats on campus. The proposed Crimes Identification System (CIS) integrates various components to enhance campus safety. These cameras capture real-time footage that is then processed using advanced video analytics algorithms.

OS21 Applications of Information Technology for Better Living (5)

Chair Tetsuro Katayama (University of Miyazaki, Japan)
Co-Chair Hiroki Tamura (University of Miyazaki, Japan)

OS21-1 Prototype of MixVRT Which Is a Visual Regression Testing Tool That Highlights Layout Defects in Web Pages

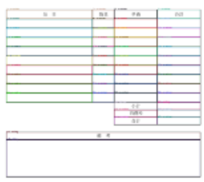
Naoki Aridome[1], Tetsuro Katayama[1], Yoshihiro Kita[2], Hisaaki Yamaba[1], Kentaro Aburada[1],
Naonobu Okazaki[1]([1]University of Miyazaki, Japan), ([2]University of Nagasaki, Japan)

As a method for detecting layout defects in Web pages, image-based visual regression testing is proposed. However, it has the problem that it takes time to detect unintended layout differences that are not based on HTML code. This study proposes a prototype of MixVRT which is a tool to detect layout defects in Web pages. MixVRT detects differences by comparing images and detects changes due to changes in HTML code on the web pages before and after changes. By comparing them, MixVRT can detect layout defects, which are unintended differences in layout. From evaluation experiments, the time required to detect layout defects can be reduced.

OS21-2 Proposal of a Method for Automatic Fill-in Fields Detection and for Labels Assignment to Generate Electronic Forms

Yuya Kimura[1], Tetsuro Katayama[1], Yoshihiro Kita[2], Hisaaki Yamaba[1], Kentaro Aburada[1], Naonobu Okazaki[1]
([1]University of Miyazaki, Japan), ([2]University of Nagasaki, Japan)

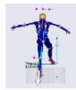
The digitalization of forms is being promoted because it is required saving as data format by amendment of Electronic Books Maintenance Act. One of the effective ways to manage contents filled in fields is using electronic forms. Several tools have been developed to generate them automatically. However, when you use a paper form, it takes time to generate electronic one because it is necessary to place fill-in fields on an electronic form by dragging them with a mouse. This paper proposes a method for automatic fill-in fields detection and labels assignment to reduce time required to place fill-in fields. From evaluation experiments, it has confirmed that the proposed method has reduced the time to generate an electronic form.



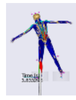
OS21-3 A Study on Methodology of Measurement for the Physical Burden on Preschool Children

Sachiko Kido, Hiroki Tamura (University of Miyazaki, Japan)

Measuring the burden placed on the body of a preschool child is difficult and has rarely been measured using motion capture. In this paper, the AnyBody Modelling System was used to verify whether it is possible to calculate the burden on the preschool child's body using motion capture. The lumbar burden value for preschool children was calculated, defining the burden on the lumbar region as the burden on the body as a whole. Few measurements have been made on young children with the AnyBody Modelling System. Therefore, the validity of the pre-school child figure was verified by comparing the lumbar burden with that of adult male. In addition, by setting the adult male data based on the preschool child's height and weight and comparing the calculated values with the preschool child's actual values, the possibility of simulating children's body burden in various movements using adult body models in the future was examined.



A preschool child's Balance with AnyBody



An adult male's Balance with AnyBody

OS21-4 Evaluation of Ankle Joint Movements in Frontal Plane for a Normal Coordinated Gait

Praveen Nuwantha Gunaratne, Hiroki Tamura (University of Miyazaki, Japan)

As per the records, around 15% of the global population is experiencing some form of disabilities in lower extremity resulting in loss of accessibility to their basic routine movements. The ankle joint complex plays an important role as a weight bearing articulation in the lower extremity and is a key contributor to the power behind human locomotion. While sagittal plane ankle movements are crucial for gait, several studies have proven that inversion-eversion, the front plane movements oversee the pressure distribution at the ankle joint to ensure a well-coordinated gait. This paper presents an evaluation of prediction of such ankle joint movements using Electromyogram (EMG), Inertial Measurement Unit (IMU) and Force-Sensitive Resistor (FSR) measurements, which can later be adapted for use in anthropometric active ankle orthosis designs to assist dynamic ankle movements during normal gait in real-time.

Sensor arrangement for data acquisition and validation

OS21-5 Development of a Real-Time Multi-Person 3D Keypoint Detection System Using Stereoscopic Cameras and RTMPose

Taufik Hidayat Soesilo, Praveen Nuwantha Gunaratne, Hiroki Tamura
(University of Miyazaki, Japan)

In this paper we consider a real time multi-person detection and analysis system using stereoscopic cameras and RTMPose, a novel high real-time pose estimation framework. RTMPose offers real time analysis of 2D key points for the individuals and this data is later augmented with depth data coming from stereoscopic imaging to give 3D skeletal data. The benefit of employing RTMPose is that the system is able to perform accurate and fast multiple persons tracking despite present occlusion scenarios. Consequently, the system overcomes the drawbacks of prior methods, including reliance on wearable devices and unsuitability for out-of-door environments, by employing stereoscopic cameras and RTMPose with low-latency and high-accurate inference. Experimental results demonstrate the system's ability to provide detailed real-time analysis of posture and movement for multiple individuals in diverse scenarios.

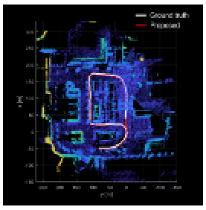
OS22 Navigation and Tracking (3)

Chair Chan Gook Park (Seoul National University, Republic of Korea)

OS22-1 Fine-registered Object LiDAR-inertial Odometry for a Solid-state LiDAR System

Hanyeol Lee and Chan Gook Park (Seoul National University, Republic of Korea)

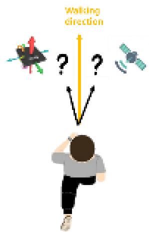
We propose the LiDAR-inertial odometry with object measurements for the solid-state LiDAR system. Although the geometric feature has been used for the precise localization with LiDAR, the measurement vanishing can lead to the localization failure in the limited field-of-view. To address this problem, we utilize objects that are sufficiently present in a man-made environment as localization measurements. The point clouds in the object are registered and the processed measurements are coupled with the geometric measurements in the estimator. The effectiveness of the object measurements is verified through a virtual environment simulator, and the proposed algorithm shows superior localization performance compared to the case of geometric measurement alone.



OS22-2 A Fusion Method for Estimating the Walking Direction of Smartwatch Users

Jae Hong Lee and Chan Gook Park (Seoul Nation University, Republic of Korea)

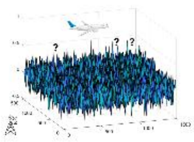
Accurately estimating the walking direction of smartwatch users is critical for applications such as exercise trajectory analysis. This study introduces a novel approach that fuses estimation direction from inertial sensors and GPS. Inertial sensors provide stable estimates as they are unaffected by environmental conditions, but their accuracy can be impacted by sensor performance and user motions, such as hand swinging. GPS, in contrast, offers higher accuracy than inertial sensors under favorable signal conditions. To leverage the strengths of both sensors, the proposed method employs an information-weighted consensus filter, integrating direction estimates and error covariances. Experimental results demonstrate that the fusion approach reduces estimation errors compared to individual sensors.



OS22-3 Multi-Frame Track-Before-Detect with Adaptive Number of Frame as Noise Level

Je Hwa Lee, Jae Hong Lee, and Chan Gook Park (Seoul National University, Republic of Korea)

Multi-frame Track-Before-Detect (MF-TBD) is a batch processing method used to enhance detection and tracking performance in low SNR environments. Unlike traditional filtering techniques, MF-TBD does not apply thresholding and instead uses all observed data to reduce the risk of target loss. By integrating observations across multiple frames, it leverages space-time correlations to improve detection robustness. However, as the number of frames increases, the computational cost grows exponentially due to the need to correlate data over a larger dataset, leading to inefficiencies. Especially in high SNR conditions, where fewer frames are sufficient for accurate detection. To address this, we propose an Adaptive MF-TBD framework that dynamically adjusts the number of frames based on SNR levels

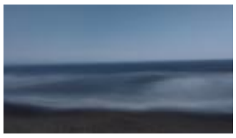


OS23 Mathematical Informatics (10)
Chair Takao Ito (Hiroshima University, Japan)
Co-Chair Amane Takei (University of Miyazaki, Japan)

OS23-1 Simplification of Rip Current Detection by Image Averaging Based on the Number of Wave Breaks

Ota Hamasuna[1], Leona Kimura[1], Satoshi Ikeda[1], Kaoru Ohe[1], Kenji Aoki[1],
Amane Takei[1], Akihiro Kudo[2], Makoto Sakamoto[1*]
([1]University of Miyazaki, Japan), ([2]National Institute of Technology, Tomakomai College, Japan)

According to a National Police Agency report, there were 1,392 water accidents in 2023, with 368 victims (dead or missing) in the sea, mainly due to rip currents. Detecting rip currents is crucial, and past studies have used image averaging, often relying on fixed-point cameras or lengthy videos, making it difficult for individuals to apply. This study proposes using smartphone videos, with durations adjusted by the number of wave breaks, to enable easier rip current detection. To test this, smartphone footage was recorded at Hitotsuba Surf Point in Miyazaki Prefecture for analysis.



OS23-2 Automated Classification of High-Grade Dried Shiitake Mushrooms Using Machine Learning

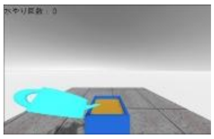
Leona Kimura[1], Ota Hamasuna[1], Kaoru Ohe[1], Satoshi Ikeda[1], Kenji Aoki[1],
Amane Takei[1], Akihiro Kudo[2], Kazuhide Sugimoto[3], Makoto Sakamoto[1*]
([1]University of Miyazaki, Japan), ([2]National Institute of Technology, Tomakomai Collage, Japan),
([3]SUGIMOTO Co., Ltd., Japan)

This study aims to automate high-grade dried shiitake mushrooms sorting using an anomaly detection system with Autoencoders (AE) trained on acceptable product data. Initial experiments using CNN approaches highlighted challenges in achieving high accuracy for acceptable product classification, necessitating improvement. The AE-based approach showed progress in detecting defective products via data cleansing, augmentation, and training optimization. However, misclassification of acceptable products with features like darker areas or complex textures remains an issue. This presentation outlines current findings and strategies, including data expansion and model improvements, to address these challenges.

OS23-3 Development of a Plant Growing Experience Application for Physically Challenged Children Using VR

Masatoshi Beppu[1], Masatomo Ide[1], Kaoru Ohe[1], Satoshi Ikeda[1], Kenji Aoki[1],
Amane Takei[1], Akihiro Kudo[2], Makoto Sakamoto[1*]
([1]University of Miyazaki, Japan), ([2] National Institute of Technology, Tomakomai Collage, Japan)

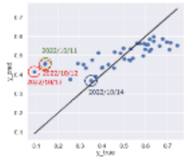
In 2016, the “first year of VR,” many VR platforms emerged, making VR technology more accessible. Currently, technology is expected to be applied and utilized in various fields. Application to the education sector is being promoted as part of the educational use of ICT. However, it is difficult to get the benefits of implementing VR due to lack of technology and equipment for teachers. Therefore, it is necessary to limit the scope of coverage. This study will focus on limb-challenged children and develop a VR application that allows them to experience plant growing. We believe that this will solve the problems that have been a concern for children with physical disabilities, such as the inability to perform exercises using soil and the lack of opportunities for trial-and-error. In this study, we also asked men and women in their teens to 40s to experience the apps we developed and obtained their evaluations through questionnaires. Within the survey, we received certain evaluations in areas such as trial and error. As for future issues, the application will be improved based on the feedback received from the survey. In addition, we believe it is necessary to evaluate the long-term effects of the application by having children with physical disabilities use it.



OS23-4 Exploring Social Media's Role in Predicting Stock Market Trends

Masatoshi Beppu[1], Masatomo Ide[1], Seita Nagashima[2], Satoshi Ikeda[1]*,
Amane Takei[1], Makoto Sakamoto[1], Tsutomu Ito[3], Takao Ito[4]
([1]University of Miyazaki, Japan), ([2]MEITEC CORPORATION, Japan),
([3]National Institute of Technology, Ube College, Japan), ([4]Hiroshima University, Japan)

This study analyzes tweets from the official Twitter accounts of NHK News and Nikkei to incorporate sentiment data into a predictive model for the Nikkei Stock Average. Adding sentiment data improved the R^2 score from 45.1% to a maximum of 70.5%, indicating the potential of SNS data in forecasting social indicators. However, no strong correlation between sentiment data and stock prices was observed. Challenges include the short data collection period and the difficulty of sentiment analysis in Japanese. Future work should focus on employing more effective methods for extracting sentiment.



OS23-5 Development of a Shrine Festival Support Application with Non-Technical Management Features: Functional Evaluation and Sustainability for Future Generations

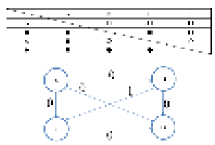
Masatomo Ide[1], Masatoshi Beppu[1], Satoshi Ikeda[1], Kaoru Ohe[1], Kenji Aoki[1],
Amane Takei[1], Akihiro Kudo[2], Makoto Sakamoto[1*]
([1]University of Miyazaki, Japan), ([2]National Institute of Technology, Tomakomai Collage, Japan)

This study focuses on developing and evaluating the usefulness of a web application that supports local shrine mythology and festivals, aiming to enhance their recognition and create lasting memories. The application incorporates features designed for both users and organizers to ensure accessibility and ease of maintenance across generations. Key functionalities include: (1) detailed event information, such as festival overviews, content highlights, and nearby facilities like parking and restrooms, enabling accurate and convenient information delivery; (2) an AR photo feature utilizing original character illustrations, allowing users to capture memorable photos; and (3) an admin-only feature for updating and editing essential information to maintain relevance over time. By combining technology with cultural heritage, this research demonstrates how digital tools can modernize traditional events and expand their appeal to broader audiences.

OS23-6 A Novel Approach to Reducing Ranking Discrepancies in Tennis Based on Tournament Choices

Masatomo Ide[1], Masatoshi Beppu[1], Kousei Yano[2],
Satoshi Ikeda[1]*, Amane Takei[1], Makoto Sakamoto[1], Tsutomu Ito[3], Takao Ito[4]
([1]University of Miyazaki, Japan), ([2]GENBASUPPORT Co., Japan),
([3]National Institute of Technology, Ube College, Japan), ([4]Hiroshima University, Japan)

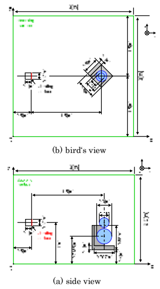
In tennis, discrepancies between rankings and head-to-head results arise because players select different tournaments for ranking. Existing methods using directed graphs cannot address discrepancies caused by varying tournament choices. This study proposes a ranking aggregation method that considers players' selected tournaments to reduce these inconsistencies. The method aggregates all chosen tournaments to form a collective ranking. Experimental results show a reduction in overall discrepancies, although some players saw an increase. This approach offers a partial solution to ranking inconsistencies caused by differing tournament selections in tennis.



OS23-7 Sound Field Evaluation on Acoustical Experiment with Several Loudspeaker Locations

Akihiro Kudo[1*], Shun Kubota[1], Amane Takei[2], Makoto Sakamoto[2]
([1]National Institute of Technology, Tomakomai College, Japan), ([2]University of Miyazaki, Japan)

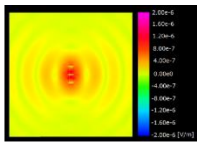
When a laboratory at a technical college is used as a site for psychoacoustic experiments, there are objects in the experimental environment such as walls, floors, ceilings, and chairs that obstruct the propagation of sound waves, which may change the accuracy of sound direction localization by altering the sound waves reaching the listener's ears. Therefore, the purpose of this study is to clarify the effects of these environments on sound localization characteristics using simulations. In this presentation, we will evaluate the effect of the reflection of sound waves by a chair set up to immobilize the subject on the subject's sound image localization for several loudspeaker positions. Fig. 1 takes one of several loudspeaker arrangements and shows the dimensions of its simulation model. The mesh size is set to 0.012m and the number of elements is 162,469,716. The acoustic impedance of the loudspeaker and a snowman as a subject and a chair is set to 4.56*109 kg/m2s to be completely reflect, the acoustic impedance of the room wall, ceil, floor is set to 445.9 kg/m2s to be completely absorbed. The sound field calculations are performed under non-stationary conditions. The time resolution is set to 1μs, and the duration time of simulation is set to 8000μs. The influence of the chair on the sound image localization is discussed by obtaining the time characteristics of the sound pressure at the subject's ear position from the simulation results.



OS23-8 Parallel High-Frequency Electromagnetic Field Analysis Based on Hierarchical Domain Decomposition Method

Amane Takei[1*], Akihiro Kudo[2], Makoto Sakamoto[1]
([1]University of Miyazaki, Japan), ([2]National Institute of Technology, Tomakomai College, Japan)

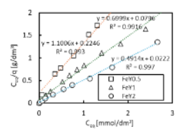
In this presentation, a parallel full-wave electromagnetic field analysis code based on an iterative domain decomposition method is explained that is named ADVENTURE_Fullwave. A stationary vector wave equation for the high-frequency electromagnetic field analyses is solved taking an electric field as an unknown function. Then, to solve subdomain problems by the direct method, the direct method based on the LDLT decomposition method is introduced in subdomains. The simplified Berenger's PML is introduced which these eight corners are given the average value of all PML's layers. And, we show a numerical example of a microwave in Fig.1. More detail will be shown in the conference.



OS23-9 Adsorption Equilibrium of Selenium Oxyanions Using FeY Mixed Oxides

Kaoru Ohe*, Amu Wakamatsu, Tatsuya Oshima (University of Miyazaki, Japan)

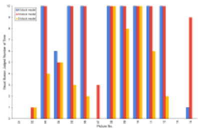
Selenium oxyanions (Se(IV) and Se(VI)) in wastewater are of concern as serious pollutants due to their easy bioaccumulation and toxicity to living organisms. In this study, mixed oxides with various Y/Fe molar composition ratios (x=0.5, 1, 2) were investigated the adsorption properties of selenium oxyanions, and were analyzed their adsorption mechanism. The experimental data best fits the Langmuir adsorption model, which is characteristic of monolayer adsorption (r^2 >0.991). The adsorption capacity ($q_{\max}$) of Se(VI) increased in the order of FeY2>FeY1>FeY0.5. The effect of ionic strength and surface charge analysis of Se(IV) on FeY2 revealed that Se(IV) was adsorbed on FeY2 formed an outer-sphere complex.



OS23-10 Influence of CNN Layer Depth on Spiral Visual Illusions

Kenji Aoki*, Makoto Sakamoto (University of Miyazaki, Japan)

Understanding how visual illusions are generated through Convolutional Neural Networks (CNNs) can contribute to elucidating the mechanisms of visual information processing in the brain. Our previous research demonstrated the potential for the spiral illusions to manifest in CNNs. In this study, we focused on the depth of the CNN and examined the effect of the number of layers on the manifestation of the visual illusion. We provided 14 types of spiral illusion images to three CNNs with varying layers and tasked them with distinguishing between concentric circles and spirals. The results indicated that CNNs with fewer layers were more prone to the illusion, whereas CNNs with more layers were less likely to exhibit the illusion. These results suggest that the number of layers in a CNN influences the manifestation of visual illusions.



OS24 New Media Interactions (3)

Chair R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)

OS24-1 Evaluation of Passive Interaction in XR Chakra Meditation Application Based on Behavioral Biometrics

P. I. A. Gayathri Bimba (Japan Advanced Institute of Science and Technology, Japan)
Chien-Tung Lin (Tainan National University of the Arts, Taiwan)
R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)
Kazunori Miyata (Japan Advanced Institute of Science and Technology, Japan)

This study explores the effectiveness of an XR chakra meditation application developed using behavioral biometrics. A direct comparison with a similar application was impossible since no commercial XR-based chakra meditation apps exist. Instead, the study compared the XR application with popular YouTube chakra meditation-videos, widely used for meditation, as indicated by their high viewership and subscriber counts. Participants provided feedback through questionnaires and brainwave data to evaluate its strengths and weaknesses of use. The experiment focused on several aspects, including the level of interference with meditation, ease of learning and use, physical and emotional effects on users, user preferences, and the perception of vibrational realism between the two methods.

OS24-2 ThoughtDiffusion: An Interactive Installation for Exploring Neuro-Art from EEG Data with Stable Diffusion Models

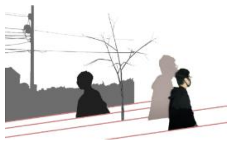
R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)

ThoughtDiffusion is an interactive installation that combines cognitive processes with the generative capabilities of AI, for generating artistic images in real-time. The installation system uses Stable Diffusion models and state-of-the-art neural decoding techniques that allow the mapping of brainwave patterns into coherent visual representations. The installation is based on a non-invasive commodity EEG headset that records users' brain signals which were fed into a stable diffusion model to output images corresponding to the intended mental state but unique to the participants. This installation uses a Kinect V2 sensor to capture users' body movements which advance multimodal interaction significantly relates to relaxation, the state of being calm, and the state of attention

OS24-3 PassBy2: Passive Interaction through the Pedestrian Counts and Real-time Weather Information

Chung Chien-Lin, R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)

There is a proverb in my country that "Constant dropping wears the stone ". it means even something as small as a drop of water can cut through a stone after a long period of accumulation. It expresses the idea that the small changes we make unconsciously in our daily lives can leave unique traces over time. This project focuses on the impact of people on their surroundings which concerned a contemporary street scene projected in an open space and captured pedestrians' pass-by counts using a Kinect sensor. Based on the measured counts of pedestrians passing by and real-time weather information, the developed application controls the color of the street scene and gradually decreases as the pedestrians cross. Using the L-system generative algorithm, the trees in the street scene grow progressively as the pedestrians cross.



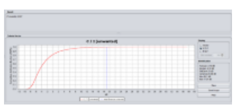
OS25 Robotic and Communications (6)

Chair Mastaneh Mokayef (UCSI University, Malaysia)
Co-Chair Takao Ito (Hiroshima University, Japan)

OS25-1 Exploring Techniques To Mitigate Interference In Drone Communication Systems

Ahmed Alsaeed Rashad[1], Mastaneh Mokayef[1]*, M.K.A Ahamed Khan[1], MHD Amen Summakieh[1], Kim Soon Chong[1], Abdul Qayyum[2], Moona Mazher[3], Sanjoy Kumar Debnath[4], Chin Hong Wong[5, 6], Chua Huang Shen[7]
([1]UCSI University, Malaysia), ([2] Imperial College London, UK), ([3] University College London, UK),
([4] Chitkara University Institute of Engineering and Technology, India), ([5] Fuzhou University, China) ,
([6] Maynooth University, Ireland), ([7] UOW University Malaysia)

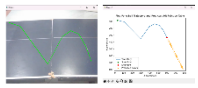
Despite the apparent advantages and strong economic efforts for cellular-connected drones, several critical challenges must be tackled for their successful implementation. Like any radio communication system interference is considered as the biggest challenge as it significantly decreases the efficiency and reliability of the drone. The Monte-Carlo simulation (MCS) strategy is based on the principle of taking samples of random variables from a given distribution. These samples are then used to assess interference in terms of the interference received signal strength compared to the desired received signal strength commonly known as C/I or SNR. The results then are derived using SEAMCAT software as probability of interference where 1 means that this system is always interfered and 0 means it's never interfered. The study has been conducted for separation distances of 2,3,4, and 6 km between the victim receiver and the interfering transmitter.



OS25-2 An Automated Tracking System for Locating Impact Points on a Table Tennis Surface Using Ping Pong Balls

Lee Wai Kit[1], Mastaneh Mokayef[1 *], MHD Amen Summakieh[1], M.K.A Ahamed Khan[1], Miad Mokayef[1], Sew Sun Tiang[1], Wei Hong Lim[1], Abdul Qayyum[2], Moona Mazher[3], Sanjoy Kumar Debnath[4]
([1] UCSI University, Malaysia), ([2] Imperial College London, UK), ([3] University College London, UK), ([4] Chitkara University, India)

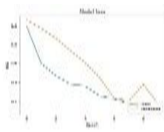
This project presents a novel system for real-time table tennis ball tracking and hitting point prediction, developed using OpenCV-Python. The system is designed to contribute to strategic analysis in the sport by accurately identifying the ball's trajectory and anticipating its landing point. The proposed system is comprised of four key modules: ball detection, ball tracking, hitting point prediction, and data visualization. Computer vision techniques are employed to effectively detect and monitor the ball's movement. Kalman filtering is utilized to refine the prediction of the ball's landing point. The generated data is then presented visually to facilitate analysis and comprehension of game dynamics.



OS25-3 An Innovative Deep Learning Technique to Identify Potato Illness

Abdul Majid Soomro[1 *], Muhammad Haseeb Asghar[2], Sanjoy Kumar Debnath[3], Susama Bagchi[3], and Awad Bin Naeem[2], M.K. A. Ahamed Khan[4], Mastaneh Mokayef [4]
([1] National University of Modern Languages, Pakistan), ([2] National College of Business Administration & Economics, Pakistan), ([3] Chitkara University, India), ([4]UCSI University, Malaysia)

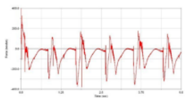
Potato cultivation is important for world food security as it itself is attacked by a great number of diseases like early blight and late blight, which cause a lot of damage to the yield and quality of the crop. But deep learning offers a great opportunity to address these disease detections; however, how effective this will be in the potato-growing environment in Pakistan is still not known. The project, therefore, tries to address data imbalance with the use of the synthetic minority oversampling technique (SMOTE) and develop a CNN architecture that is optimized to provide high diagnostic accuracy. This research can give innovative and productive locally useful solutions, which might transform the management of diseases for Pakistani farmers while improving food security and economic stability.



OS25-4 A Wearable Walking Support System Design And Simulation

Omar Ayaman Yehiya[1], M. K. A. Ahamed Khan[1 *], Mastaneh Mokayef [1], Ridzuan A[1], Abdul Qayyum[2], Moona Mazher [3], Susama Bagchi[4], Sanjoy Kumar Debnath[4]
([1]UCSI University, Malaysia), ([2] Imperial College, London, UK), ([3]University College London, UK), ([4]Chitkara University, India)

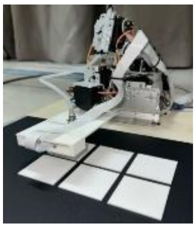
This research article is focused on the development of a Robotic lower limb exoskeleton model using MATLAB Simulink. The primary aim is to design a dynamic and flexible exoskeleton capable of assisting individuals with lower limb impairments, thus enhancing their mobility and overall quality of life. The model incorporates realistic representations of the lower limb anatomy, encompassing thigh, shank, and foot segments, with carefully integrated joints, constraints, and actuators to emulate natural human motion. A closed-loop control strategy optimizes the exoskeleton's performance, ensuring safe and stable operation during walking and other activities. Extensive simulations are conducted to evaluate the exoskeleton's efficacy, analyzing key parameters such as joint angles, joint torques, and power consumption



OS25-5 A Floor Tiling Robotic System

Hue Chau Jieng[1], M. K. A. Ahamed Khan[1] *, Mastaneh Mokayef[1], Ridzuan A[1], Abdul Qayyum[2], Moona Mazher[3]. Susama Bagchi[4], Sanjoy Kumar Debnath[4]
([1]UCSI University, Malaysia), ([2] Imperial College, London, UK), ([3] University College London, UK), ([4]Chitkara University, India)

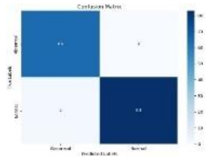
With the accelerated advancement of robot technology and sensor technology, construction challenges have become less difficult. The construction industry has been revolutionized by innovations in materials, equipment, and procedures, making it more efficient and safer.This paper proposes a Floor Tiling Robot robotic system that uses a vision-based solution to minimize labour-intensive, improve productivity, and increase the precision of the floor tiling process in order to reduce the material cost. The Floor Tiling Robot has implemented several systems, including a pneumatic vacuum suction system as a method for grasping floor tiles, a finite state machine as a method for robotic arm movement control algorithm and Canny Edge Detector algorithm as a method for floor tile positioning.



OS25-6 Evaluation of Heart Disease Risk Using Deep Learning Technique with Image Enhancement

Abdul Majid Soomro[1] *, Asad Abbas[2], Susama Bagchi[3], Sanjoy Kumar Debnath[3], Awad Bin Naeem[2], M. K. A. Ahamed Khan[4], Mastaneh Mokayef[4]
([1]National University of Modern Languages, Pakistan), ([2]National College of Business Administration & Economics, Pakistan), ([3]Chitkara University, India), ([4] UCSI University, Malaysia)

This study emphasizes the significance of the heart in the human body. Numerous serious vascular conditions exist in the heart and the blood. The dataset, study goals, methodology, approach, and efficient algorithms for identifying and classifying electrocardiogram (ECG) data are all covered in this paper. To assess ECG images, researchers used a convolutional neural network. Iterations in model training increase the accuracy. The system was constructed in Python using Matplotlib, NumPy, and Keras. The GPU-based machine learning platform was Google Colab. Photos were analyzed, categorized, and processed using MobileNet-V2. With a remarkable accuracy rate of 99.3 %, the developed model offers a viable basis for further hyperparameter investigation. Overall, this study combines advanced machine learning algorithms, to enhance the diagnosis of heart-related disorders.



OS26 Navigating the Digital Frontier: Innovations in the Age of Industry Revolution 4.0 (13)

Chair Wei Hong Lim (UCSI University, Malaysia)
Chair Takao Ito (Hiroshima University, Japan)

OS26-1 Driver State Monitoring Using Pose Estimation: Detecting Fatigue, Stress, and Emotional States for Safer Roads

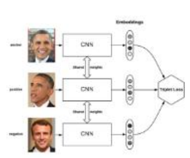
Hao Feng Chan[1], Dexter Sing Fong Leong[1], Shakir Hussain Naushad Mohamed[1], Wui Chung Alton Chau[1], Takao Ito[2], Zheng Cai[1], Xinjie Deng[1], Yit Hong Choo[1*]
([1]Deakin University, Australia) ([2]Hiroshima University, Japan)

Driving under fatigue, stress, or emotional impairment poses significant risks to road safety. This paper proposes a custom pose estimation framework designed to detect driver states, such as fatigue and stress, by analyzing body posture, head pose, and gesture dynamics. Using a novel deep learning approach trained on diverse driving scenarios, the model identifies physiological and behavioral markers associated with impaired states. Unlike existing methods, this system integrates pose estimation with real-time emotional and movement analysis, enabling robust performance in challenging conditions, including poor lighting and occlusions.

OS26-2 Optimizing Face Embedding Sizes and Accuracy in Facial Recognition Systems

Wui Chung Alton Chau[1], Dexter Sing Fong Leong[1], Shakir Hussain Naushad Mohamed[1], Hao Feng Chan[1], Takao Ito[2], Zheng Cai[1], Xinjie Deng[1], Yit Hong Choo[1*]
([1]Deakin University, Australia) ([2]Hiroshima University, Japan)

Face recognition technology is integral to security, access control, and identity verification in finance, healthcare, and transportation. It protects personal data, secures online transactions, controls access to areas, and helps prevent identity theft. This paper proposes a novel optimization algorithm to enhance face embedding sizes and facial recognition accuracy. We employ the FaceNet architecture, a deep neural network, alongside Triplet Loss metrics for efficient recognition. The model is tested with the Labelled Faces in the Wild (LFW) dataset, showcasing the effectiveness of the proposed algorithm.



OS26-3 Suspicious Behavior Detection Using Computer Vision

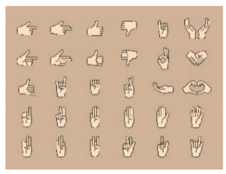
Dexter Sing Fong Leong[1], Hao Feng Chan[1], Shakir Hussain Naushad Mohamed[1], Wui Chung Alton Chau[1], Takao Ito[2], Zheng Cai[1], Xinjie Deng[1], Andi Prademon Yunus[2], Yit Hong Choo[1*]
([1]Deakin University, Australia) ([2]Hiroshima University)

Detecting suspicious activity is a crucial task for public safety. To be classified as suspicious, a person must exhibit specific body language, movement patterns, and reactions based on the context. Research has been conducted in this field using computer vision tools such as MediaPipe and YOLO. However, these approaches often face challenges, including high false positive rates and errors in detecting suspicious behavior in crowds. This paper aims to explore a novel approach to detecting suspicious activity using deep learning models, incorporating psychological behavior to classify suspicious behavior in crowded environments.

OS26-4 Sign Language Recognition Algorithms Using Hybrid Techniques

Shakir Hussain Naushad Mohamed[1], Hao Feng Chan[1], Dexter Sing Fong Leong[1], Wui Chung Alton Chau[1], Takao Ito[2], Zheng Cai[1], Xinjie Deng[1], Yit Hong Choo[1*]
([1]Deakin University, Australia) ([2]Hiroshima University)

Sign language recognition is a vital tool for enabling communication with individuals who are hearing impaired. This paper proposes a custom gesture recognition framework designed specifically for sign language interpretation. The proposed model uses pose estimation and gesture dynamics, incorporating a deep learning approach trained on diverse datasets from Roboflow. The system achieves robust recognition of complex gestures while maintaining efficiency. This framework emphasizes adaptability to variations in sign language styles.



OS26-5 Geographic and Risk Factor Analysis of Non-Communicable Cardiovascular Diseases in Central Java using Machine Learning

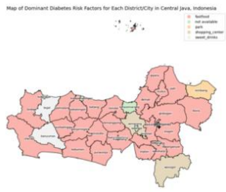
Nurhasanah, Andi Prademon Yunus*
(Telkom University, Indonesia)

Non-communicable diseases (NCDs), especially cardiovascular diseases, are a major health problem in Indonesia and globally with a mortality rate of 17.9 million people per year. Central Java Province is one of the regions with a high number of cardiovascular patients, especially hypertension in 2023. NCD risk factor analysis is needed to map the distribution and identify high-risk areas. This study uses Machine Learning and Geo-Mapping to analyze risk factors such as unhealthy food consumption, physical inactivity, alcohol, smoking, stress, and access to public transportation. The results of the study are expected to be the basis for planning appropriate intervention programs to reduce the prevalence of cardiovascular disease and improve the quality of life of the people of Central Java Province, Indonesia.

OS26-6 Analysis of Geographical Characteristics and Risk Factors for Non-Communicable Diseases: Diabetes in Central Java Using Random Forest and SHAP

Ambar Arum Prameswari, Andi Prademon Yunus*(Telkom University, Indonesia)

Non-Communicable Diseases (NCDs) account for 71% of global deaths, posing a significant public health challenge, particularly in low- and middle-income countries where 77% of these deaths occur. In Indonesia, NCDs contribute to 76% of total mortality, with diabetes responsible for 7%, making it one of the leading causes of death. Central Java, a province with high population density, has experienced a concerning rise in diabetes cases, reaching 618,546 in 2021. This chronic condition significantly increases the risk of complications such as cardiovascular disease and stroke, underscoring the urgent need for effective prevention strategies. This study utilizes machine learning to analyze lifestyle behaviors and environmental factors, aiming to identify the key risk factors for diabetes in Central Java and map their geographical distribution.



OS26-7 Geographic Analysis of Risk Factors for Chronic Respiratory Non-Communicable Diseases Using Machine Learning

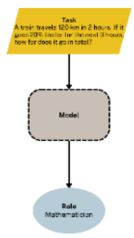
Ayu Susilowati , Andi Prademon Yunus*(Telkom University, Indonesia)

Non-communicable diseases (NCDs) account for approximately 74% of global deaths, with specifically 13% attributable to chronic respiratory diseases (such as asthma and COPD) in Central Java by 2023-a figure that is projected to continue to rise. The World Health Organization (WHO) has set a goal to reduce the incidence of NCDs by targeting behavioral risk factors, including alcohol consumption, air quality, smoking, and physical activity. This study aims to geo-mapping the distribution of non-communicable disease (NCD) risk factors offers crucial insights for pinpointing high-risk areas and developing targeted prevention strategies for chronic respiratory diseases by using machine learning.

OS26-8 Role-Play Prediction Using Ontology-Based Graph Convolutional Network Model

Asyafa Ditra Al Hauna[1], Andi Prademon Yunus[1*], Siti Khomsah[1], Fukui Masanori[2]
([1]Telkom University, Indonesia, [2]Iwate Prefectural University, Japan)

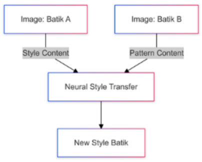
Current applications of large language models often assign tasks without consideration of how LLMs understand a given prompt. Simple commands sometimes do not guarantee desired responses, as LLMs are systems based on mathematical modeling and cannot cognitively be capable of understanding commands. Hence, a method is required to guide LLMs in performing tasks appropriately. This paper presents a method to develop model-based automation of role selection supported by ontology. This can allow for more accurate and relevant role recommendations than if done manually. As such, this optimization at hand improves the performance of LLMs for specific tasks and overcomes the limitations of previous works that define the roles by hand.



OS26-9 Experimental Exploration of Neural Style Transfer: Hyperparameter Impact and VGG Feature Dynamics in Batik Motif Generation

Happy Gery Pangestu[1], Andi Prademon Yunus[1*], Ratih Alifah Putri[1], Takao Ito[2]
([1]Telkom University, Indonesia) ([2]Hiroshima University)

Batik, a traditional Indonesian art form that employs wax-resist dyeing to create intricate patterns, holds deep cultural and historical significance. Despite its value, the complexity of its production has led to a decline in interest, particularly among younger generations. This study explores the application of Neural Style Transfer (NST), a subset of Artificial Intelligence, to innovate batik design by transferring styles from images onto content. In our research, we utilized three pretrained Convolutional Neural Networks (CNN) such as VGG, ResNet, and Inception. We compared the performance of these models using their original weights and fine-tuned versions. The results demonstrate significant variations in the quality of the batik designs, depending on the models and their hyperparameters. Additionally, the selected images were found to play a crucial role in influencing the final results.



OS26-10 Exploring Non-Communicable Disease Risk Factors on Cancer Rates in Central Java Using Random Forest and SHAP

Novi Ramadani, Andi Prademon Yunus* (Telkom University, Indonesia)

Non-Communicable Diseases (NCDs), particularly cancer, remain a significant public health concern in Indonesia and globally. Central Java Province is one of the regions with a substantial number of cancer cases in Indonesia, emphasizing the need to understand its contributing factors. Mapping the distribution of NCD risk factors provides valuable insights for identifying high-risk areas and designing targeted preventive strategies. This study aims to analyze the geographical distribution of NCD risk factors and identify the most influential factors associated with cancer prevalence across cities and regencies in Central Java Province using Machine Learning and Geo-Mapping.

OS26-11 Addressing Noise Challenges in CNN-based Pneumonia Detection: A Study Using Indonesian Thoracic Imagery

Wahyu Andi Saputra[1], Andi Prademon Yunus[1]
([1]Telkom University, Indonesia)

Accurate pneumonia diagnosis is vital, especially in resource-limited areas like Indonesia. While CNNs show promise for automated detection using chest X-rays, real-world image quality affects their performance. This study addresses this challenge by using a primary dataset—images directly from Indonesian patients—to avoid the biases of pre-processed secondary data. This ensures our findings are relevant to the Indonesian context. We tested how different noise types (salt-and-pepper and Gaussian) impact the accuracy of several common CNN architectures. These noise types mimic common image imperfections. Our analysis reveals how much noise degrades the CNN's ability to correctly identify pneumonia. This highlights the need for better pre-processing methods and potentially specialized CNN designs to handle noisy images. Ultimately, our work improves our understanding of deploying CNNs for pneumonia diagnosis in real-world settings, leading to more reliable and helpful diagnostic tools. Using primary data from diverse populations is crucial for building trustworthy AI in healthcare.

OS27 Industrial Revolution (4)
Chair Hazry Desa (UniMAP, Malaysia)

OS27-1 Digital Guardians: The Role of AI and Robotics in Protecting Construction Heritage

Muhammad Azizi Azizan, Nurfadzillah Ishak, Hazry Desa, (UniMAP, Malaysia)

Preserving construction heritage is vital for safeguarding cultural legacies. Traditional methods often face resource, environmental, and structural challenges. This paper explores AI and robotics in heritage conservation, highlighting their efficiency and sustainability. Tools like 3D modeling, digital twins, and predictive analytics are examined, with applications in structural monitoring, restoration, and documentation. The Sultan Abdul Samad Building in Malaysia serves as a key case study, demonstrating how technology preserves authenticity while reducing invasive interventions. Challenges, including costs, data limitations, and ethical concerns, are addressed. By merging tradition with innovation, AI ensures the protection of cultural landmarks in a digital age.

OS27-2 Architectural Memories: AI Redefines Dilapidation Analysis and Conservation

Muhammad Azizi Azizan, Nurfadzillah Ishak, Hazry Desa, (UniMAP, Malaysia)

Heritage conservation faces challenges from urbanization, environmental degradation, and resource constraints. AI and advanced technologies are revolutionizing traditional practices by offering efficient, data-driven solutions. Focusing on Georgetown, Penang, a UNESCO World Heritage Site, this paper examines AI-driven tools like 3D scanning, predictive maintenance models, and drones for structural analysis and restoration. Results show a 60% reduction in inspection time and targeted interventions for 40% of buildings. While AI enhances accuracy and efficiency, high costs and ethical concerns limit widespread adoption. By merging AI with traditional methods, the study highlights how innovation can protect architectural heritage while honoring cultural authenticity.

OS27-3 The Future of Robotics in Contract Management.

Muhammad Firdzaus Mat Ros, Muhammad Azizi Azizan, Hazry Desa,
(UniMAP, Malaysia)

Robotics and AI are revolutionizing contract management by automating tasks, improving accuracy, and enhancing efficiency. This paper explores the role of technologies like RPA, natural language processing, and predictive analytics in streamlining contract creation, review, and compliance. Real-world case studies highlight benefits such as faster processes and better decision-making. It also addresses challenges, including ethical and workforce implications. The integration of smart contracts and blockchain signals a transformative future, requiring continuous innovation and adaptation in contract management practices.

OS27-4 Leveraging AI to Enhance Extended Producer Responsibility Compliance in Construction Waste Management

Mohamed Fuad Shahariman, Muhammad Azizi Azizan, Hazry Desa, Nur Amierah Harun
(UniMAP, Malaysia)

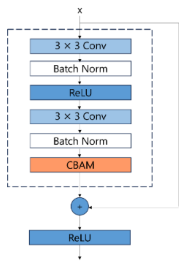
Extended Producer Responsibility (EPR) holds producers accountable for the environmental impact of their products, including waste management. In construction, EPR compliance is challenging due to material diversity and project complexity. This paper explores using AI technologies to enhance EPR compliance in construction waste management. AI-driven systems automate data collection, analysis, and reporting, improving efficiency, accuracy, and transparency. The study highlights AI's potential to revolutionize EPR compliance, delivering substantial environmental and economic benefits while addressing industry-specific challenges.

GS Abstract
GS1 Machine Learning & Autonomous Driving (5)
Chair Obada Al aama (Kyushu Institute of Technology, Japan)

GS1-1 Automatic classification of respiratory sounds by improving the loss function of ResNet

Ryusei Oshima[1], Tohru Kamiya[1], Shoji Kido[2]
([1]Kyushu Institute of Technology, Japan), ([2]Osaka University, Japan)

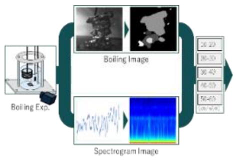
Respiratory diseases cause 8 million deaths annually, and this number is expected to increase. Breath auscultation, a primary diagnostic method, is noninvasive, repeatable, and immediate, but faces challenges such as reliance on skilled practitioners, difficulty in quantitative assessment, and limited accessibility in developing regions or disaster sites. To address these issues, we developed a deep learning-based breath sound classification system using the ICBHI 2017 dataset. Our method classifies breath sounds into four categories: Normal, Crackle, Wheeze, and Crackle and Wheeze. We use ResNet-34 as the base model, which is enhanced with CBAM for better spatial and channel feature extraction. To deal with class imbalances, we incorporate Focal Loss. The system achieves accuracy of 0.732, SE of 0.607, SP of 0.843, and ICBHI of 0.725.



GS1-2 Classification of Heat Transfer Coefficient Using Deep Learning with Information from Boiling Images

Fuga Mitsuyama, Ren Umeno, Tomohide Yabuki, Tohru Kamiya (Kyushu Institute of Technology, Japan)

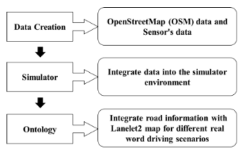
Recently, the high integration of electronic devices in many products has led to an increase in heat generation density. Boiling-cooling is attracting attention as a method that can efficiently cool heat sources. In this paper, we propose a method to predict the heat transfer coefficient (HTC), which is important for the design of cooling systems, from boiling sound data and boiling images using deep learning. This model learns the physical law of boiling by considering the information of bubbles, which significantly affects the HTC value, and improves the prediction accuracy compared to the conventional method using only the boiling sound data.



GS1-3 A Data Format Integration of Open-Street-Map and Lanelet2 Toward the Ontology Framework for Safety Autonomous Driving systems

Obada Al aama[1], Takahiro Koga[1], Tomoki Taniguchi[1], Davaanyam Jargal[1], Junya Oishi[2], Shigeru Nemoto[2], Wataru Mizushina[2], Kazuki Hirao[2], Hakaru Tamukoh[1], Hiroaki Wagatsuma[1]
([1]Kyushu Institute of Technology, Japan, [2]Aisan Technology Co., Ltd., Japan)

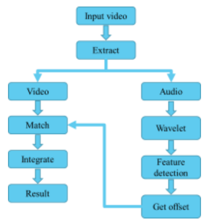
This study proposes a framework integrating OpenStreetMap (OSM) data with ontology-based systems to enhance autonomous driving systems. OSM provides static geographical data, while the Lanelet2 mapping framework incorporates lane-level road information and topological relationships, enabling advanced testing of vehicle behavior. Ontology-based integration offers semantic representations of road elements and traffic rules, supporting realistic modeling of complex driving scenarios. This structured approach ensures accurate simulation and testing, facilitating applications such as traffic analysis and route optimization, while improving scalability, precision, and safety in autonomous vehicle development.



GS1-4 Developing a Sound-Based Method to Synchronize Multiple Videos Recorded by Multiple Sound Sources

Davaanyam Jargal, Rena Kato, Tomoki Taniguchi, Kosei Shibata, Takahiro Koga, Obada Al aama, Hakaru Tamukoh and Hiroaki Wagatsuma (Kyushu Institute of Technology, Japan)

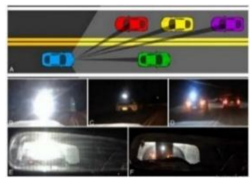
The environmental monitoring from a vehicle on the road is important for the driving behavior analysis as well as the monitoring of driver's operations. A light weight mobile camera is useful for the record of multiple directions from the vehicle simultaneously; however the timing synchronization is an issue need to be solved. In this purpose, we proposed the sound-based method to synchronize different videos recorded with environmental sounds. In this task, the extraction of common sound features and amplifying of the features are necessary to superimpose those sound profiles to find consisting time points. In the validation the effectiveness, we used recoded videos from the bus driven by an expert driver.



GS1-5 Intelligent Path Planning for Robots and Practical Implementation of Programmable Headlights for Autonomous Vehicles remote presentation

Farkad Adnan, Abdul Samad Bin Shibghatullah, Mohd Radzi Bin Aridi
(Universiti Tenaga Nasional (Uniten), Malaysia)

The ability to move is essential for the development of intelligent robots for autonomous navigation. Neural networks outperform traditional methods in modeling complex relationships and identifying patterns, but current systems are limited to specific robots and sensors. This paper presents a universal method for interpreting data from different 2D sensors, predicting distances between robots and walls, and using neural networks for navigation. The goal is to create a versatile algorithm that can be applied to different robots and programmable lamps, reducing accidents. The thesis also explores programmable lamps that block light from reaching the eyes of passengers, using one network to determine free space using odometry data and another to find safe paths while avoiding obstacles. Simulated path examples will be presented.



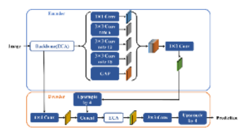
GS2 Image Processing I (4)

Chair Amane Takei (University of Miyazaki, Japan)

GS2-1 Recognition of Plastic Bottles Region Using Improved DeepLab v3+

Yusuke Murata, Tohru Kamiya (Kyushu Institute of Technology, Japan)

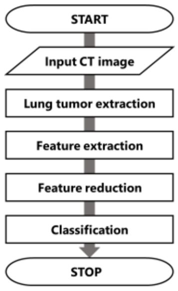
The shortage of factory workers is a major problem in Japan. This paper focuses on sorting plastic bottles at waste disposal plants as one of the solutions. We try to realize automation by recognizing plastic bottles from images and gripping them with a robot arm. In this paper, we propose an image analysis method limited to plastic bottles of 500 ml capacity. As a basic study, experiments were conducted to investigate how accurately a single plastic bottle can be recognized in an image. The method is semantic segmentation, and DeepLab v3+ was used as the deep learning model, with improvements by ECA block and Mish activation function. We compared the methods and found that the proposed method showed improvement in the area of misrecognition in the base model.



GS2-2 Non-Invasive Classification of EGFR Mutation from Thoracic CT Images Using Radiomics Features and LightGBM

Reo Takahashi[1], Tohru Kamiya[1], Takashi Terasawa[2], Takatoshi Aoki[2]
([1]Kyushu Institute of Technology, Japan), ([2]University of Occupational and Environmental Health, Japan)

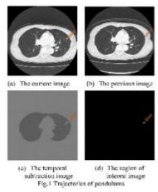
Cancer caused 9.7 million deaths in 2022, including 1.8 million from lung cancer, the leading cause of cancer death. EGFR mutation testing is essential for lung cancer treatment planning, but it is invasive and visual identification from chest CT images is difficult. This paper proposes a computer-aided diagnosis system to identify EGFR mutation status. Lung tumor regions were automatically extracted and radiomics features were obtained. Dimensionality reduction was performed using null importance, variance inflation factor, and recursive feature elimination. The method was applied to 143 cases and achieved 84.6% accuracy, 94.8% true positive rate and 25.5% false positive rate. The results suggest that CAD systems can improve the non-invasive detection of EGFR mutations in lung cancer.



GS2-3 Detection of Lung Nodules from Temporal Subtraction CT Image Using Elastic Net-Based Features Selection

Natsuho Baba, Tohru Kamiya (Kyushu Institute of Technology, Japan)
Takashi Terasawa, Takatoshi Aoki (University of Occupational Health)

CT (computed tomography) is mainly used to diagnose lung cancer. Many CT images impose a heavy burden on visual screening, so a CAD (computer-aided diagnosis) system is expected to reduce the burden. In this paper, we propose an image analysis method to detect lung nodules from chest CT images using machine learning techniques. The best results were obtained for the method using LightGBM with feature reduction by Elastic Net.



GS2-4 Detection of Lung Nodules from CT Image Based on Ensemble Learning

Natsuho Baba, Tohru Kamiya (Kyushu Institute of Technology, Japan)
Takashi Terasawa, Takatoshi Aoki (University of Occupational Health)

Lung cancer is the most diagnosed cancer worldwide and the leading cause of cancer-related deaths, making early detection and treatment crucial. One method used in computer-aided diagnosis (CAD) systems is the temporal subtraction of images by performing a different operation between the current and previous images of a patient. In this study, radiomic features are extracted as explanatory variables from the temporal subtraction images. Feature selection is performed using Elastic Net, followed by the application of machine learning methods. Finally, ensemble learning is applied to classify unknown data into two categories: positive and negative lung nodules.

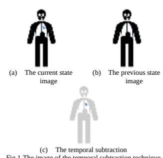


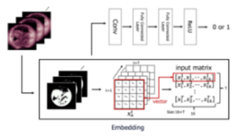
Fig.1 The image of the temporal subtraction technique

GS3 Image Processing II (4)
Chair Kuo-Hsien Hsia (National Yunlin University of Science and Technology, Taiwan)

GS3-1 Shape-Preserving Embedding Technique for Binary Classification of Video Image of the Solar Surface

Iori Tamura, Akiko Fujimoto, Soichiro Kondo, Reiri Noguchi (Kyushu Institute of Technology, Japan)

We study the embedding technique on the binary classification of video images as the explanatory variable. In this study, we assume the shape on video frame image have high sparsity and strong characteristic time evolution. In the embedding process, 2-dimensional image is resized keeping shape characteristics of the image and converted to a vector. The embedding allows dimensionality reduction from a 3-dimensional array (video image) as input data for machine learning to a 2-dimensional array of time sequences of embedded vectors. Using solar surface video images in the space weather field, we present evaluation experiments on multiple models with different embedding sizes, transformation formulas, and number of layers in the CNN.



GS3-2 Seated Posture Estimation Based on Monocular Camera Images

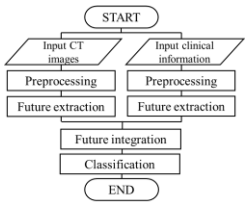
Hitoshi Shimomae, Tsubasa Esumi, Noriko Takemura (Kyushu Institute of Technology, Japan)

Poor seated posture significantly strains the body, leading to symptoms like shoulder stiffness and back pain. While research on seated posture estimation using images has been active, many studies focus on extreme postures not typically seen in daily desk work. This study aims to estimate common postures, such as slouching, which are often experienced in everyday settings. Due to the lack of medical quantitative metrics for evaluating posture quality, we manually annotated some of our collected posture data and used semi-automatic annotation by SVM to build a dataset. Using this dataset, we trained deep learning models for posture estimation with different input data types: RGB images, silhouette images, and posture key points, and compared their performance.

GS3-3 Identification of lung nodules based on combining multi-slice CT images and clinical information

Yuto Nishitaki, Tohru Kamiya (Kyushu Institute of Technology, Japan) Shoji Kido (Osaka University, Japan)

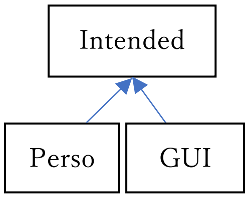
In recent years, a CT scan of the chest has played an important role in the diagnosis of lung cancer. However, the large number of CT images generated in a single examination places a heavy burden on the physician reading them. To solve this problem, computer-aided diagnosis (CAD) systems are being introduced. Conventional CAD systems are based on methods that use only image information. Recently, however, research on systems that incorporate clinical information in addition to image information has attracted attention. In this study, we propose a method for identifying nodular shadows that integrates a composite image created from multi-section CT images with clinical information such as the patient's age, sex, and medical history as recorded in the medical record.



GS3-4 Graphical User Interface (GUI) Design for Mobile Commerce Site for Women Seller in Rural Area remote presentation

Kho Irene[1], Shayla Islam[1], Abdul Samad Shibghatullah[2]
([1]UCSI University, Malaysia), ([2]University Tenaga Nasional, Malaysia)

GUI is a medium that allows users to communicate with electronic devices. Good GUI design is important because it allows users to learn how to use a system in the shortest amount of time and effectively operate it. Hence, a well-designed GUI for m-commerce sites is crucial to aid women sellers in rural areas to sell their products directly to customers. Unfortunately, the GUI of m-commerce sites caters only to IT-literate users with fast internet connections. Existing m-commerce sites' GUIs are compact with information. This may lead to an interface design that looks busy and messy unless women sellers are given proper guidance by IT experts. A research review about GUI requirements caters for women sellers' according to their personality types and hopes that they will be willing to explore m-commerce sites.

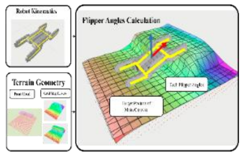


GS4 Robotics (5)
Chair

GS4-1 Obstacle-Aware Autonomous Flipper Control Method Based on Terrain Geometry

Kotaro Kanazawa, Noritaka Sato, Yoshifumi Morita (Nagoya Institute of Technology, Japan)

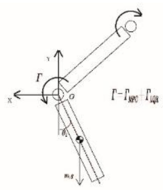
Rescue robots play crucial roles in search and rescue missions at disaster sites. Crawler mechanisms, which are valued for their high traversability on unstructured terrain, can be equipped with flipper arms featuring four single-rotation joints to enhance their performance. However, operating these flippers is challenging. Although various control methods have been developed recently, difficulties persist in adapting to three-dimensional uneven terrains and in avoiding interference between flippers and the environment. In this study, we propose an obstacle-aware autonomous flipper control method that actively adjusts the flippers based on the progress of the robot and terrain geometry.



GS4-2 An Adaptive Control Method for a Knee-Joint Prosthetic Leg Toward Dynamic Stability and Gait Optimization

Ge Yiqian[1], Purevdorj Choisuren[1], Shintaro Kasai[1], Hiroaki Wagatsuma[1]
([1]Kyushu Institute of Technology, Japan)

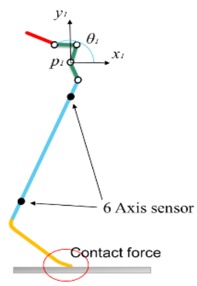
This study presents a control strategy for the knee joint of a prosthetic leg designed for below-knee amputation people. Model Predictive Control (MPC) and Linear Quadratic Regulator (LQR) are considered to be useful for the achievement of dynamic stability and adaptive motion control. MPC is employed to optimize the knee joint's trajectory in real-time, ensuring smooth and efficient motion while accommodating gait characteristics. LQR is integrated to regulate the center of mass (CoM) dynamics, maintaining overall stability by leveraging feedback from the prosthetic leg and user inputs. In the computer simulation, the feasibility of those control approaches will be examined whether they enhance a smooth control, in the sense of the stability and adaptability in human daily activities.



GS4-3 A Gait Analysis with Multibody Dynamics Toward Energy-Efficient Active Knee Prostheses

Purevdorj Choisuren[1], Ge Yiqian[1], Shintaro Kasai[1], Batbaatar Dondogjamts[2], Erdenesuren Naranbaatar[2] and Hiroaki Wagatsuma[1]
([1]Kyushu Institute of Technology, Japan; [2]Mongolian University of Science and Technology, Mongolia)

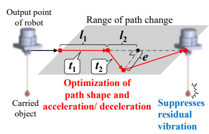
In general, prosthetic knee users have a large stress in the locomotion due to less smoothness and unnecessary energy consumption. In the passive prosthesis, it is difficult to regulate the stiffness depending on the ground contact force. In consideration of designs for such an adaptivity to improve passive systems, we propose an artificial knee kinematics design to absorb the redundant contact force for the smooth and stable walking and explore necessary constraints for the proposed mechanics to be able to have multi-functions not only for walking but also knee flexion accumulating the power for jumping. In the analysis, we used Multibody Dynamics (MBD) to investigate. This result will contribute to design an integrated dynamic model by incorporating a flexible body and ground contact forces in various purposes.



GS4-4 Suppressing of Multi-Axial Vibration Caused in Carried Objects by Robot Using a Heuristic Algorithm Based on Evaluation of Actual Machine Information

Yusuke Ueno[1], Hiroki Noguchi[1], Fumitoshi Shimono[1], Hiroshi Tachiya[2]
([1]Komatsu University, Japan) ([2]Kanazawa University, Japan)

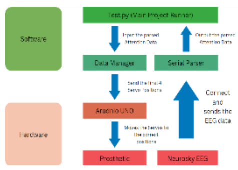
Residual vibrations induced in objects carried by robots cause the positioning accuracy to deteriorate, which makes the installation of carried objects difficult. This study proposes a method for determining a trajectory that can suppress residual vibration using a heuristic algorithm, based on the behavior of an object to be carried, which is measured by actually operating a robot. The method does not require a kinetic analysis of the carried object and can be applied to robots still in operation on the production line. Trajectories were generated for a commercially available industrial robot to suppress residual vibrations in two axial directions for a pendulum-shaped object to be transported. A significant reduction in vibration amplitude was achieved by optimizing path shape, acceleration and deceleration.



GS4-5 Developing Low-Cost BCI-Based Brain-Limb Interaction Device with Prosthetic Hand remote presentation

Nethika Jayith Rajapakse(International Bilingual International Science Park, Taiwan)

Inventions that are designed to heal the body and/or mind should always be sought. Currently, too many people are lacking access to cheap prosthetic devices, especially those that allow neural connections to be gained with such devices. This is the reason why this study intends to propose a low-cost invention that is able to enable people to gain back concentration with limb use using Python Programming language, an EEG Neurosky headset, an Arduino, and a 3D-printed prosthetic hand. With our implementation, the proposed method can explore the boundaries of improving attention and continue to develop higher-level BCIs for the mind-limb connection, improving on current-day prosthetics and concentration/limb connection rehabilitation.



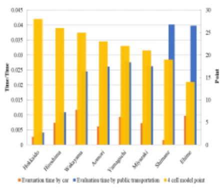
GS5 Applications I (6)

Chair Tsutomu Ito (Ube National College of Technology, Japan)

GS5-1 A Study on Local Airports Contributions to Tourism Industry in Japan

Tsutomu Ito[1], Seigo Matsuno[1], Makoto Sakamoto[2], Satoshi Ikeda[2], Takao Ito[3]
([1] Ube National College of Technology, Japan), ([2]University of Miyazaki, Japan)
([3]Hiroshima University, Japan)

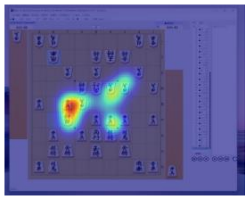
This study examines the impacts of local airports on the regional tourism industry in Japan. A series of indexes of the tourists and a new measure of the time index have been developed based on our four-cell model to analyze the correlation between different indexes such as the number of airport tourists and the time used by public transportation. During the verification process, a correlation has been discovered between the number of tourists in the city where the airport is located and the number of airport users for specific regionally managed airports. Furthermore, regarding access to tourist attractions from airports, a strong negative correlation is confirmed between the evaluation index and the state of development of public transport, suggesting the importance of developing public transport, including airports, in attracting tourists.



GS5-2 Analysis of Careless Mistakes Using Gaze Information

Ryota Yabe, Noriko Takemura (Kyushu Institute of Technology, Japan)

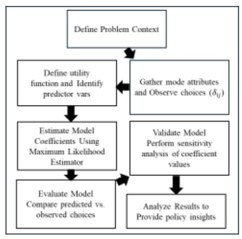
Careless mistakes are caused by a lack of concentration, time pressure, and information overload due to multitasking. If we predict careless errors, they could be helpful in various situations, such as learning support, business efficiency improvement, and medical diagnosis support. However, it is challenging to reproduce careless error situations, and few studies have been on predicting careless errors. Therefore, we focus on Shogi (Japanese chess), where situational awareness is complex, and players must maintain long periods of concentration, making careless mistakes. In this study, we collected eye-tracking data of players during a game and annotated careless errors based on the game's contents and surveys from the players. By analyzing this data, we have examined the conditions under which careless mistakes occur.



GS5-3 A Mathematical Framework for Logit Model in Transportation Mode Choice Analysis

Ahmad Altaweel[1], Kazuhito MINE[1], Bo-Young Lee[2], Jang-Sok Yoon[2], Hiroaki Wagatsuma[1]
([1]Kyushu Institute of Technology, Japan; [2]Logistics Revolution Korea, Korea)

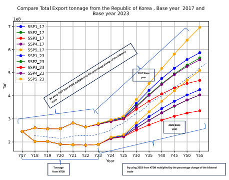
Traditional transportation demand forecasting has relied on massive zone-specific aggregations, which assume a linear demand increase. Such models may lack the flexibility needed to perform dynamic and context-sensitive analyses. Recently, disaggregated behavioral models have gained prominence for requiring less data and enabling sensitivity analyses in policy decisions. This study explores the feasibility of a model focusing on the mathematical formulation and validation of the transportation mode choice model. The study uses the logit model with a non-linear probability distribution function represented by a logistic curve and incorporates a linear combination of independent predictor variables. The mathematical model is examined for its ability to estimate choice probabilities. The methodology is formulated to be adaptable to diverse contexts that provide an analytical framework for transportation systems independently of geographic or demographic considerations.



GS5-4 A Computational Approach for Global Trade Analysis in Korea Contributing to the Forecasting of Future Efficacy in Global and Domestic Korean Transportations

Bo-Young Lee[1], Ahmad Altaweel[2], Kazuhito Mine[2], Jang-Sok Yoon[1], Hiroaki Wagatsuma[2]
([1]Logistics Revolution Korea Co., Korea; [2]Kyushu Institute of Technology, Japan)

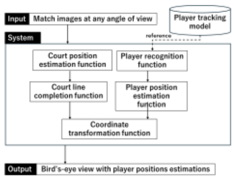
Economic forecasting studies are integral for shaping strategic policy decisions by providing data-driven insights that guide resource allocation, logistics and transportation, and long-term planning. This study investigate the trade dynamics of the Republic of Korea through the Global Trade Analysis Project Recursive dynamic GTAP-RD model with the GTAP v11 database to forecast economic scenarios and Shared Socioeconomic Pathways (SSPs) serve as growth trajectories. The analysis centers on the Republic of Korea's key trading partners, as identified by the GTAP database, and top trading sectors from the Korea Transport Database (KTDB) to compare the key influencers on Korea's trade thereby providing deeper strategic economic planning. This study investigates further the involvement of Korea's logistics and transportation by focusing on changes in import/export tonnage to inform infrastructure planning and strategic transport development. The evaluation of Korea's trade in the global context is of the essence to ensure adaptive logistics, backing economic resilience, and aligning with changeable global trade conditions.



GS5-5 Fundamental Research on Athlete Positions Estimation in Indoor Sports at Various View

Iori Iwata, Yoshihiro Ueda, Kazuma Sakamoto, Riku Kaiba (Komatsu University, Japan)

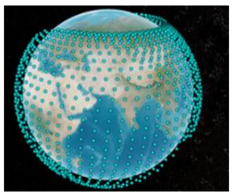
In sports such as volleyball, basketball, and soccer, the positioning of players is of great importance for the purposes of tactical analysis. In many cases, this is researched using image recognition techniques, which facilitate the creation of visual representations of player positions. Typically, images captured by cameras are transformed into a bird's-eye view through the application of algorithms such as projective transformation. Traditional methods rely on four reference points such as court line intersection to transform images into a bird's-eye view, which can be difficult to obtain because camera angles and camera angles of view change during a match. This research proposes a new method to overcome these limitations by automating the selection of reference points using net and court line coordinates. This approach enables accurate player position estimation even with partial court images and historical video data, expanding the possibilities for tactical analysis in volleyball.



GS5-6 Overview of the development of low earth orbit satellite navigation enhancement technology remote presentation

Dingcheng Tang*, Jinliang Wang, Jianfeng Shan, and Guoji Zou
(Space star technology co, LTD, Beijing, China)

Low Earth Orbit (LEO) satellites are expected to become a new increment in the development of the new generation of satellite navigation systems. The development direction of satellite navigation enhancement is gradually tilting towards the low orbit field, becoming a new growth and empowerment point for the next generation of satellite navigation. This article was on discussing the new development opportunities brought by low orbit navigation enhancement technology for building a Global Navigation Satellite Systems (GNSS) global space-based monitoring network, providing global quasi real time high-precision services, and providing global high integrity monitoring services.

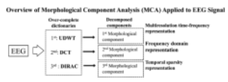


GS6 Applications II (5)
Chair Marion Oswald (TU Vienna, Austria)

GS6-1 Signal Decomposition and Noise Reduction in Single-Channel EEG: A Morphological Component Analysis (MCA) Approach

Kosei Shibata[1], Yide Yang[1], Rena Kato[1], (Kyushu Institute of Technology, Japan),
Hendry Ferreira Chame[2], Laurent Bougrain[2] (Université de Lorraine, CNRS, LORIA, France),
Tomohiro Shibata[1], Hiroaki Wagatsuma[1](Kyushu Institute of Technology, Japan)

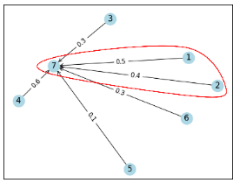
This study applies the Morphological Component Analysis (MCA) to single-channel EEG data obtained during human-to-human interactions in a board game (Hex-game). MCA, a dictionary-based signal decomposition method, separates signals into distinct morphological components. It enables the extraction of plausible brain activity and the removal of noise, such as ocular and muscular artifacts. By focusing on neural dynamics in interactive settings, this approach highlights the relationship between cognitive processes and social behavior. The approach suggest that MCA offers a promising framework for EEG analysis in complex, dynamic environments, combining effective feature extraction with robust artifact removal.



GS6-2 Variable Selection Methods for Multivariate Time Series Data Using Multivariate Granger Causality

Keita Ohmori[1,2], Toshiki Saitoh[1], Akiko Fujimoto[1], Eiji Miyano[1]
([1]Kyushu Institute of Technology, [2]SUMCO, Japan

In this paper we study variable selection methods for multivariate time-series data. Hmamouche et al. (2018) proposed a method that first constructs a causal graph based on Granger causality among time-series data, and then selects variables from clusters formed by clustering the vertices corresponding to each variable. However, this method only performs pairwise Granger causality tests, which may not fully capture the interactions among variables. To address this issue, we propose a variable selection method that performs multivariate Granger causality tests on all combinations of explanatory variables with respect to the target variable, selecting the combination with the strongest causality. Our new method successfully constructs a predictive model with a higher accuracy compared to the previous method.



GS6-3 A Support System for a Visually Impaired Person Finding Bus Route Numbers Employing MY VISION

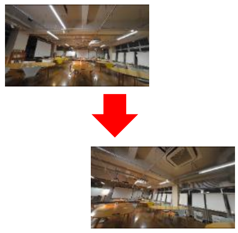
Daichi Nanaura, Seiji Ishikawa, Yui Tanjo (Kyushu Institute of Technology, Japan)

A bus is not a very convenient means for a visually impaired person because of the difficulty in identifying its route number , although it is an economical tool for travel. This paper proposes a method identifying a bus route number using MY VISION system which employs an ego camera worn by a visually impaired person. The method finds a frontal area of an approaching bus using the video which the camera provides using optical flow and the random forest based on Haar-like features. It then extracts the upper destination panel area followed by detecting a route number at the right-hand side of the panel. Finally, the detected route number is identified by template matching. In the experiment, various kinds of videos containing busses were captured and the effectiveness of the proposed method was shown.

GS6-4 Human Pose Estimation from Egocentric Videos

Shunya Egashira, Yui Tanjo (Kyushu Institute of Technology, Japan)

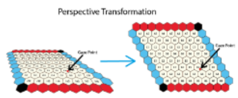
According to a survey conducted by the Ministry of Health, Labour and Welfare in 2019, about 30% of patients suffer from back pain and stiff shoulders. Although researches on pose estimation have been conducted for a long time, they cannot be used for daily pose estimation, because they need fixed cameras to capture target/subject motion. To solve this problem, the present paper, proposes a novel pose estimation method from egocentric videos using Epipolar Geometry. It computes three rotational angles, i.e., pitch, yaw and roll, from the egocentric motion videos to evaluate differences from his/her normal motion. In the experiment, three egocentric videos were used to verify the performance and effectiveness of the proposed method and reasonable/satisfactory results were obtained.



GS6-5 Analyzing Eye-Tracking Data to Detect Joint Attention in Hexgame Experiments

Yide Yang[1], Rena Kato[1], Kosei Shibata[1] (Kyushu Institute of Technology, Japan), Hendry Ferreira Chame[2], Laurent Bougrain[2] (Université de Lorraine, France),
Hiroaki Wagatsuma[1](Kyushu Institute of Technology, Japan)

This study aims to explore the mechanisms of joint attention in Hexgame experiments by analyzing Tobii eye-tracking data. In the experiment, two participants engaged in a strategic game, during which their gaze data was recorded. Through perspective transformation, the gaze points were mapped onto the coordinates of the game board to determine whether joint attention occurred during critical game states. The primary focus of this study is the analysis of eye-tracking data to identify attention coordination between players during the progression of the game. Future work will expand this framework to assess win probabilities and predict subsequent moves, providing deeper insights into strategic decision-making.



AUTHORS INDEX

Notation of session name

PS: Plenary Session IS: Invited Session, OS: Organized Session, GS: General Session,

Note: 33/90 = (page no. in Technical Paper Index) / (page no. in Abstracts)

High-fidelity Multi-agent Simulations for Social Systems

Shinobu Yoshimura
Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8589, Japan

Hideki Fujii
School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-8656, Japan
Email: syoshi@edu.k.u-tokyo.ac.jp, fujii@sys.t.u-tokyo.ac.jp

Abstract

For realizing the rational and quantitative design of social systems, we have been developing intelligent multi-agent based simulations together with high-fidelity models of social systems. The one is MATES (Multi-Agents based Traffic and Environmental Simulator), and the other is a virtual nursing care process simulator based on a multi-agent model. In this talk, I first describe the objectives of the research and some key technologies, and then introduce their practical applications with verification and validation, i.e. Tramway extension problem in an actual middle sized-city, Okayama, Japan, and an excretion care process in an actual day-care facility for elderly persons.

Keywords: Multi-agent simulation, High-fidelity model, Traffic simulation, Excretion care process

1. Introduction

Together with the rapid progress of computer technologies, computer simulations such as computational mechanics and computational science have been popular in various science and engineering fields. Simulation-based design of artifacts is now a central concept in CAE areas, because its superior capability of quantitative evaluation of artifacts' mechanics and physics behaviors. Such a capability of quantitative evaluation is also strongly demanded for the rational and efficient design of social systems, i.e. Simulation-based design of social systems.

Since 2000, we have been developing intelligent multi-agent based simulations together with high-fidelity models of social systems. The one is MATES (Multi-Agents based Traffic and Environmental Simulator) , and the other is a virtual nursing care process simulator based on a multi-agent model.

Road traffic is a key portion of the infrastructure supporting mobility through the transportation of humans and goods. At the same time, it is the cause of various types of urban and environmental issues including traffic jams, accidents, heavy energy consumption, as well as air pollution and global warming due to engine emissions. Since it is very difficult to restore a road environment to a previous condition once it has been changed, it is strongly desirable to accurately estimate the impact of transportation policies quantitatively. This is why simulations have been playing an important role in the field of traffic systems.

On the other hand, the Government of Japan has taken several measures to secure human resources for long-term care, including improving work efficiency by care workers, improving a working environment, and providing effective training for new employees. However, there are major barriers in resolving the issues, such as difficulties in public surveys due to a variety of situations involving personal data in long-term care facilities, and difficulties in quantifying information and features on care workers and care receivers in actual nursing care processes. Thus, as the first step, specifically focusing on "excretion care, "which is regarded to be a particularly high burden on care workers, we newly construct a virtual nursing care process simulator based on a multi-agent model.

In this lecture, we mainly describe our high-fidelity multi-agent traffic simulator for mixed traffic of cars, pedestrians, and trams. This simulator is then applied to a real-world case study on a tramway extension plan in an actual city. Based the results with other considerations, the Okayama-city has decided to extend the tramway into the station square, and it is under construction. Then, we briefly describe a virtual nursing care process simulator based on a multi-agent model. Applying it to simulate care processes in an actual day-service facility, we have confirmed that the simulated results agree with observed ones reasonably well.

2. Tramway Extension Simulation in Okayama-city

2.1. Target problem

In Okayama-city, which is the capital of Japan's Okayama Prefecture, many people have chosen cars as their primary mode of transportation, which causes various traffic problems. There are two tramlines operating in the city. The Okayama-city government had a plan to extend the tramway into the station square to improve access and convenience of movement for rail users around downtown areas. Because the tramway needs to cross the large and busy intersection in front of the rail station, and since there are plans to reduce the duration of green light time for car traffic, there are concerns that the project will have a negative impact on car traffic flow. The Okayama-city government asked us to perform the traffic simulation on this issue. We attempted to quantitatively clarify the impact of the proposed tramway extension.

2.2. Data preparation

The road network used in this research is shown in Fig. 1. Even if the extension plan were limited to around the rail station, we considered it probable that its impact would spread widely. The node positions and connections were set by importing data from OpenStreetMap2, and the precise lane configuration and signal control information were provided by the Okayama Prefectural Police (OPP).

Fig.1 Road network in Okayama city

The time period to be simulated was set as one hour from 13:00 to 14:00 on a holiday, which is one of the time periods with the highest traffic volumes. The OD (Origin-Destination) matrix for this time period was estimated iteratively via the gravity model by referring to the observed link traffic volume, which was provided by OPP. A comparison of link traffic volume in the simulation and that in the real world is shown in Fig. 2. The correlation coefficient was 0.939 and the slope of the regression line was 0.923. Based on these simulated results, we judged that the estimated OD matrix was reliable enough to be used as the baseline in the subsequent simulation scenarios.

The observed traffic volume of pedestrians at the intersection was also provided by OPP. The actual routes and timetables were provided by the bus and tram agents.

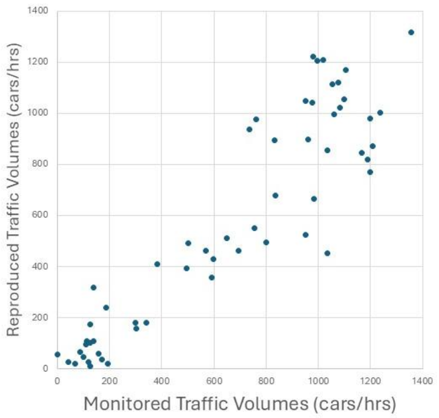


Fig. 2 OD matrix estimation result

2.3. Simulated scenarios

The road environments around the intersection with and without the tramway extension, are shown in Figs 3(a) and 3(b), respectively. The main difference is that the tram stop is relocated from the east side of the intersection to the station square. As a result, the signal pattern at the intersection needs to be changed, and it is also expected that the volume of pedestrians on the east side crosswalk that are getting on and off of the tram will be reduced. In this research we conducted simulations according to the following four scenarios, before and after tramway extension, and measured the distance from the intersection to the rearmost car agent, which is stopped by a red signal in each scenario.

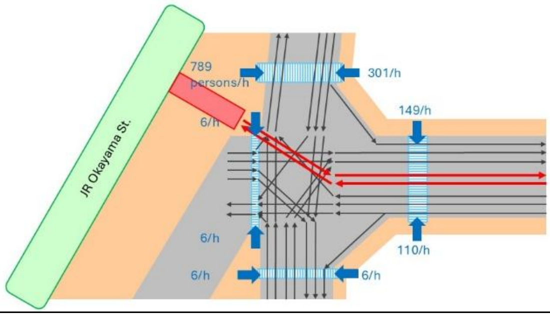


Fig. 3(a) Intersection with tramway extension

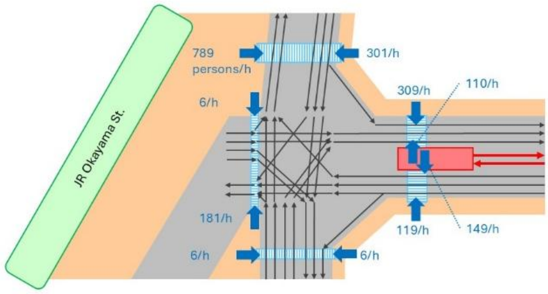


Fig.3(b) Intersection without tramway extension

Baseline scenario 1 (baseline_1): without tramway extension, without pedestrians

Baseline scenario 2 (baseline_2): without tramway extension, with pedestrians (current status)

Extension scenario 1 (ext32): with tramway extension (green light for tram: 32 seconds, 1 time per cycle), with pedestrians

Extension scenario 2 (ext64): with tramway extension (green light for tram: 32 seconds, 2 times per cycle), with pedestrians

The signal cycle at the intersection is fixed at 150 seconds for all scenarios to maintain coordination with neighboring signals. The minimum time required for a tram to cross the intersection is 32 seconds.

2.4. Simulation results

Figure 4 shows a simulation screenshot around the intersection for ext32. The red rectangles represent passenger cars, black rectangles represent trams (which are not seen in the screenshot), and green circles represent pedestrians. Small squares on the lanes indicate traffic light colors or passage permissions.

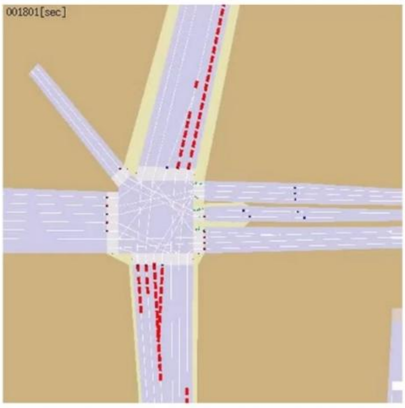


Fig. 4 Simulation screenshot around the intersection in front of Okayama station

As an example of the simulation results, Fig. 5 shows a comparison of the position of the rearmost stopped car agent in the first lane (where cars are permitted to turn left or go straight) on the north side road of the intersection. The mean, minimum, and maximum values of the positions that were measured once per signal cycle are indicated in the figure. A comparison between baseline_1 and baseline_2 shows that the existence of pedestrians significantly affects congestion levels, thereby also implying that the influence of pedestrians must be considered in those scenarios. In the comparison between baseline_2 and ext32, even though the average value is declining slightly, it can be seen that the change is sufficiently small as compared with the daily fluctuation level. In ext64, it is understood that car traffic is significantly affected by the tramway extension since the traffic jam length is becoming much longer.

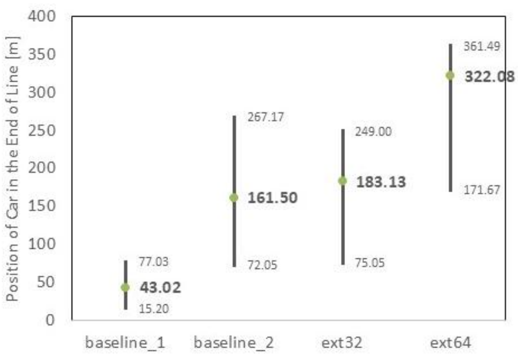


Fig. 5 Position of rearmost stopped car agent in the first lane of north-side road

2.5. Discussions

We developed a simulation framework for mixed traffic by implementing pedestrians and tram agents for the multi-agent-based traffic simulator. The results show that the layered road structure could be effectively reused in existing simulation models while maintaining their independence.

We simulated the tramway extension plan in front of Okayama Station with MATES, and quantitatively assessed its impact on road traffic. From the results, it could be determined that there was hardly any negative influence on car traffic patterns at the intersection in the case of ext32, in which the duration of the green light for the tram is 32 seconds × 1 time per cycle. In the case of ext64, in which the duration of green for the tram is 32 seconds × 2 times per cycle, car traffic on the north side road was noticeably degraded. We also checked to ensure that there are no remarkable changes on the west and south side roads of the intersection. Even though the simulation results in this research were obtained from a limited number of scenarios, they showed that this kind of simulation has the potential to provide objective data to stakeholders for use in discussions about transportation policies. Based the results with other considerations the Okayama-city has decided to extend the tramway into the station square, and it is now under construction.

3.Excretion Care Process Simulation

Specifically focusing on ”excretion care, which is regarded to be a particularly high burden on care workers, we newly construct a virtual nursing care process simulator based on a multi-agent model. As the first step, applying it to simulate care processes in an actual day-service facility, we have confirmed that the simulated results agree with observed ones reasonably well. Next, we evaluated the work efficiency of care workers quantitatively, parametrically varying the number of elderly people and that of care workers. As the results, it is confirmed that the total amount of assistance by care workers increases as the number of elderly people increases, and that the work time per care worker decreases as the number of care workers increases. We can conclude that although the developed multi-agent simulator is still limited to excretion care processes, such a simulation-based approach will be a powerful tool to discuss issues on nursing care processes quantitatively and to search for new solutions, by extending it to various situations in the nursing care processes.

Authors Introduction

Dr. Shinobu Yoshimura

He received Dr. Eng. in 1987 from the University of Tokyo (UTokyo). He is currently Emeritus Professor and Project Professor of Graduate School of Frontier Sciences in UTokyo. His research areas are high-performance and intelligent computational mechanics.

Dr. Hideki Fujii

He received Ph.D. in 2009 from UTokyo. He is currently Associate Professor of School of Engineering in UTokyo. His research areas high-performance social systems simulation based on a multi-agent approach.

Toward 2035: Renewable Energy Innovations Transforming Our Future

Firas Basim Ismail
Power Generation Research Centre, College of Engineering, Universiti Tenaga Nasional (UNITEN), Kajang, 43000, Malaysia

Email: Firas@uniten.edu.my

Abstract

This paper explores the transformative role of renewable energy innovations in addressing global challenges such as climate change and energy security. By 2035, significant milestones are expected to be achieved through targeted policies and technological advancements. Key innovations in solar technology, wind energy, bioenergy, and energy storage are discussed, along with the role of digitalization and AI in optimizing renewable energy systems. The paper also highlights global and regional efforts, including Malaysia's National Energy Transition Roadmap (NETR), and addresses challenges such as intermittency, high costs, and material safety. The conclusion emphasizes the importance of strategic investments and collaboration to ensure a sustainable energy future.

Keywords: renewable energy, solar technology, wind energy, bioenergy, energy storage, AI, Malaysia, NETR

1. Introduction

Renewable energy has become a cornerstone in the global effort to combat climate change and ensure energy security. As the world faces increasing environmental challenges, the transition to renewable energy sources is no longer optional but imperative. By 2035, significant milestones are expected to be achieved through targeted policies and technological innovations. This paper aims to provide a comprehensive overview of the key innovations in renewable energy, the role of digitalization and artificial intelligence (AI), and the global and regional efforts driving this transformation. Additionally, the paper will address the challenges that need to be overcome to ensure a seamless transition to renewable energy.

The urgency of this transition is underscored by global agreements such as the Paris Agreement (2016), which aims to limit global warming to below 1.5°C. Countries around the world are setting ambitious targets to increase the share of renewable energy in their energy mix. For example, the United States aims to achieve a carbon-free power sector by 2035, while China plans to expand renewable energy to account for 50% of its total power generation by the same year. In Southeast Asia, Malaysia has developed the National Energy Transition Roadmap (NETR) to decarbonize its electricity sector and promote renewable energy adoption.. Fig. 1 illustrates the Global Average Temperature from 1850 to 2023.

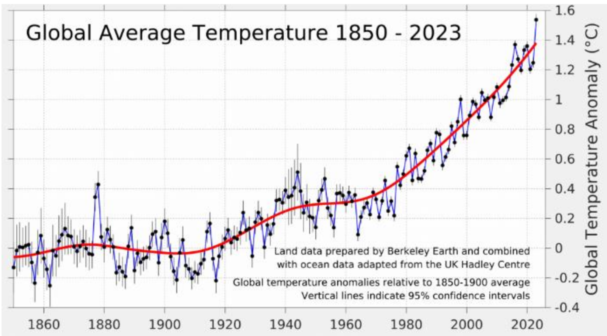


Fig.1 Global Average Temperature.

2. Key Innovations in Renewable Energy

2.1. Solar Technology

Solar energy continues to be one of the most promising renewable energy sources, with significant advancements in technology driving its adoption. One of the most notable innovations in recent years is the development of perovskite solar cells. These cells have shown higher efficiency (+8%) compared to conventional silicon-based solar cells, making them a game-changer in the solar industry. The world's first perovskite solar cell factory was established in Wroclaw, Poland, in 2021, marking a significant milestone in the commercialization of this technology. However, challenges remain in terms of stability, cost, and scalability, which need to be addressed before perovskite solar cells can be widely adopted [1]. Perovskite solar cells – higher efficiency (+8%) is shown in Fig. 2.

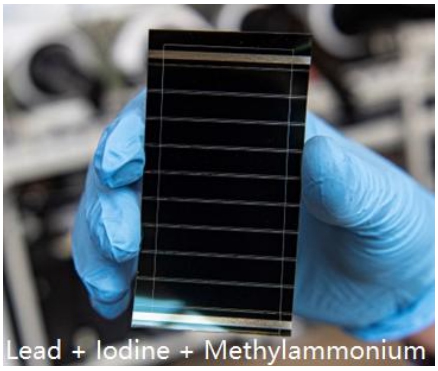


Fig.2 Perovskite solar cells.

Another promising innovation in solar technology is spectral splitting concentrated solar power (CSP) as shown in Fig.3. This hybrid technology combines CSP and photovoltaic (PV) systems to optimize electricity generation by splitting the solar spectrum. By doing so, it achieves a 10% increase in power output compared to conventional CSP systems. While still in the experimental stages, ongoing research is focused on improving the economic viability of this technology [2].

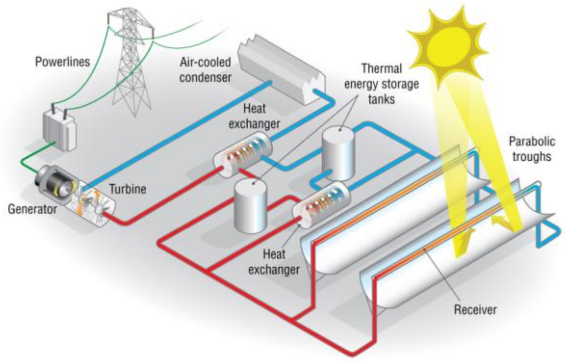


Fig. 3 Spectral splitting CSP.

2.2. Wind Energy

Wind energy has also seen significant advancements, particularly in turbine technology. The development of larger and more efficient wind turbines has dramatically increased energy generation capacity. For instance, the MySE 22 MW turbine, developed by MingYang Wind Power in China, features a rotor diameter of 310 meters, making it one of the largest and most powerful wind turbines in the world. This breakthrough in turbine technology has the potential to significantly reduce the cost of wind energy and increase its competitiveness with traditional energy sources [3]. analyzed.

2.3. Next-Generation Bioenergy

Bioenergy, particularly algae biofuels, is emerging as a promising alternative for sustainable energy production. Algae biofuels are derived from microalgae, which can be cultivated in large quantities and have a high lipid content, making them an excellent source of biofuel. While still in the research and development phase, pilot projects have demonstrated the feasibility of small-scale production. However, challenges such as high production costs and the need for large-scale cultivation facilities remain barriers to widespread adoption [4]. analyzed.

2.4. Energy Storage

Energy storage is a critical component of renewable energy systems, addressing the intermittency issues associated with solar and wind energy. Smart grid technologies have emerged as a key innovation in this area, facilitating communication between energy suppliers and consumers. These technologies integrate renewable energy sources through automation, sensors, and predictive analytics, enabling more efficient energy management. For example, Malaysia has implemented smart grid systems that integrate solar farms, electric vehicle (EV) charging stations, and battery storage systems, creating a more resilient and flexible energy grid [5].

3. The Role of Digitalization and AI

3.1. Applications of AI in Renewable Energy

Artificial intelligence (AI) is playing an increasingly important role in optimizing renewable energy systems. AI enhances the efficiency, reliability, and economic viability of renewable energy by improving resource assessment, energy forecasting, and system monitoring. Machine learning algorithms, deep learning techniques, and optimization methodologies are being used to predict energy generation patterns, optimize grid operations, and reduce maintenance costs. For example, AI-powered predictive maintenance systems can identify potential failures in wind turbines or solar panels before they occur, reducing downtime and improving overall system performance [6].

3.2. Status of AI Initiatives

Several AI initiatives are already underway in the renewable energy sector. In Malaysia, Tenaga Nasional Berhad (TNB) has implemented AI-driven technologies such as Digital Twin and IoT sensors for predictive maintenance and real-time analytics. These technologies enable TNB to monitor the condition of transformers and transmission lines more effectively, reducing the risk of power outages and improving grid reliability. Additionally, hyperscale data centers, such as the AirTrunk JHB1 in Johor Bahru, are utilizing renewable energy and advanced AI systems to optimize their operations, further demonstrating the potential of AI in the renewable energy sector [7]. Asset Performance Management System is shown Fig.4.

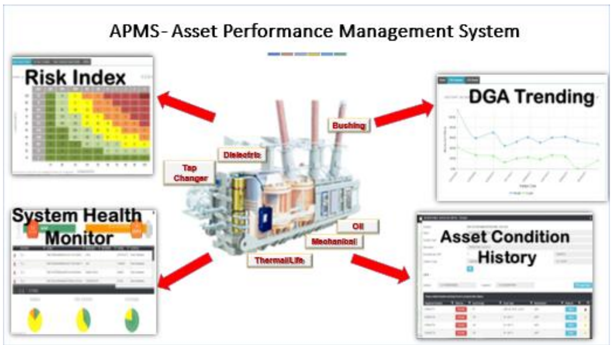


Fig. 4 Asset Performance Management System.

4. Renewable Energy Policies and Roadmap

4.1. Global Efforts

Global efforts to promote renewable energy are being driven by international agreements and national policies. The Paris Agreement (2016), signed by over 194 countries, aims to limit global warming to below 1.5°C by reducing greenhouse gas emissions. In Europe, the EU's REPowerEU Plan (2022) seeks to phase out Russian fossil fuels and accelerate the adoption of renewable energy. Similarly, China's 14th Five-Year Plan targets a 50% increase in renewable energy generation by 2025, reflecting the country's commitment to transitioning to a low-carbon economy [8].

4.2 Malaysia's National Energy Transition Roadmap (NETR)

Malaysia has developed the National Energy Transition Roadmap (NETR) to decarbonize its electricity sector and promote renewable energy adoption. The roadmap includes several flagship projects, such as the Hybrid Hydro Floating Solar (HHFS) initiative, which combines hydropower with floating solar panels. Feasibility studies for the first phase of this project have been completed, with operations planned to begin between 2025 and 2040. Additionally, Malaysia is developing large-scale solar parks with a combined capacity of 500 MW, in collaboration with local SMEs. Another key initiative is the green hydrogen and ammonia co-firing project, which aims to decarbonize power plants by using hydrogen and ammonia as alternative fuels. Feasibility studies for this project have been completed for key facilities such as the Jimah East Power plant [9].

5. Challenges and Limitations

Despite the significant progress in renewable energy, several challenges remain. One of the most pressing issues is the intermittency of renewable energy sources such as solar and wind, which are dependent on weather conditions. This intermittency affects the reliability of renewable energy systems and necessitates the development of advanced energy storage solutions. Additionally, geographic limitations restrict the feasibility of certain renewable technologies in regions with limited solar or wind resources.

High costs also pose a significant barrier to the widespread adoption of renewable energy. The initial investment required for renewable energy infrastructure, such as solar panels and wind turbines, remains substantial, particularly in developing countries. Furthermore, policy and collaboration gaps hinder the development of innovative solutions, as effective collaboration between policymakers, industries, and researchers is essential for driving progress in the renewable energy sector.

Finally, material safety concerns associated with certain renewable technologies, such as the use of toxic materials like lead in perovskite solar cells, pose environmental risks. Addressing these challenges will require continued research and development, as well as the implementation of robust safety regulations [10].

6. Success Stories

Several success stories highlight the potential of renewable energy to transform the global energy landscape. One notable example is the @DisCo Software, developed to enhance cybersecurity for energy utilities. Funded by the US Department of Energy (DOE) with a $4.5 million grant, this software uses machine learning to identify vulnerabilities in devices such as solar inverters, reducing detection time from months to hours. The software won the 2023 R&D World Award, underscoring its potential to improve the security and reliability of renewable energy systems (US Department of Energy, 2023).

In Malaysia, the 2010 National Renewable Energy Policy set a target of achieving 20% renewable energy in the energy mix by 2025. By 2023, Malaysia had already surpassed this goal, achieving a 25% renewable energy mix. This success is attributed to the implementation of large-scale solar projects, the adoption of smart grid technologies, and the development of innovative solutions such as hybrid hydro floating solar systems (Malaysia Ministry of Energy and Natural Resources, 2023).

7. Conclusion

The global shift toward renewable energy is not just essential for sustainability but also vital for energy security. Innovations in technology, supported by proactive policies and collaboration, are paving the way for significant milestones by 2035. However, addressing challenges such as intermittency, high costs, and material safety is crucial to ensure a seamless transition. With strategic investments and advancements in AI, renewable energy systems can achieve optimal efficiency and reliability, securing a sustainable future for generations to come.

References

1. Green, M. A., Ho-Baillie, A., & Snaith, H. J. (2014). The emergence of perovskite solar cells. Nature Photonics, 8(7), 506-514.
2. Zhang, X., Zhao, Y., & Wang, Z. (2021). Spectral splitting concentrated solar power: A review. Renewable and Sustainable Energy Reviews, 141, 110807.
3. Zhang, Y., Liu, W., Wang, X., & Gu, H. (2022). A Novel Wind Turbine Fault Diagnosis Method Based on Compressed Sensing and DTL-CNN. Renewable Energy, 194, 249-258.
4. Chisti, Y. (2007). Biodiesel from microalgae. Biotechnology Advances, 25(3), 294-306.
5. Abdullah, M. A., Hassan, M. K., & Ismail, F. B. (2016). A Review on Comparison of Technologies and Progress of a Smart Grid Development in Malaysia and Singapore. Renewable and Sustainable Energy Reviews, 59, 419–428. doi:10.1016/j.rser.2016.01.041.
6. European Commission. (2022). REPowerEU: A plan to rapidly reduce dependence on Russian fossil fuels and fast forward the green transition. European Commission.
7. https://www.efchina.org/Blog-en/blog-20220905-en.
8. Malaysia Ministry of Energy and Natural Resources. (2021). National Energy Policy 2022-2040. Malaysia Ministry of Energy and Natural Resources.
9. US Department of Energy. (2021). Solar Energy Technologies Office: Perovskite Solar Cells. US Department of Energy.
10. International Renewable Energy Agency (IRENA). (2021). Renewable Energy Statistics 2021. International Renewable Energy Agency.

Author Introduction

Dr. Firas Basim Ismail

He is an Associate Professor at Universiti Tenaga Nasional (UNITEN), Malaysia, specializing in renewable energy and sustainable technologies. With a strong academic background and extensive research experience, Dr. Ismail has contributed significantly to the field of renewable energy, particularly in solar technology, wind energy, and bioenergy. His work focuses on advancing innovative solutions to address global challenges such as climate change and energy security. Dr. Ismail has been actively involved in various national and international projects, including Malaysia's National Energy Transition Roadmap (NETR), where he has played a key role in promoting renewable energy adoption and decarbonization strategies. He is also a sought-after speaker at international conferences, sharing insights on the latest advancements in renewable energy technologies and policies. In addition to his research and teaching responsibilities, Dr. Ismail is committed to fostering collaboration between academia, industry, and policymakers to drive the global transition toward a sustainable energy future.

Integration of Human-Device Interface: Transforming the Future of Interaction

Norrima Mokhtar
Department of Electrical Engineering, Universiti Malaya 50603 Kuala Lumpur, Malaysia

Takao Ito
Graduate School of Advanced Science and Engineering, Hiroshima University, Japan
Email: norrimamokhtar@um.edu.my, itotakao@hiroshima-u.ac.jp

Abstract

The integration of Human-Device Interfaces (HDIs) is revolutionizing the way human interact with technology for daily life, paving the way for transformative advancements across industries such as healthcare, communication, education, entertainment, gaming, industry and automation. Brain-computer Interfaces (BCIs), wearable devices, gestures, speech recognition, haptic feedback and tactiles interfaces, web-based interfaces and mobile applications become the new normal in massive integration of HDIs. Despite the positive usage of HDIs, these advancements come with significant challenges including data security, reliability, durability, safety, ethics, accessibility, costs, design and technology that either to be accepted or rejected by certain people. These challenges need to be addressed and managed globally for further development and robust regulations to avoid misuse of technology.

Keywords: User experience, Human-Device Interface, Future Technology

1. Introduction

Rapid technology direction has paved the way for seamless interaction between humans and machines, adding more usability, efficiency and overall experience [1], [2], [3], [4], [5]. Traditionally, inputs such as keyboards, buttons and touch screens dominate user interactions[6]. However, new technologies like gesture recognition, eye tracking, speech recognition and brain computer interfaces (BCIs) are the next direction of how humans interact with devices. This direction of technology making the interfaces requires less effort, more responsive and enhances user experience.

In particular to Human-Device Interfaces (HDIs), Artificial Intelligence (AI), either supervised, semi-supervised and unsupervised learning algorithms has become huge and popular in HDIs development. AI enables systems to process input and respond to input data which allows for real time decision making and system automation. These advancements not only enhance accessibility for individuals with disabilities but also open experience across diverse fields, including healthcare, education, entertainment, automation and gaming industry.

For HDIs, the development of contactless and effortless input approached has become popular in recently [7]. Eye movement tracking using electrooculography(EOG) or wearable camera have enabled hands free device control hence contributes to significant achievement in medical and automation industries [7], [8], [9]. In this sense, Brain-Device Interfaces (BDIs) or Brain Machine Interfaces (BMIs) allow mental commands, redefining user-device collaboration and more importantly user experience. The aim of this paper is to explore current advancement, example of case studies conducted by ACRLAB, Faculty of Engineering, Universiti of Malaya, Malaysia.

2. ACRLAB, Faculty of Engineering, University Malaya

2.1. Aim and Mission

ACRLAB, Faculty of Engineering aimed to contribute to HDIs technology advancement, effortless interfaces between human, devices and systems to enhance user experience. The facilities available will be discussed under the case studies section. The location of ACRLAB, Faculty of Engineering, Universiti Malaya, national and international collaborations are shown below in (Fig. 1).

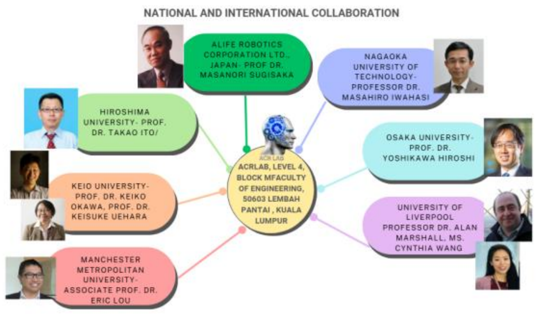


Fig.1 ACRLAB location with national and international collaborators.

2.2. *The importance of Human-Device Interfaces (HDIs)*

HDIs is set to revolutionize how we live, work, drive, travel, exercise, sleep, shopping and interact with technology. Eventually, technological advancements have rapidly progressed, moving from basic tools like keyboards and mice to more sophisticated, interconnected devices such as smartphones, wearable technology (Fig. 2), [6], [7] and AI driven assistants. This shift is fundamentally changing how we engage with technology and without realizing we are so dependent on it, integrating into our everyday routines and surroundings. The HDIs also enable adaptability and personalization of devices according to our need, hence enhancing user experience and preferences.

Fig. 2. BCIs using Electroocoulogram for wheelchair navigation by ACRLAB, Universiti Malaya [6], [7].

3. Recent Works on HDIs

3.1. *Wearable Technology*

Major developments include smartwatches, health monitor, voice and gesture-based interfaces such as SIRI, google assistant, AR/VR systems and Brain Computer Interfaces (BCIs) [9] are offering possibilities in many industries such as healthcare, gaming, shopping, driving, monitoring and others. There are two examples of these technology that was implemented in ACRLAB, Faculty of Engineering, Universiti Malaya as shown in Fig. 2 and Fig. 3, [6], [7].

Fig.3 Eye driven wheel chair by ACRLAB, Universiti Malaya

Fig. 2 shows BCIs using Electrooculogram for wheel chair navigation by ACRLAB, Universiti Malaya. Fig. 3, [6], [7] shows eye driven wheelchair using wearable camera for navigation. Both works shown the proof on concept for possibilities enhancing user experience and effortless interface for the disabled communities to be more independent and expressing their need.

Fig. 4 shows the modes and state for BCIs navigation by Fig. 2, [6], [7].

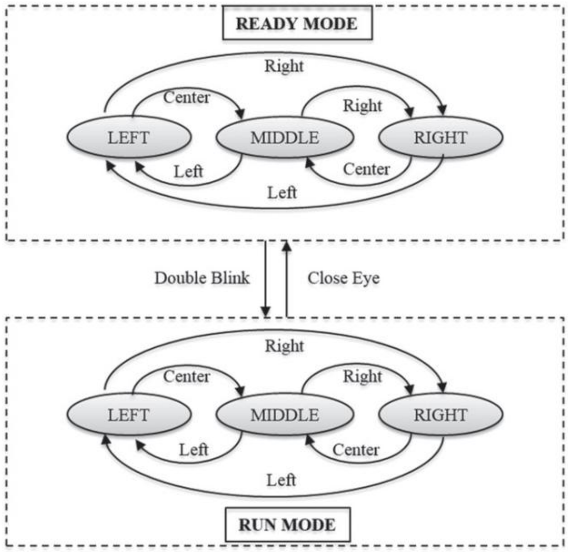


Fig. 4. The modes and state BCIs navigation by Fig. 2, [6], [7]

Other commercially available wearable devices include Apple Watch, Fitbit, Garmin and Samsung Galaxy Watch which are popular for tracking fitness, health metrics and facilitating communication. Apple watch for example, can detect irregular heart rhythms and perform ECG tests, potentially identifying early signs of heart disease. Additionally, health monitoring devices such as Dexcom G6 and Abbot Freestyle Libre, can continuously monitor blood glucose levels and sends data to a smartphone or compatible receiver every five minutes. Timely diabetes management, improving patient wellbeing and risk of complications are among countless benefits that HDIs can offer.

3.2. *Voice and Gesture-Based Interfaces*

Voice assistants like Siri, Google Assistant and Alexa are becoming smarter thanks to natural language processing (NLP) and machine learning. These interfaces simplify tasks such as setting reminders, controlling smart home devices and retrieving information. Gesture control like Microsoft Kinect or motion sensors in smartphones enable gesture-based interactions, enhancing user experience for gaming or other means to control other systems effortlessly. Fig. 5 shows interactive design with gesture and voice recognition for virtual teaching environments [5]. Augmented Reality (AR) and Virtual Reality (VR) are usually integrated with gesture, voice and adaptive virtual environments to enhance user experience [5].

Fig. 5. Gesture and voice recognition for virtual teaching environments [5].

3.3. *Haptic Feedback and Tactile Interfaces*

Haptic Feedback and Tactile Interfaces provide sensory feedback through touch, simulating textures or resistance. Haptic sensors depend on the mechanisms such as piezoresistive, capacitive, piezoelectric and triboelectric with respect to functionalities in force sensing, gesture translation and touch identification. Fig. 6 shows haptic sensing and feedback techniques [10].

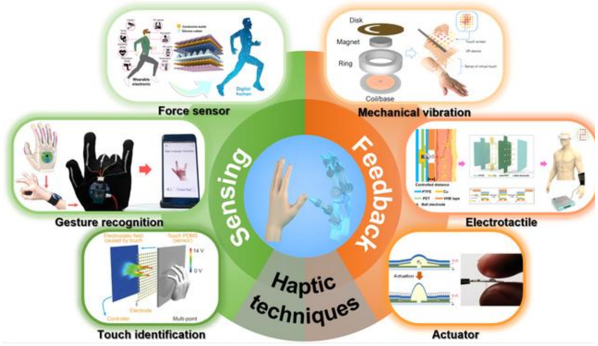


Fig. 6. Haptic sensing and feedback techniques [10]

3.4. *Web-Based HDIs*

Remote monitoring systems allow flexibility to interact with medical devices, wearable devices, education, robotics, industries, gaming and entertainment. This technology has become the new normal and essential for user interfaces and services which can be done via mobile applications and web-based HDIs. Fig. 7 and Fig. 8 shows drone location and virtual maps for potential usage in integrated monitoring and wildfire mitigation developed by ACRLAB, Faculty of Engineering, Universiti Malaya. This work combines machine learning, vision-based system, GPS, control system to stabilize the drone flight and Internet of Things Technology to enable remote monitoring, streaming data of interests online via web-based applications. The same concept can be applied to continuously monitor patient health status, by streaming the data to medical officers.

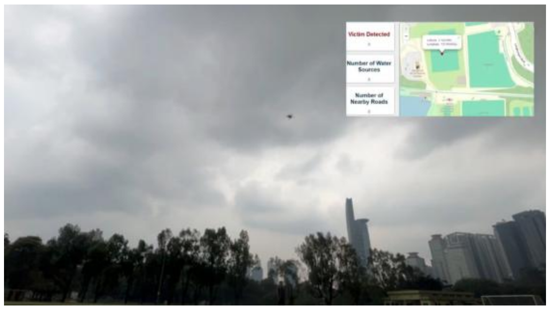


Fig. 7. Drone and web-based interface (ACRLAB, Faculty of Engineering, Universiti Malaya)

Fig. 8. Illustration of drone and web-based interface

4. Ethical Considerations and Challenges in Human-Devices Interfaces (HDIs)

HDIs evolve and become an integral part of daily life, which several ethical considerations, challenges, data privacy, data security, side effects in term wearable devices, physiological and psychological effect both human, adults and children should be taken into our major concern. There is a need to draw the line between enhancement and alteration, augmenting human abilities about the implications of changing natural human traits and the potential societal consequences. The dilemma is the use for good and betterment of society or the other way [10], [11], [12], [13], [14], [15].

5. Challenges in HDIs Integration

Further concern in HDIs integration consists of hardware limitations, system compatibility, reliability (error and failure rates), safety, public perception, trust issues(misuse of personal data) and regulation which ultimately determined by government decisions. For instance, wearable devices that rely on sensors to monitor vital signs or BCIs that interpret brain activity may not be capable of delivering high accuracy and performance without large, bulky components or frequent battery charging. The other important factor is the development costs to support this advanced and sophisticated HDIs technology which depends on the community's financial sustainability [10], [11], [12], [13], [14], [15].

Notably public perception is also crucial in HDIs integration. Some people are wary of wearable devices that integrate closely into the body. This can be invasive or non-invasive HDIs.

In this context, although HDIs play major roles in managing our daily life, which cannot be avoided as the world and future direction are towards digitization, the is a need for comprehensive regulations, ethical standards, accountability and global cooperation for these emerging issues.

6. Conclusions

In conclusion, the integration of Human-Device Interfaces (HDIs) holds immense potential to transform various industries and become an essential part of daily life. However, the associated challenges, from data security and reliability to ethical considerations and accessibility, highlight the need for comprehensive global efforts to address these issues. By developing robust regulations, ethical practices, and ensuring inclusive design, society can fully benefit HDIs while mitigating risks, for responsible and sustainable technological advancements.

References

1. Rubio-Drosdov E, Díaz-Sánchez D, Almenárez F, Arias-Cabarcos P, Marín A. Seamless human-device interaction in the internet of things. IEEE Transactions on Consumer Electronics. 2017 Nov;63(4):490-8.
2. Söldner R, Rheinländer S, Meyer T, Olszowy M, Austerjost J. Human–device interaction in the life science laboratory. InSmart Biolabs of the Future 2022 Jan 29 (pp. 83-113). Cham: Springer International Publishing.
3. Jaber W, Jaber HA, Jaber R, Saleh Z. The convergence of AI and BCIs: A new era of brain-machine interfaces. Artificial Intelligence in the Age of Nanotechnology. 2024:98-113.
4. Mazhar N, Gaballa A, Pandey AK, Cabibihan JJ. Virtual Reality Hand Tracking for Immersive Telepresence in Rehabilitative Serious Gaming. InInternational Conference on Social Robotics 2023 Dec 3 (pp. 3-12). Singapore: Springer Nature Singapore.
5. Fang K, Wang J. Interactive Design with Gesture and Voice Recognition in Virtual Teaching Environments. IEEE Access. 2024 Jan 1.
6. Aziz F, Arof H, Mokhtar N, Mubin M. HMM based automated wheelchair navigation using EOG traces in EEG. Journal of neural engineering. 2014 Sep 4;11(5):056018.
7. Ramli R, Arof H, Ibrahim F, Mokhtar N, Idris MY. Using finite state machine and a hybrid of EEG signal and EOG artifacts for an asynchronous wheelchair navigation. Expert Systems with Applications. 2015 Apr 1;42(5):2451-63.
8. Casadio M, Ranganathan R, Mussa-Ivaldi FA. The body-machine interface: a new perspective on an old theme. Journal of Motor behavior. 2012 Nov 1;44(6):419-33.
9. Takara S, Kida H, Inoue T. Development of implantable devices for epilepsy: research with cats, dogs, and macaques in biomedical engineering. Advanced Robotics. 2024 May 14:1-25.
10. Shi Y, Shen G. Haptic Sensing and Feedback Techniques toward Virtual Reality. Research. 2024 Mar 23;7:0333.
11. Şahinel D, Akpolat C, Görür OC, Sivrikaya F, Albayrak S. Human modeling and interaction in cyber-physical systems: a reference framework. Journal of Manufacturing Systems. 2021 Apr 1;59:367-85.
12. Farias da Costa VC, Oliveira L, de Souza J. Internet of everything (IoE) taxonomies: A survey and a novel knowledge-based taxonomy. Sensors. 2021 Jan 14;21(2):568.
13. Sun T, Tsaava T, Peragine J, Crosfield C, Lopez MF, Modi R, Sharma R, Li C, Sohal H, Chang EH, Rieth L. Flexible IrOx neural electrode for mouse vagus nerve stimulation. Acta biomaterialia. 2023 Mar 15;159:394-409.
14. Pöhn D, Hommel W. Towards an Improved Taxonomy of Attacks Related to Digital Identities and Identity Management Systems. Security and Communication Networks. 2023;2023(1):5573310.
15. Wirtz BW. Artificial intelligence, big data, cloud computing, and Internet of Things. InDigital Government: Strategy, Government Models and Technology 2022 Oct 8 (pp. 175-245). Cham: Springer International Publishing.

Authors Introduction

Associate Prof. Ir. Dr. Norrima Mokhtar

She earned a Bachelor of Engineering (B.Eng) in Telecommunication Engineering from Universiti Malaya in 2000. She was awarded the Panasonic Scholarship to pursue her Master of Engineering (M.Eng) at Oita University, Japan (2003-2006). Additionally, she received the SLAB/SLAI scholarship to complete her Ph.D. in Electrical Engineering at Universiti Malaya (2008-2012). Between 2000 and 2002, she worked as a Telecommunication Engineer at Echobroadband Sdn. Bhd., where she contributed to upgrading cable TV networks to hybrid fiber coaxial networks in Köln, Germany. Since 2003, she has built a distinguished academic career at Universiti Malaya, where she currently serves as an Associate Professor.

Professor Dr. Takao Ito

He received his M.S., and Ph.D. He is Professor of Management of Technology (MoT) in Graduate School of Engineering at Hiroshima University. He is serving concurrently as Professor of Harbin Institute of Technology (Weihai) China. He has published numerous papers in refereed journals and proceedings, particularly in the area of management science, and computer science. He has published more than eight academic books including a book on Network Organizations and Information (Japanese Edition). His current research interests include automata theory, artificial intelligence, systems control, quantitative analysis of inter-firm relationships using graph theory, and engineering approach of organizational structures using complex systems theory.

Non-Invasive Glucose Monitoring Based on Mid-Infrared Spectroscopy

Puteri Nur Sofea Mohd Zakki, Nani Fadzlina Naim, Hasnida Saad, Wan Norsyafizan Wan Muhamad, Suzi Seroja Sarnin, Norsuzila Ya'acob
Universiti Teknologi MARA (UITM), 40450 Shah Alam, Selangor.

Noor Fitrah Abu Bakar
Universiti Teknologi MARA (UITM), 40450 Shah Alam, Selangor.

Email: sofeazakki@gmail.com, nanifadzlina@uitm.edu.my, hasnidasaad@uitm.edu.my, syafizan@uitm.edu.my, norsuzilayaacob@uitm.edu.my, fitrah@uitm.edu.my

Abstract

This paper presents on non-invasive method of glucose monitoring using Mid-Infrared (M-IR) spectroscopy. Glucose samples are prepared and analyzed using M-IR spectroscopy. Using Fourier-transform of the M-IR spectroscopy, we experimentally track variations in the mid-infrared glucose absorption peak. The glucose samples and the relation with diabetic people is also presented in this paper. It is found that as the glucose concentration increases, the wavelengths at which absorbance peaks occur also increase particularly for wavelength range 1400-1470nm.

Keywords: glucose, distilled water, FTIR spectroscopy, absorbance

1. Introduction

Researchers have been working to develop diabetic self-care assessment techniques over time. Many of these therapies call for a blood sample. The number of diabetics has been gradually increasing in recent years due to changes in individual behavior. We place a great deal of importance on it. The 140 million people who require insulin injections globally produce very little insulin on their own. It is referred to as the insulin therapy [1]. They must keep an eye on their own blood sugar levels. They must specifically monitor blood glucose levels all day long. Plus, they required an insulin injection if their blood sugar level increased [1][2].

However, today's technologies for measuring glucose all require patients to prick their fingers and take a blood sample. It causes pain for diabetics and increases their chance of getting an infection [3]. These will thus make the burden heavier.

Since people avoid taking measurements with the uncomfortable finger-stick glucose monitors that are currently available, the development of non-invasive (or minimally invasive) glucose monitoring technologies is inevitable. The most recent efforts made in this area were thoroughly discussed in several review papers [4]. Among the many non-invasive (NI) glucose detection methods, optical sensing has received the most attention due to some potential benefits in terms of sensitivity, response time, and patient comfort.

Several researchers have shown that FT-IR spectroscopy may be utilized to detect glucose in the MIR range [5]. Hardly a few research teams have adapted MIR transmission measurement using FT-IR spectroscopy due to blood's viscosity, high particle content, and strong water background absorption.

In [6], diffrerent concentration of ethanol solution is obtained. Using Agilent Cary 630 Fourier Transform Infrared (FTIR) spectroscopy, the absorbance of various concentrations of ethanol solution was measured.

Researchers in [7] deployed A Fourier transform infrared spectroscopic method with attenuated total reflectance (FTIR-ATR) for the prediction of sugar contents in honey samples.

The study of the interaction between matter's electromagnetic radiation in the infrared (IR) range is known as infrared (IR) spectroscopy. Additionally, this spectroscopy has long been a powerful resource for recognizing organic samples, even those in complicated forms. One of the improvements in IR spectroscopy where this method is more widely used in the research of various areas is the Fourier Transform Infrared (FTIR) spectroscopy.

There are several benefits to performing FTIR spectroscopy since it is considered to be a straightforward approach that only needs little sample preparation and can produce immediate results. The FTIR spectroscopy also has a high sensitivity or spectral signal to noise (S/N) ratio [8], which enables the detection of the component in glucose even at extremely low concentrations. It is clear from the reported results that this method is appropriate for determining blood glucose levels [8].

The Beer's Lambert Law is a law which governs the relationship between the absorbance and the concentration of samples. This law states that there is a liner relationship between the absorbance and their concentration [6]. The general equation is written as:

$$A = \varepsilon l c \qquad (1)$$

Where:
A = Absorbance of light
ε = Wavelength-dependent molar absorptivity with coefficient (M-1cm-1)
l = Length of solutions this light passes through (cm)
c= Concentration of solution (v/v%)

Wavenumber formula:

$$W = 1/\lambda \qquad (2)$$

Where:
W= wavenumber in cm^{-1}.
λ= wavelength in nm.

The range of the IR bands covers from 70nm to 1000µm. The IR region is divided into three sub-regions, namely the Near-IR region (700-2500nm), the middle-IR region (2500-25000nm) and Far-IR region (2500-1000µm). In this study, we investigate non-invasive blood glucose testing with infrared radiation to minimize the burden on diabetics. FTIR equipment operates at the Mid-IR region is used for this project [9]. However, due to project limitations, blood glucose solution is replaced with glucose solution. Various glucose concentration represents various diabetic conditions as shown in Table 1.

2. Research Methodology

In this experiment, different concentrations of glucose were prepared by adding the mass of the glucose and diluting the mass in 50 mL of distilled water at room temperature. The prepared concentrations are depicted as in Table 1. Fig. 1 shows the weighting scale and the beaker used to produce the glucose solution.

Fig. 1. Chemical weighting scale and beaker

The highest concentration (64,000 mg/dl=3.2 g in 50 ml) of the six various concentrations of aqueous glucose solution made in distilled water. Sample 1 is the highest concentration is at this level which indicates a very high risk diabetic person. Sample 6 represents a low glucose concentration and it is considered high risk due to hypoglycemia condition. Table 1 shows the aqueous solutions samples with different concentration. Each sample represents different characteristics of blood glucose for diabetic people. Fig. 2 displays the schematic diagram of the FTIR spectroscopy that is connected to a personal computer for further analysis.

Table 1. Aqueous Solutions Sample with Different Concentration and Characteristics

Samples	Characteristic of concentration	In mg/dl	In 50 ml	Risk
Sample 1	Very high concentration	6400	3.2g	Very high
Sample 2	Hyperglycaemic blood	540	0.27g	High
Sample 3	Clinical accuracy	400	0.2g	No risk
Sample 4	Normal human blood	160	0.08g	No risk
Sample 5	Normal human blood	70	0.035g	No risk
Sample 6	Low concentration	30	0.015g	High

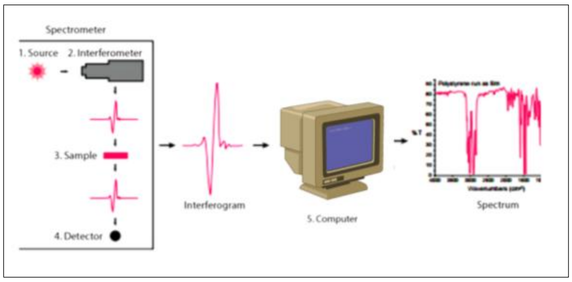


Fig. 2. Schematic diagram of the FTIR spectroscopy

The following weight of aqueous glucose solution are prepared: 3.2g, 0.27g, 0.2g, 0.08g, 0.035g and 0.015g. First and foremost, study into the various glucose concentrations must be conducted by reviewing earlier research papers. This procedure must be followed to guarantee that the glucose measurement is based on optical absorbance and that the result is nearly identical to the blood concentration reading obtained using a glucometer.

Following that, data analysis can be carried out by varying the optical absorbance intensity in distilled water for various concentration values. The relationship between absorbance intensity and glucose concentrations can then be plotted using Excel for all samples after receiving the Cary 630 FTIR spectroscopy results. Fig. 3 shows the Cary 630 FTIR Spectroscopy used in this experiment.

By examining the absorbance on the selected peaks, it is then possible to validate the data with regard to its various peaks at various concentrations. This makes it possible to confirm the validity of the experiment and the accuracy of the results at all concentrations.

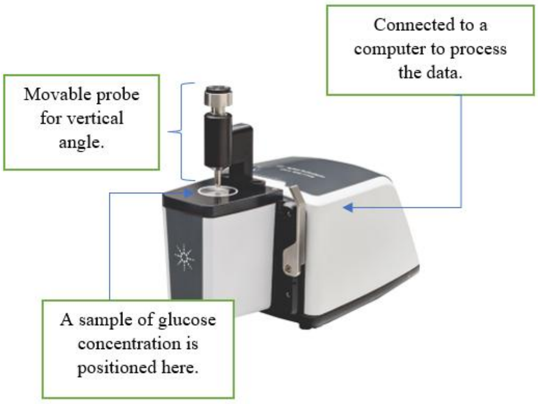


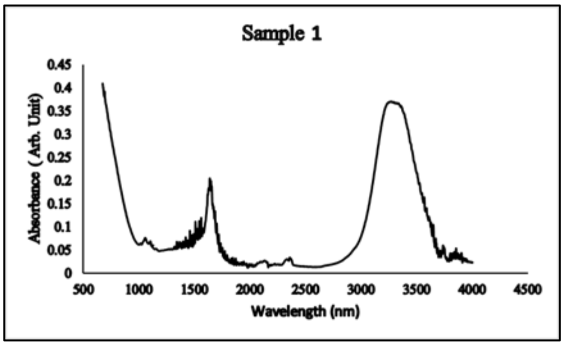
Fig. 3. Cary 630 FTIR Spectroscopy

The Cary 630 FTIR instrument needs to be connected to the computer that has its data processing software loaded before it can measure the optical of glucose. The glucose concentration sample needs to be set up on the platform as shown in the figure in order to begin the measuring process. The instrument's platform needs to be cleaned beforehand since it will affect the undesirable elements that are captured during data collection. The probe accessory for liquid samples can be adjusted up and down without contacting the sample. Typically, a solvent of the chemical in which the researcher is interested is used as the background sample. Air serves as another backdrop example. When the platform detects no background sample, it is automatically set.

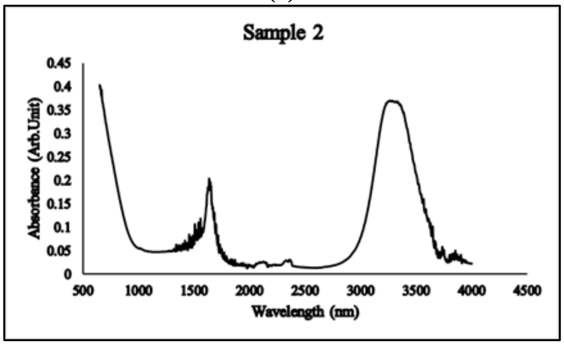
Once the background sample is recognized, the glucose sample can be placed on the same platform for the measurement process by dropping a few drops of the sample. However, the platform must be cleaned again before placing another concentration to make sure the measurement is not mixed with other concentration. The spectroscopy will automatically remove the background data from the testing sample. Therefore, the researcher will obtain the measurement results without the solvent of the glucose.

3. Result and Discussion

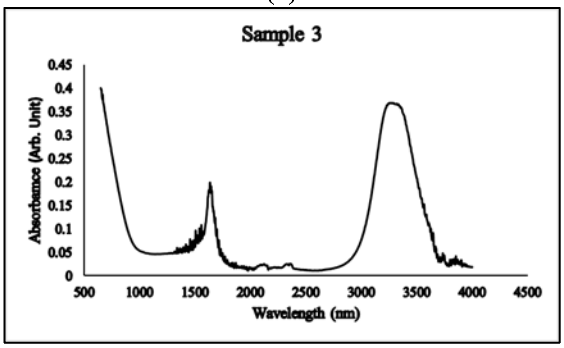
Using MIR-spectroscopy, the absorbance of the samples are obtained. Fig. 4 shows the graph of each sample with various absorbance intensity and glucose concentration. Each sample has different peak value, whereas the sample with the highest concentration having the highest peak value.



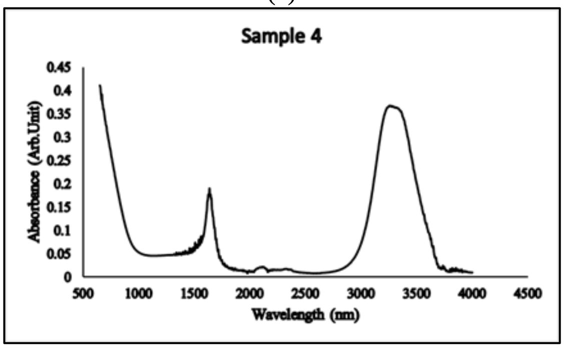
(a)



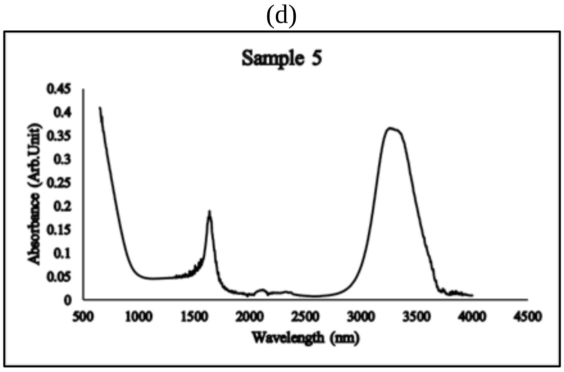
(b)



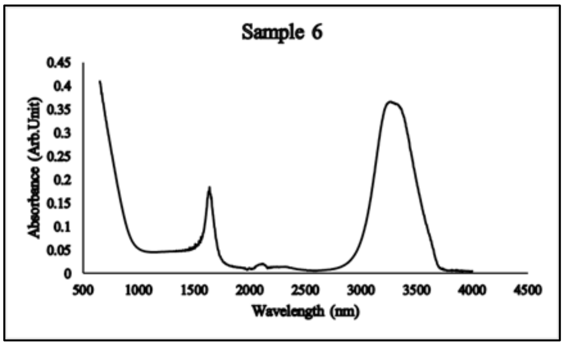
(c)



(d)



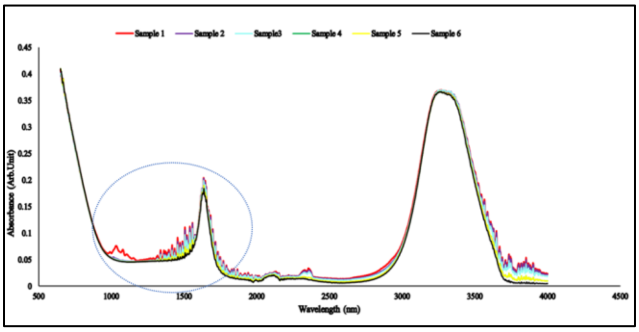
(e)



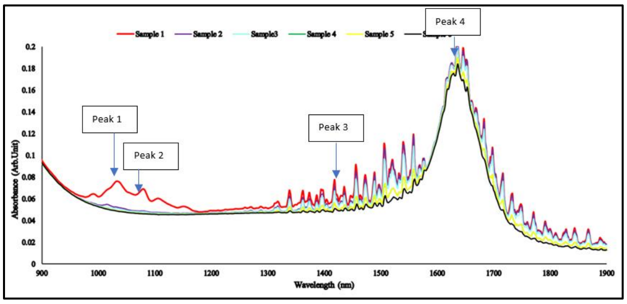
(f)

Fig. 4. Relationship between the absorbance intensity and glucose concentration for (a) 3.2g, (b) 0.27g, (c) 0.2g, (c) 0.08g, (d) 0.035g & (d) 0.015g

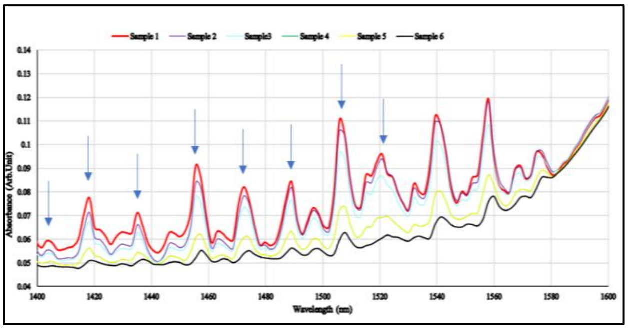
When all the graphs for all the samples have been obtained, the data must then be merged to analyze the peaks for various aqueous glucose solutions. Fig. 5 (a) shows the combination of all spectra and the dotted circles shows the spectra that will be analyzed as it shows significance difference of the samples. Fig. 5 (b) shows the zoom in spectra that will be analyzed. Four peaks are identified and can be further analyzed. Fig. 5 (c) shows the close up look of Peak 3. The wavelength ranges for Peak 3 is from 1400 nm until 1530 nm.



(a)



(b)



(c)

Fig. 5. (a) Absorbance spectra of combination of all glucose samples, (b) Zoom in the peaks spectra of the glucose concentration, (c) Spectra of Peak 3 with close-up look

Fig. 6 shows the absorption change of small peaks in broader Peak 3, which are marked as arrows in Fig. 6 from left to right designed as Peak3-1, Peak3-2, Peak3-3, Peak3-4, and Peak3-5. Each peak corresponding to the absorption wavelength of 1400nm, 1420nm, 1440nm, 1460m, and 1470nm, respectively. It can be observed that the higher the concentration, the higher the absorption power is. Table 2 shows the concentration and the absorbance for all samples. The findings demonstrate that the -CHO bond present in the FTIR spectrum transmission spectrum from glucose-water solution at different concentrations can be used to identify the glucose component.

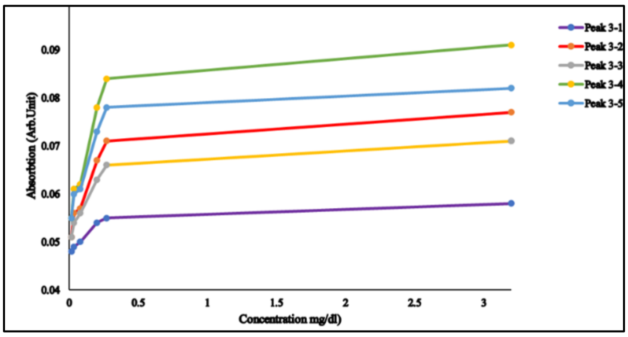


Fig. 6. The change of selected Peak 3 as function of glucose concentration

Table 2. Absorbance Intensities Of Glucose For Various Concentration

Samples	Total Absorbance Intensity (Au)
Sample 1	175.3459
Sample 2	171.3098
Sample 3	168.0368
Sample 4	159.3074
Sample 5	159.3074
Sample 6	154.7048

4. Conclusion

A non-invasice glucose monitoring has been successfully demonstrated. It demonstrates that the -CHO bond present in the FTIR spectrum transmission spectrum from glucose-water solution at different concentrations can be used to identify the glucose component. It is found that as the glucose concentration increases, the wavelengths at which absorbance peaks occur also increase particularly for wavelength range 1400-1470nm. The outcomes demonstrated that various glucose-specific peaks had various absorption values; therefore, by measuring these values, we can establish the correct correlations for the concentration-to-absorption relation, which can be applied for a high-precision glucose measurement.

References

1. K. Fujita et al., "Noninvasive Measurement of Blood Glucose Based on Optical Sensing and Internal Standard Method," 2005 IEEE Instrumentationand Measurement Technology Conference Proceedings, Ottawa, ON, Canada, 2005, pp. 1433-1437, doi: 10.1109/IMTC.2005.1604387.
2. S. Koyama, Y. Miyauchi, T. Horiguchi and H. Ishizawa, "Non-invasive measurement of blood glucose of diabetic based on IR spectroscopy," Proceedings of SICE Annual Conference 2010, Taipei, Taiwan, 2010, pp. 3425-3426.
3. Y. Chen, J. Liu, Z. Pan and S. Shimamoto, "Non-invasive Blood Glucose Measurement Based on mid-Infrared Spectroscopy," 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 2020, pp. 1-5, doi: 10.1109/CCNC46108.2020.9045160.
4. D. -H. Kim, I. K. Ilev and J. U. Kang, "Using Mid-Infrared Glucose Absorption Peak Changes for High-Precision Glucose Detection," LEOS 2007 - IEEE Lasers and Electro-Optics Society Annual Meeting Conference Proceedings, Lake Buena Vista, FL, USA, 2007, pp. 226-227, doi: 10.1109/LEOS.2007.4382359.
5. S. Liakat, K. A. Bors, T. -Y. Huang, A. P. M. Michel, E. Zanghi and C. F. Gmachl, "In vitro measurements of physiological glucose concentrations in biological fluids using mid-infrared spectroscopy," CLEO: 2013, San Jose, CA, USA, 2013, pp. 1-2, doi: 10.1364/CLEO_SI.2013.CW1L.2.
6. Saad, H., Zolkapli, M. and Rahman, H.A. (2019) "Quantification analysis of ethanol mixtures using Cary 630 FTIR spectrometer," 2019 IEEE 9th International Conference on System Engineering and Technology (ICSET) [Preprint]. Available at: https://doi.org/10.1109/icsengt.2019.8906377.
7. Ofélia Anjos, Maria Graça Campos, Pablo Contreras Ruiz, Paulo Antunes, "Application of FTIR-ATR spectroscopy to the quantification of sugar in honey,"Food Chemistry, 2015, pp. 218-223
8. K. Yang, Weizhong Tan, Xincong Zhou and Chao Wang, "Validation on relationship between the absorption spectrum of mid-infrared and water content in mechanical oil," Proceedings of the IEEE 2012 Prognostics and System Health Management Conference (PHM-2012 Beijing), Beijing, 2012, pp. 1-5, doi: 10.1109/PHM.2012.6228968.
9. K. V. Pozhar, E. L. Litinskaia, M. O. Mikhailov and E. A. Polyakova, "Reflection Near-infrared Spectroscopic Method For Non-invasive Blood Glucose Measurement," 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), St. Petersburg and Moscow, Russia, 2020, pp. 2524-2527, doi: 10.1109/EIConRus49466.2020.9039005.

Authors Introduction

Puteri Nur Sofea Binti Mohd Zakki

She is a fresh graduates (BE) Degree in Electronic Engineering majoring in Communication at Universiti Teknologi Mara UiTM Shah Alam, Malaysia. Her interests include Radio Frequency Engineering & Networking.

Ir Ts Dr Nani Fadzlina Naim

She received her B.Eng in Electrical-Electronics and M.Eng in Electrical (Electronics and Telecommunication) from Universiti Teknologi Malaysia in 2005 and 2007, respectively. She received her PhD in Optical Communications from Universiti Kebangsaan Malaysia in 2015. Currently, she is working as a Senior Lecturer at School of Electrical Engineering, Universiti Teknologi MARA (UiTM). Her research interests include Optical Communications, Optical Fiber Sensor, Fiber Bragg Grating Based Sensor and spectroscopy.

Researcher ID Link: Nani Fadzlina Naim - Google Scholar

Dr. Hasnida Bt Saad

She is a senior lecturer from the School of Electrical Engineering at the College of Engineering, Universiti Teknologi MARA, Malaysia. She received her Ph.D. in Electrical Engineering from Universiti Teknologi MARA in 2018. Her research interests include spectroscopy studies and optical sensors.

Dr Wan Norsyafizan Wan Muhamad

She received her Bachelor in Electrical Engineering from the Universiti Malaya, Malaysia in 2002. She completed her Master of Electrical Engineering from Universiti Malaya, Malaysia in 2009. She obtained her Ph. D in 2017 from University of Newcastle, Australia. Currently, she is a Senior Lecturer in Faculty of Electrical Engineering at Universiti Teknologi MARA, Shah Alam, Malaysia. Her current research interests are in the area of wireless communication (Physical and MAC cross layer optimization).

Ir Ts Dr Suzi Seroja Sarnin

Suzi Seroja Sarnin received a Bachelor of Electrical and Electronics degree from Universiti Teknologi Malaysia in 1999. After receiving her degree, she worked for two years (1999–2001) as a quality control engineer at Memory Tech (M) Sdn. Bhd. In March 2001, she started working as a contract lecturer at Universiti Teknologi MARA. She then continued her studies at University Kebangsaan Malaysia for an Msc in Microelectronics, obtaining her degree in 2005. She pursued a PhD in Electrical Engineering, which she completed in 2018. She continued to work for Universiti Teknologi MARA as a Senior Lecturer at the Department of Electrical Engineering after obtaining her PhD. Her research interest includes Wireless Communication, Multiple Access, Space Time Coding and Coding Theories, Signal Processing and Internet of Things. She has collaborated actively with researchers in several other disciplines of Electrical Engineering and industries.

Associate Prof. Ir. Gs. Ts. Dr. Norsuzila Ya'acob

She is an Associate Professor at the Department of Communication Engineering, School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM). Her current research interest are in the areas of Satellite Communication, Space Weather, Remote Sensing & GIS, Wireless Communication, Signal Processing..

Associate Professor Dr Noor Fitrah Abu Bakar

She is an Associate Professor at the School of Chemical Engineering, College of Engineering, UiTM, Shah Alam, Malaysia. Her research interests are nanomaterials designs and processing in various applications including chemical, pharmaceutical, biomedical, environmental, food, agriculture, sensors, oil and gas engineering.

Investigation of Electromagnetic Radiation (EMR) Before and After Super Brain Yoga Exercise Comparing with Short-Term Memory

Ros Shilawani S Abdul Kadir, Suzi Seroja Sarnin*, Muhammad Afiq Kamil Arif, Faizul Hafizzi Ahmad, Wan Norsyafizan Wan Muhamad, Aziati Husna Awang

Universiti Teknologi MARA (UiTM) Shah Alam, 40450 Shah Alam, Selangor, Malaysia.

E-mail: ros885@uitm.edu.my, suzis045@uitm.edu.my, afiqarif97@yahoo.com, fhahmad94@gmail.com, syafizan@salam.uitm.edu.my, aziatihusna@uitm.edu.my*

Abstract

This research is concerned with the investigation of the Electromagnetic Radiation (EMR) on a human body before and after performing Super Brain Yoga (SBY) exercise compared with short-term memory. SBY is one of the earliest yogic practices that is designed to strengthen, energize, and activate the brain cells and their wave. 20 participants were involved in this research and asked to perform SBY exercise for about two weeks, 20 times in the morning and 20 in the evening at their own comfortable place. The frequency detector will be used to obtain frequency measurements before SBY and after two weeks of the workout being practiced. Initially, a digit span test will also be tested to obtain their short-term memory condition before and after SBY is performed. After all of the data had been collected, statistical analysis was used to analyze the result of this experimental work. From the analysis, it shows that the right side improved by 76% and the left side by 62%. Also, there are increments of 80% of the Digit Span test by the participant after performing SBY. Other than that, 3% of the participants decreased while 17% of them remained the same. In conclusion, this research finding shows that performing SBY exercise gives some benefits to an individual; the EMR of the human body is improving significantly with a better short-term memory of a participant.

Keywords: Electromagnetic Radiation (EMR), Super Brain Yoga (SBY), Short-term memory, Frequency Detector, Digit Span Test.

1. Introduction

Super Brain Yoga is often considered "the newest form of yoga" that stimulates the brain" and yet, while it appears to be a modern style of yoga, it is, in reality, a contemporary term for an old technique [1]. This form, which is based on ancient Indian practices, is traditionally known as Thoppukaranam. SBY focuses on the ancient philosophies of yoga and acupuncture and facilitates the release of energy from the brain, across the spine, and across the entire peripheral nervous system [1]. This allows direct flow of energy into energy centers, better known as chakras [1]. This ancient method has, however, continued in a curious way [1].

SBY is a simple and effective brain recharging and energizing procedure. The concepts of subtle energy and ear acupuncture are based on [2]. This is a scientifically validated technique to help super-energize the brain and increase its sharpness and comprehension, a fast, simple, drug-free way to boost mental capacity and promote brain activity [2]. In addition, place pressure on each earlobe with your thumb and finger while bending the knee and breathing.

Pilot research on the impact of SBY on impaired schoolgirls, such as Attention Deficit Hyperactivity Disorder (ADHD)/Attention Deficit Disorder (ADD), behavioral and emotional disorders, down syndrome, and unique intellectual difficulties. Children also recorded major gains in academic and behavioral performance, increased interest in the classroom, and improved social skills, based on prior studies [3]. The result of an electroencephalograph indicated reduced activation in the parieto-occipital area of the brain with one test after SBY [3]. This indicates diminished brain neural activation during exercise [3]. Further tests on the effect of SBY are underway. Evidence has found that these activities intensify the attention of youngsters. Further studies on the influence of SBY are ongoing. Information has found that these exercises are intensifying the concentration of kids.

SBY is one of a kind of exercise. Exercise is important for overall physical and mental health [4]. Research has established that it's beneficial for the brain and may help improve memory in people of all ages, from children to older adults [4]. 15 minutes of moderate exercise on a stationary bike led to improved cognitive performance, including memory, across all ages [4]. In Malaysia, this SBY exercise is also known as Ketuk Ketampi. It is a type of physical punishment at school done from the 80s to the year 2000.

The digit span test includes both forward and backward digit conditions. The test requires a participant to read a series of digit strings, then they must repeat them in the same or reverse sequence of presentation. The digit span test involves both aural attention and short-term verbal memory [5].

2. Literature Review

2.1. *Human Body EMR*

Human Body Electromagnetic radiation (EMR) is often referred to as auras and bio-energies in the human body [7]. Such energy propagates electromagnetic waves in and out of free space [6]. A previous study suggests that SBY transfers energy trapped through the main centers in the simple and sex chakra and eventually into the crown chakra that regulates the pineal gland and general brain health [3]. For the human body, there are 16 points of frequency measurement around the human body, there are eight points on the right and eight points on the left which were regarded as R1-R8 and L1-L8, as shown in Fig. 1. Based on previous research, EMR systems have been used to manage doctrines, improve patient care more easily, and enhance the quality of healthcare delivery. However, it has not been frequently utilized in medical studies for therapeutic purposes [8]. The next research states that a magnetic field is created in the surrounding space by the flow of electrical current. As a result, the electrical current within the human body creates an electromagnetic field around the human body [9].

Past research states that these EMRs encircle the human in an oval shape, emitting signals outside of the physical body at a distance of 2-3 feet [10]. Based on other research, the EMR of the human body after exercise has different types of patterns. From the analysis, after intense exercise higher frequency of EMR is detected [6]. Besides, earlier research found that a non-smoker has better EMR compared to a smoker. Therefore, generally, non-smokers are healthier and fall in the excellent category [11].

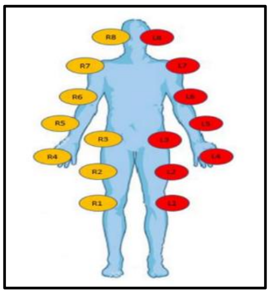


Fig. 1. 16 points of human body [5].

2.2. *Super Brain Yoga (SBY)*

SBY is a simple, balanced exercise that doesn't involve complicated body twists or turns [12]. It is a basic exercise involving the touching and squatting of the ears. As Master Chao Kok Shui suggested, the right earlobe is softly squeezed with the left thumb and the left pointer finger while the left earlobe is squeezed with the right thumb and right pointer finger while performing SBY. The ear creates the requisite relation to each side of the brain for electricity. The left cerebrum and pituitary organ are invigorated and initiated by this interaction [3]. The workout and exercise technique of SBY is very important for the development of direct flow energy called chakras in the human body SBY is often referred to in some parallels as squat. The exercise in squats allows individuals to keep their arms forward. It focuses mostly on the thigh, where there are hamstrings, quadriceps, and glutes [13]. While these two exercises can look the same, the key emphasis and benefits vary. The research shows that SBY has a major influence on both concentration control and working memory, as opposed to simple squat [14]. Fig. 2 illustrates the technique and the need to travel and execute the SBY workout. One of the most common punishments for Malaysians during childhood is SBY.

Previous research aims to evaluate the effect of SBY on the short-term memory and concentration of the students, for which they need to respond to a survey and the findings have been reported, statistically evaluated, and interpreted on the basis of the research [15]. The result suggests a major improvement in short-term memory and concentration before and after the completion of the SBY study [15]. There are Alpha, Beta, Theta, and Delta waves in our human brain. SBY is primarily based on the Alpha waves in Fig. 3 that make the brain operation process calmer. This calm, meditative mindset is related to motivation, increased focus, increased performance, and inner harmony. Mental awareness, alertness, calmness, body/mind alignment, and learning in general are supported by alpha wave support [16].

Another research found that the activity of the alpha wave in the brain increased rapidly after conducting SBY for one minute [16]. This proves it supports kids who are challenged by attention deficit hyperactive disorder (ADHD), speech delay, and autism [16].

Fig. 2. Super Brain Yoga position and steps.

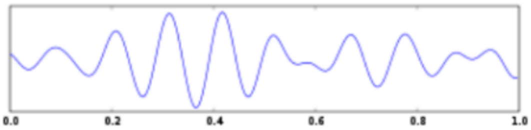


Fig. 3. Alpha Wave [16].

2.3. *Frequency Detector*

As in Fig. 4, human EMR can be measured over a large spectrum using a portable frequency meter and is known as a frequency detector. The dipole antenna is fitted in this frequency detector [17]. The frequency detector has a channel unit used to avoid the existence of abnormal commotion within it [17]. A broad spectrum of human electromagnetic radiation frequencies in the Mega Hertz (MHz) range can be measured by the detector. It is often used as one of the important tools for the identification and measurement of electromagnetic frequency or radiation.

Based on previous research, it proved that the frequency detector does provide complete information on patients' health problems, as well as technical and specific information [6]. In another research, the researcher used a frequency detector to measure human psychological stress [18].

Fig. 4. Frequency detector [6].

2.4. *Short-term memory (STM)*

In Fig. 5, STM refers to information retrieval over a very brief amount of time. Via identification by sensory organs, information reaches the STM inventory. Restricted capacity, limited duration, and encoding are important areas of STM. Small capability refers to the preservation of about seven facts at a time and evidence is quickly lost due to time or distraction. STM and Long-term memory function as two independent stores of memory. Rehearsal of the data is required for the conversion from STM to LTM. The lost information is referred to as decay and the transition of the recalled information to LTM [19]. Previous research found that in India, people with low education levels have a bad short-term memory [5].

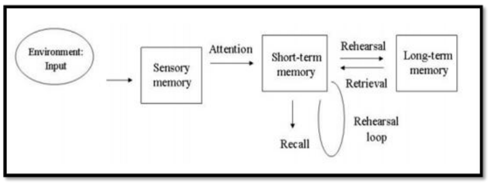


Fig. 5. Short-term memory model [19].

2.5. *Short-term memory loss*

Short-term memory loss is when you forget things you heard, saw, or did recently. It's a normal part of getting older for many people. But it can also be a sign of a deeper problem such as dementia, a brain injury, or a mental health issue. Short-term memory can lead to asking the same question repeatedly, forgetting recent events, forgetting something you saw or read recently, and forgetting where you just put something. One of the home remedies for short-term memory is exercising regularly [20]. Based on a previous study, the peak level of memory performance was found to be 24 and kept decreasing by age [21].

2.6. *Digit span test*

The Digit span test comprises both digit forward and digit backward conditions. The test involves reading out a series of strings of digits to the participants who are required to repeat them in the same or reverse order of presentation. Span is sometimes proverbial to be the number of sequences properly remembered [22]. Performance on the Digit span test requires auditory attention as well as short-term verbal memory [5]. In the previous research, researchers used the Digit span test in Indian contexts to measure attention and working memory [5].

3. Methodology

Fig. 6 shows the flows for this research. Firstly, we have to find a group of volunteers in order to complete this research. The demographic data of participants such as name, age, and information about the participant's health will be recorded. After that, the EMR of participants will be collected using a frequency detector and a digit span test will be done before performing the SBY exercise. After the data had been collected, the participant will repeat SBY 20 times in the morning and 20 times in the evening for about 2 weeks. After completing the exercise for 2 weeks, the duration of 2 weeks has ended, the participant will once again carry out the EMR measurement to check the radiation of the body and the digit span test to check the participant's short-term memory condition. The result will be recorded and will be analyzed. Lastly, the final result will be concluded.

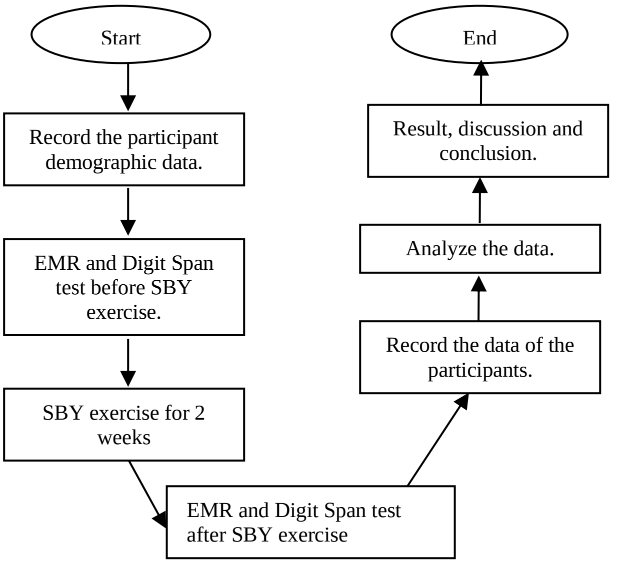


Fig. 6. Flowchart of the research experiment

3.1. *Step of SBY Exercise*

Here are the steps that are involved in SBY [23].

1. Firstly, we have to stand up and make sure that our body is straight.
2. Shift your upper body around with your left hand. Leave the palm in place by using the thumb and forefinger to reach for the right earlobe.
3. In your right hand, cross your upper body and aim for your left earlobe. Once again, put your thumb above your forefinger.
4. Next up, all earlobes should be pressed softly. At this point, do not turn your body position. Make sure your left arm is tucked within your right arm, close to your shoulder.
5. Inhale slowly when moving into the squat position toward the ground.
6. Stay at this position for 1 or 2 seconds if we can. Go up back to a standing place, then we exhale.
7. Repeat all the steps twenty times twice a day this cycle during the morning and evening.

3.2. *Data Collection*

The data were collected involving 30 participants which consist of all genders. Data acquisition is taken before participants begin and after 2 weeks of performing SBY has been completed. This method has been conducted at participant place as shown in Fig. 7 due to pandemic.

Fig. 7. Data collection at participant house.

3.3. *Data Analysis*

There are 15 of color variation of bioenergy. Fig. 8 shows the frequencies and properties of color that identified by the frequency detector [24]. The data was analyzed at 16 points of human body measuring from both side [6]. In order to analyze the data, the Microsoft Excel was used. Table 1 shows all of the total score range which will decide the category of a participant, whereas in Table 2 shows the health score for the participant physical. Based on the two of table above, this illustrates that various type of categories exist. For the satisfactory group, there are two categories which are good and excellent while for the inadequate group, there are moderate and poor [6].

Colour	MHz	Colour	MHz	Colour	MHz	Colour	MHz
Gold	4.0-4.1	Gold	15.7-16.5	Gold	62.5-66.0	Gold	249.8-264.2
Yellow	4.2-4.3	Yellow	16.6-17.4	Yellow	66.1-69.9	Yellow	264.3-279.7
Green	4.4-4.6	Green	17.5-18.5	Green	70.0-74.1	Green	279.8-296.7
Cyan	4.7	Cyan	18.6-18.7	Cyan	74.2-74.9	Cyan	296.8-299.7
Gray/Black	4.8	Gray/Black	18.8-19.3	Gray/Black	75.0-77.5	Gray/Black	299.8-310.2
Cyan	4.9	Cyan	19.4-19.6	Cyan	77.6-78.5	Cyan	310.3-314.2
Blue	5.0-5.1	Blue	19.7-20.7	Blue	78.6-83.1	Blue	314.3-332.7
Navy	5.2-5.5	Navy	20.8-22.0	Navy	83.2-88.1	Navy	332.8-352.7
Purple	5.6-5.8	Purple	22.1-23.3	Purple	88.2-93.4	Purple	352.8-373.7
Orchid	5.9-6.1	Orchid	23.4-24.6	Orchid	93.5-98.6	Orchid	373.8-394.7
Silver	6.2	Silver	24.7-25.0	Silver	98.7-100.0	Silver	394.8-400.2
White	6.3	White	25.1-25.3	White	100.1-101.1	White	400.3-404.9
Burgundy	6.4-6.5	Burgundy	25.4-26.2	Burgundy	101.2-104.8	Burgundy	405.0-419.4
Red	6.6-6.9	Red	26.3-27.7	Red	104.9-111.1	Red	419.5-444.5
Rose	7.0-7.3	Rose	27.8-29.4	Rose	111.2-117.8	Rose	444.6-471.4
Orange	7.4-7.8	Orange	29.5-31.2	Orange	117.9-124.8	Orange	471.5-499.4
Gold	7.9-8.2	Gold	31.3-33.0	Gold	124.9-132.1	Gold	499.5-528.5
Yellow	8.3-8.7	Yellow	33.1-34.9	Yellow	132.2-139.8	Yellow	528.6-559.4
Green	8.8-9.2	Green	35.0-37.0	Green	139.9-148.3	Green	559.5-593.4
Cyan	9.3	Cyan	37.1-37.4	Cyan	148.4-149.8	Cyan	593.5-599.4
Gray/Black	9.4-9.6	Gray/Black	37.5-38.7	Gray/Black	149.9-155.1	Gray/Black	599.5-620.5
Cyan	9.7-9.8	Cyan	38.8-39.2	Cyan	155.2-157.1	Cyan	620.6-628.5
Blue	9.9-10.3	Blue	39.3-41.5	Blue	157.2-166.3	Blue	628.6-665.4
Navy	10.4-11.0	Navy	41.6-44.0	Navy	166.4-176.3	Navy	665.5-705.4
Purple	11.1-11.6	Purple	44.1-46.7	Purple	176.4-186.8	Purple	705.5-747.4
Orchid	11.7-12.3	Orchid	46.8-49.3	Orchid	186.9-197.3	Orchid	747.5-789.4
Silver	12.4-12.5	Silver	49.4-50.0	Silver	197.4-200.1	Silver	789.5-800.5
White	12.6	White	50.1-50.6	White	200.2-202.4	White	800.6-809.9
Burgundy	12.7-13.1	Burgundy	50.7-52.4	Burgundy	202.5-209.7	Burgundy	810.0-838.9
Red	13.2-13.8	Red	52.5-55.5	Red	209.8-222.2	Red	839.0-889.1
Rose	13.9-14.7	Rose	55.6-58.9	Rose	222.3-235.7	Rose	889.1-942.0
Orange	14.8-15.6	Orange	59.0-62.4	Orange	235.8-249.7	Orange	942.0-1000

Fig. 8. The colors and frequencies identified by EMR [24].

Table 1. Total score range which decides the categories of a participant.

Category	Color	Score
Excellent	Gold, Silver	5
Good	White, Orchid, Cyan	4
Moderate	Rose, Purple, Blue	3
	Navy, Green, Yellow, Orange, Black	2
Poor	Burgundy	1
	Red	0

Table 2. The physical health score of a participant [6].

Range of total score	Category
>75	Excellent
56-75	Good
31-55	Moderate
15-30	Poor

4. Results and Discussions

4.1. *Color and Score Analysis*

Table 3 shows the analysis table for the average on the right and left sides of the human body using Microsoft Excel software. Based on Table 4 the percentage for before performing SBY is 82%. The physical health score for this percent falls into the 'Excellent' category. The percentage after performing the SBY exercise is 89%. Although the category is still the same category, there is an increase of percentage of 7%. This shows that after performing SBY has an improvement in the health score of participants.

Table 4 shows the results of the analysis table for 16 point frequencies for the right and left side of one participant with increasing results. Based on Table 4, the percentage before performing SBY is 6% which falls into the 'Poor' category, whereas the percentage after performing SBY exercise is 47% and falls into the 'Moderate' category. This shows an increment of 41%. This show that after performing SBY, there is an improvement in term of the health score of the participant.

Table 3. Average Score, percentage, and category before and after performing SBY.

	Score	Percentage(%)	Category
Total Average Before SBY	65	82	Excellent
Total Average After SBY	72	89	Excellent

Table 4. Score, percentage, and category of one of the increasing participants before and after performing SBY.

	Score	Percentage(%)	Category
Increasing Participants Before SBY	2	6	Poor
Increasing Participants After SBY	37	47	Moderate

Table 5 shows the analysis table for 16 point frequencies for the right and left side of one participant with the decreasing result. The analysis shows that the percentage before performing SBY is 59% which falls into the 'Good' category, whereas the percentage after performing SBY exercise is 40% and falls into the 'Moderate' category. This shows a decrement of 19%. It proves that not all participants give positive results after doing the SBY exercise. This may be due to the participant's emotional or body condition. The participant might be tired or unhappy doing exercise. Some participant that overweight may have muscles torn at their thigh due to sudden exercise if they are not exercising for a long time.

Table 5. Score, percentage, and category of one of the decreasing participants before and after performing SBY.

	Score	Percentage(%)	Category
Decreasing Participants Before SBY	47	59	Good
Decreasing Participants After SBY	32	40	Moderate

Table 6 shows the analysis table for 16-point frequencies for the right and left side of one participant with the unchanged category. Based on Table 4.4, the percentage before performing SBY is 48% which falls into the 'Moderate' category, whereas the percentage after performing SBY exercise is 54% and still falls into the 'Moderate' category. This shows an increment of 6% in term of percentage but still remain the same in term of category. Based on the table, it shows that performing SBY exercise can give a positive result.

Table 6. Score, percentage, and category of one of the participants that have no changes in result before and after performing SBY.

	Score	Percentage(%)	Category
Maintain Participants Before SBY	38	48	Moderate
Maintain Participants After SBY	43	54	Moderate

Table 7 shows the analysis table for 16 point frequencies for the right and left side of another participant with the highly increasing result. Based on Table 7, the percentage before performing SBY is 8% which falls into the 'Poor' category, whereas the percentage after performing SBY exercise is 75% and falls into the 'Excellent' category. This shows an increment of 67%. Based on the table, it shows that performing SBY exercise can improve the overall mental and physical well-being of a person.

Table 7. Score, percentage, and category of one of the highly increasing participants before and after performing SBY.

	Score	Percentage(%)	Category
Increasing Participants Before SBY	6	8	Poor
Increasing Participants After SBY	60	75	Excellent

4.2. *Result of Human Body Electromagnetic Radiation (EMR)*

Fig. 9 and Fig. 10 shows the output results before and after the SBY exercise for the right and left side. Each of the points represents their average frequency. Based on the graph, it shows that the majority after SBY on the right side was in the Excellent category. For R6, R7, and R8, although the frequency decreases, it still falls into the Excellent category. While on the left side has an unusual pattern but still, the frequency is in the Excellent category. The results might be influenced by the surrounding environment as not all participants have the same location during data collection. Further analysis was done in the next part to examine these differences.

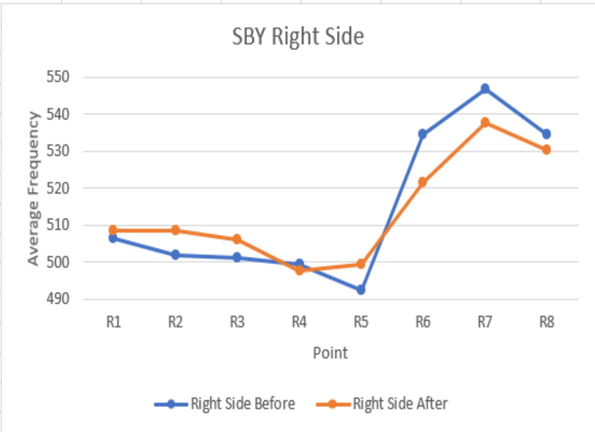


Fig. 9. Comparison of average frequency before and after SBY at the right side point.

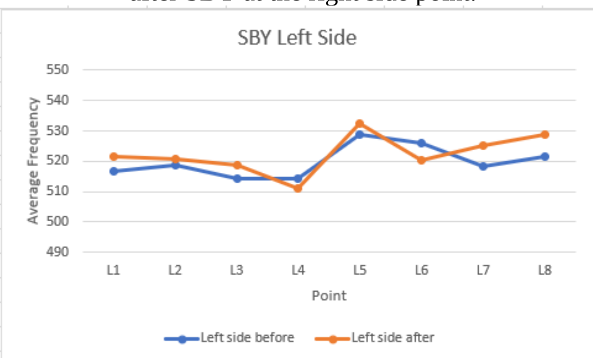


Fig. 10. Comparison of average frequency before and after SBY at the right side point.

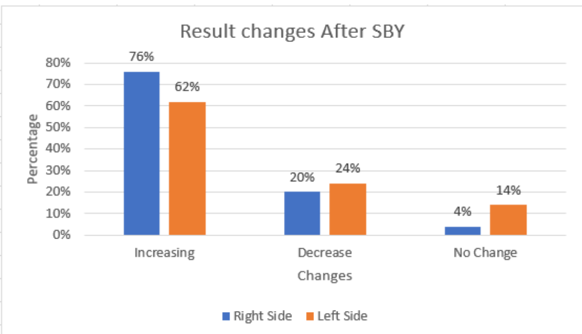


Fig. 11. Right side point percentage after performing SBY for 30 participants.

Fig. 11 shows the right and left side results after performing SBY. The percentage represents the number of participants. Based on the bar graph, the increasing result had the highest percentage for both sides 76% and 62%. The decreasing result had 20% for the right side and 24% for the left side while the no change result had 4% for the right side and 14% for the left side. It shows that SBY can boost the overall physical and mental well-being of a person [25]. SBY also synchronizes the left and right sides of the brain, contributing to an increase in overall brain activity [25] [26].

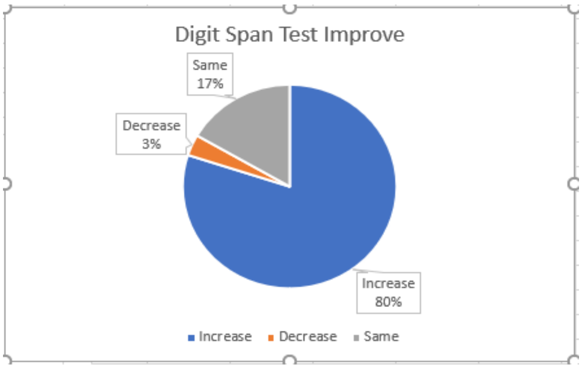


Fig. 12. Number of participants that improved on digit span test.

Fig. 12 shows the percentage of Digit Span test improved for all 30 participants which increased by 80%. Other than that, 3% of the participants are decreasing while 17% remain unchanged. This also proved that SBY appears to be energizing and can help facilitate the overall functioning of brain power [27].

5. Conclusion

In conclusion, this research finding shows that performing SBY exercise has some benefits to an individual and the EMR of the human body is improving significantly. For left and right sides, it has a certain pattern and behavior whereby the frequencies vary between variables. Meanwhile, the percentage of participant that improved their EMR for the right side is 76% while 62% on the left side. This proves that SBY gives positive results and has the potential to significantly improve group dynamics [28]. Other than that, the percentage of short-term memory from the digit span test also increased by 80%. This also proves that SBY can improve short-term memory.

Acknowledgment
The authors would like to thank members School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia for their guidance and support during this project.

References

1. R. Flynn (2018) “What is Superbrain Yoga and How Does It Work?” Mindvalley.
2. D. Raftery (2010) “ Super Brain Yoga” Dave’s wiki.
3. S.N. Jois, L. D’ Souza “ The effectiveness of Superbrain Yoga on concentration, memory and confidence in school students,” Indian Journal of Traditional Knowledge, vol17(4), pp. 741-744, 2018.
4. Jillian Kubala, MS, RD ((2018) ” 14 Natural Ways to Improve Your Memory” healthline.
5. Ravikesh Tripathi1 , Keshav Kumar2 , Srikala Bharath3 , Marimuthu P4 , Vikram Singh Rawat5 , Mathew Varghese “ Indian older adults and the digit span” Dement Neuropschol 2019 March; 13(1):111-115.
6. R.S.S.A.Kadir, Z.H. Murat, N.S.Johari “Electromagnetic Radiation (EMR) of Human body before and after intense exercise” Indonesian Journal of Electrical Engineering and Computer Science,. vol 17, no. 1, pp. 141-147,2020.
7. R.S.S.A. Kadir, Z.H. Murat, N.I.N.M Suhaimi “ Electromagnetic Radiation (EMR) of Human Body before and after jogging,” Indonesian Journal of Electrical Engineering and Computer Science., vol 9, no. 3, pp.643-649,2018.
8. Institute of Medicine. Relieving pain in America: A blueprint for transforming prevention, care, education, and research. Washington DC: The National Academies Press; 2011.
9. S. Z. A. Jalil, M. N. Taib and H. Abdullah Idris, "Investigation of frequency radiations of the human body," 2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), 2010, pp. 377-380.
10. R. S. S. A. Kadir, Z. H. Murat, M. N. Taib, and S. Z. A. Jalil, “ Investigation of electromagnetic radiation for stroke patients and non-stroke participant,” Proc. – 2015 6th IEEE Control Syst. Grad. Res. Colloquium, ICSGRC 2015, pp. 130-134, 2016, doi; 10.1109/ICSGRC.2015.7412479.
11. R. S. S. A. Kadir, Z. H. Murat, H. Hashim and W. N. W. Muhamad, "A preliminary study on human body aura of smoker and non smoker using Resonant Field Imaging (RFI)," 2011 IEEE International Conference on System Engineering and Technology, Shah Alam, 2011, pp. 209-214.
12. S. Jonex, C-IAYT(2020)” How to do Superbrain Yoga”, Wikihow.
13. Hana Medeysek (2019) , “Squat 101: How to do squats and which muscle they activate” Adidas.
14. Thomas JI, Venkatesh D.” A comparative study of the effects of superbrain yoga and aerobic exercise on cognitive functions,” Natl J Physiol Pharm Pharmacol 2017; 7(9); 895-900.
15. S.N.Jois, L.D’ Souza, R. Moulya “Beneficial effects of Superbrain Yoga on short term memory and selective attention students,” Indian Journal of Traditional Knowledge, vol 16 (Suppl), pp. S35-S39, 2017.
16. Wikipedia contributors (2020) “Alpha waves,” Wikipedia.
17. S.Z.A. Jalil, A.N. Shamsuddin, S. A. M. Aris, N. A . Bani, H.M Kaidi, M.N. Muhtazaruddin, S. Usman, M. A. M. Izhar, R. Yusuf, R.S.S.A. Kadir “ Investigation of Human Electromagnetic Radiation Characteristic For Kidney Disease Patients,” International Journal of Engineering & Technology, 7 (4.11) (2018) 40-43.
18. N. I. Abdullah, O. C. Bakar, A. Omar, Z. H. Murat and M. I. Selamat, "The screening of human psychological stress using Resonant Field Imaging (RFI) frequency counter," 2013 IEEE International RF and Microwave Conference (RFM), Penang, pp. 382-386, 2013.
19. Ellison, Angela Foil, “ Short-term Memory Capacity and Recall with student with and without Intellectual Giftedness: An Empirical Inquiry” Dissertations. 1464.
20. Erica Hersh (2018) “ Short Term Memory Loss: Symptoms, Diagnosis, Treatment and More”.
21. Berkauzer, Alisha "The Effects of Age on Short-Term Memory Loss due to Proactive Interference," Undergraduate Journal of Mathematical Modeling: One + Two: Vol. 4: Iss. 1, Article 2. (2011).
22. Elliot, C., D. Murray, and L. Pearson. "British abilities scales." Windsor, England: nfer-nelson (1983).
23. “What is SuperBrain Yoga? | Metagenics Blog”.
24. M. Rosdi, R.S.S.Ros Shilawani, Z. Hj. Murat, and N. Kamaruzaman, “The comparison of human body Electromagnetic radiation between Down Syndrome and Non Down syndrome person for brain, chakra and energy field stability score analysis,” Proc.- 2012 IEEE Control System. Grad. Res. Colloqium, ICSGRC 2012, no. Icsgrc, pp. 370-375, 2012.
25. The Art of Living, “5 super brain yoga exercises to boost your gray matter,” Art of Living (Global).
26. Akshar, G. M. (2020, June 10), “3 super brain yoga exercises for children: Steps, health benefits and precautions,” Times Now.
27. F. (2018, February 21), “15 Unbelievable Benefits Of Super Brain Yoga,” Indian Fashion Blog.
28. Yoga Vidya Pranic Healing Foundation of West Bengal, (2019), “Superbrain Yoga: The Smart Way To Become Smart,” PranicKolkata.Com.

Authors Introduction

Dr. Hjh Ros Shilawani S Abdul Kadir

She is working as a Senior Lecturer at the School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She received PhD in Wireless Communication UiTM, a Master of Science in Microelectronics and a Degree in Electrical Engineering UTM. Her interests include Wireless Communication and IoT.

Ir. Ts. Dr Suzi Seroja Sarnin

She is working as a Senior Lecturer at the School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She received PhD in Wireless Communication UiTM, a Master of Science in Microelectronics and a Degree in Electrical Engineering UTM. Her interests include Wireless Communication and IoT.

Muhammad Afiq Kamil Arif

He is currently serving as a Traction Power Supply Engineer at Keretapi Tanah Melayu. He holds a degree in Electrical Engineering majoring in Power from UiTM. His interests are electrical power engineering.

Faizul Hafizzi Ahmad

He is currently working as a Junior Engineer at the Department of Biomedical Engineering, Hospital Sultan Ismail Johor Bahru. He holds a degree in Electronic Engineering majoring in communication from UiTM. His interests are biomedical engineering.

Dr Wan Norsyafizan Wan Muhamad

She is working as a Senior Lecturer at the School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She received PhD in Electrical, Electronic and System Engineering UKM, Bangi. Her interests include Optical sensor, Fiber Bragg Grating (FBG) based sensor, Optical Network.

Dr. Aziati Husna Awang

She is working as a Senior Lecturer at School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She received her PhD in Electrical, Engineering at Leeds University, UK. Her interests include THZ Detection, Photonic Sensors Optical Devices Microwave/RF Devices and IoT Technologies

Preliminary Investigation of Electromagnetic Radiation (EMR) Between Adults and Children

Ros Shilawani S Abdul Kadir, Aziati Husna Awang*, Muhamad Azizularif Mohamad Azizan, Suzi Seroja Sarnin, Suhaila Subahir, Roshakimah Mohd Isa

Universiti Teknologi MARA (UiTM) Shah Alam, 40450 Shah Alam, Selangor, Malaysia.

E-mail: ros885@uitm.edu.my, aziatihusna@uitm.edu.my, azizularif97@yahoo.com, suzis045@uitm.edu.my, suhailas@uitm.edu.my, roshakimah@uitm.edu.my*

Abstract

This research is concerned with the initial investigation of electromagnetic radiation (EMR) between adults and children which involved 30 participants from the adult group and 32 participants from the children group. The range of age of the children group is set at 3-12 years old and for adult group is set at ages of 20-25 years old. The EMR is created by an electrical current within the human body and it vibrates on its own characteristics radiation of frequency around the body. A frequency detector was used to measure the EMR frequencies (in MHz) around 16 points of the human body. The obtained data were assessed by examining the pattern and behavior of captured frequencies, as well as comparing the frequencies of adults and children. Microsoft Excel was used to analyze the data by generating graphs and charts. The collected data were also converted into color code hence; physical health condition was obtained. From the data analysis, adults have higher frequencies of reading in all body points as compared to children. Consequently, adults and children also have different frequency patterns, color-coded and health scores in right-left comparison and segmentation comparison. The analysis also showed that children have better health scores as compared to adults. In conclusion, the EMR emitted from adults and children are significantly different in frequency and have their own characterized frequency patterns and children have better health scores compared to adults.

Keywords: Electromagnetic Radiation (EMR), Adults, Children, Frequency Detector

1. Introduction

Adults and children are different physically and psychologically. A child is a person who is in the developmental stage between infancy and puberty. Researchers often refer to two further stages of development that occur between infancy and adolescence: early childhood and middle childhood. Early childhood is defined as the period between the ages of two and five years. Middle childhood is commonly defined as the years between the ages of six and twelve [1]. This research concerns children from early and middle childhood.

The development from childhood to adulthood takes place on several levels: biological, cognitive, social, emotional, and behavioral. Internal variables such as heredity drive development, whereas extrinsic factors such as nutrition and culture shape it [1]. Then, children transition into adolescence and then transition to adulthood depending on the culture's definition of adults. Adolescence to adulthood is called emerging adulthood, which it's an important time for very important development in the brain especially in terms of strengthening interconnections across brain structures [2] where it is involved in high-level thinking which affects the whole body.

Theoretically and scientifically, each person has an EMR vibration in their body, and this vibration can reveal information on a variety of aspects of a person's physical health, and emotional and psychological status [3]. Right now, the world is moving faster than ever before causing most adults to experience a continuous stream of pressure to keep up, stress to do well, and fatigue in their body [4][5][6]. These negative thoughts and negative energy within an adult's body have caused the EMR of the human body to become unstable [7]. These findings prompted the initial investigation of EMR using a frequency detector, with the aim of determining if there are differences in EMR frequency between children and adults. Hence, the main target of this research is to analyze, compare, and characterize the pattern of EMR between adults and children in 16 points of the left-right side of the human body.

2. Literature Review

2.1. *Human Electromagnetic Radiation*

The concept of energy fields has a long history in Eastern cultures, where health, healing, self-defense, and life are all characterized in terms of energy flow [8]. In the Western world, it is thought that every living creature has its own radiation, which radiates into space surrounding the body in an oval form and is represented as endogenous energy fields created by and contained inside the body [9]. Human electromagnetic radiation varies depending on the health and activity of the body as a biological system [10] [11]. Furthermore, earlier research has demonstrated that the human body has varied electromagnetic radiation characteristics depending on gender [9] [10] [11]. This occurs because of small electrical currents, or bioelectricity, which is a component of the electrochemical energy system that exists because of chemical reactions that occur naturally in the human body [12] [13]. As a result, the electromagnetic field that surrounds the body is generated by an electrical current within the human body, and it vibrates on its own characteristics radiation of frequency around the body [14]. The human body is divided into several sections in biomechanics research, including the head, neck, upper torso, lower torso, upper arm, forearm, hand, thigh, calf, and foot [15].

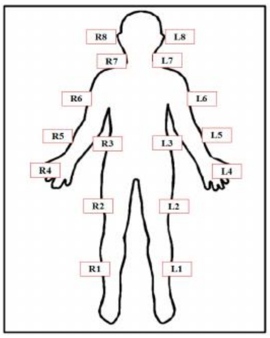


Fig. 1. 16 points surround the body [14].

This EMR could be captured using various detector instruments like BEO GDV (Gas Discharge Visualization) [16], Dipole Antenna, Lecher Antenna, Frequency Wave Detector, Biopulsar Relfexograph, Aura Color Space Visualizer algorithm, Quantum resonance magnetic analyzer [17], Biophoton detector [18] and Kirlian photography[19]. In this research, the human body consists of 16 points surrounding the body, namely the left side and right side with 8 points representing each part [19], particularly R1-R8 and L1-L8 as shown in Fig. 1.

2.2. *Adult*

An adult is a biologically human being that has reached maturity in puberty for women, having menstruation and blossoming breasts while for men, it growth of facial hair, and deep voice, and the development of pubic hair in both genders [2]. Physical strength, coordination, and dexterity peak in both genders in their late twenties and early thirties, then gradually decline throughout life, even if they continue an active lifestyle [2]. Sensory acuity is likewise at its best in the early 20s, while visual acuity stays high until middle age when people tend to become farsighted and require reading glasses [2]. Hearing begins to deteriorate slightly by the late twenties, particularly for high-pitched tones [2].

Males and females differed greatly in much research, particularly in terms of the human physical body and emotions. It was discovered that gender disparities in emotions and self-control were evident as early as the infancy stage [20]. Males and females have various frequency radiation properties, and frequency radiation analysis of the human body, particularly the thoracic region, can be utilized to classify and categorize human gender [21]. It was also discovered that due to differences in brain structure and activity, males and females emit distinct frequencies of EMR in the upper body region, where neuroendocrine mechanisms that are engaged in the combination of genes and hormones broadcast various frequencies of EM [20]. Previous studies also reveal that humans have various frequency radiation characteristics depending on gender and that it is possible to identify humans based on frequency radiation analysis of the human body, notably in the arm segment [22].

2.3. *Children*

In terms of biology, a group of children are considered human beings that progress from stages of birth and head to the stages of puberty. The development of children is divided into three sections, which are early childhood development, middle childhood development, and adolescent development. The growth pattern of middle childhood children in the age of 7 – 10 years old, is gaining in weight and increasing in height. During this development period, boys and girls have similarities in physical like gaining height and weight as is accompanied by an increase in muscular strength. The relative proportion of muscle and fat tissue are the same for both genders, but this will begin to change in age about 11 years old where boys generally develop more muscle tissue and girls develop more fatty tissue. Throughout this middle childhood, boys and girls perform similarly in most activities related to motor skills. The boys show a slightly greater difference in overall strength especially in forearm strength than girls. Girls, on the other hand, have slightly better limb coordination and general body flexibility, which is advantageous in dance, balancing, and gymnastics [1].

The child's energy field is entirely open and exposed to his surroundings. The energy environment has a constant influence on the child, and the child is always reacting to it. This idea, since infants are born with open and susceptible human energy fields, suggests that children have more physical and psychological energy and are healthier. In the study's preliminary findings, it was discovered that younger children's energy fields were less dense and had more translucency than older children's energy fields. The energy field was also narrower or shorter in these younger children, with more frequent square openings, more observed rings, and a more dynamic field overall. The discovery that the EMR of the youngest children is smaller, less dense, and more dynamic is interesting because it supports the notion of children's fields being more fully open but also more vulnerable [8].

Another study found differences in movement coordination between children with attention-deficit/hyperactivity disorder (ADHD) and healthy children when doing simple and complex tasks. This is because children with ADHD have poor coordination and perform poorly in both simple and complicated tasks.

However, there is no difference in reaction time between children with ADHD and children without ADHD, indicating that there is no substantial difference in the speed of visual-motor integration between the two groups, implying that some components of cognitive processing may be the same [23].

Consequently, there has been research on dyslexia and normal children based on writing words on paper, and it has been found that the frequency spectrum pattern and frequency ranges of dyslexic and normal children in relax and writing are different based on EEG signals. During relaxing or writing activities, dyslexic children produce a higher frequency of EEG signal than normal children [24]. Another study discovered that dyslexic children spend more energy, especially if they have autism and respond improperly while writing. Even if they write correctly, dyslexic children's log energy entropy of EEG data suggests that they are still impaired because they consume more energy in the right hemisphere of the brain [25].

2.4. *Frequency Detector*

Since the human radiation wave is invisible and highly defined, it is difficult to measure and record the readings. Fig. 2 shows an antenna-equipped hand-held detector that provides a realistic, real-time readout of EMR frequencies in natural states at the time of testing. This detector additionally incorporates an ultrasensitive detector and filter module for determining the relative field strength of electromagnetic waves that interact with antennas. [3]. This system recognizes and defines 15 distinctive colors of the visual spectrum to define 15 distinct electromagnetic fields in various parts and distances from the body [4].

Fig. 2. Frequency Detector [13].

In previous research, it was demonstrated that the Frequency Detector may be used as an acceptable instrument for measuring stress, with Heart Rate Variability (HRV) and Galvanic Skin Response (GSR) as references [4] [5]. It's also can be used to detect fatigue in human body differentiate the frequencies of the human body before and after exercise on the subjects [6].

Another study showed that the Frequency Detector may be used to distinguish between post-stroke patients and non-stroke participants by comparing frequency acquired from the left and right sides, as well as chakra points. Non-stroke individuals had higher EMR for left and right-side frequency but lower EMR for chakra than post-stroke patients [19].

3. Methodology

3.1 *Flowchart of Research*

Fig. 3 shows the flowchart of this research. The first step is to find a group of volunteers from adults and children. Then, a record of volunteers' demographic data will be collected such as name, age, and health condition. In order to avoid any interruptions, the participants were told to remain calm and steady during the EMR measurement. Data was collected on the EMR from 8 points on the left and right sides of the body, respectively. Lastly, the EMR data will be analyzed to conclude the results of the research.

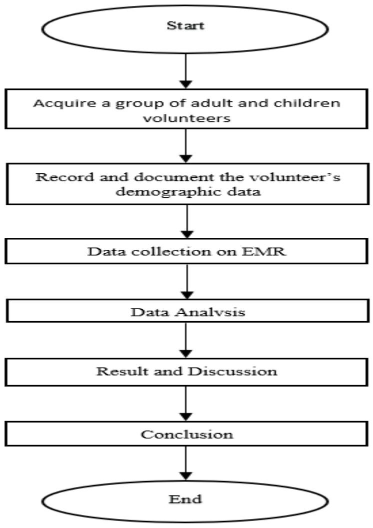


Fig. 3. Flowchart of the research

3.1. *Data Collection and Data Analysis*

30 healthy adults (20-25 years old) and 32 healthy children (3-12 years old) consisting of both genders participated in this study. The study was conducted in a room or space that comfortable and free of any electronic signal that may interrupt the measurement. The data was analyzed using a right-left comparison of adults and children, as well as body segmentation of the upper, middle, and lower body, as shown in Fig. 4, Fig. 5, and Fig. 6. Microsoft Excel was utilized for data analysis, and the data was categorized into 15 color variants of bioenergy, as shown in Fig. 7. This color variation will be converted into the score presented in Table 1 and used to determine a participant's health score category in Table 2.

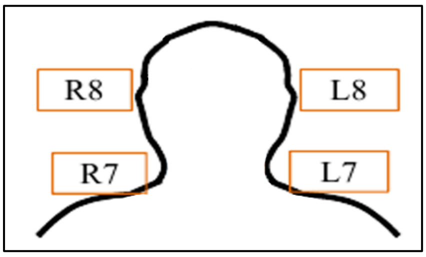


Fig. 4. Upper Segment Points [21]

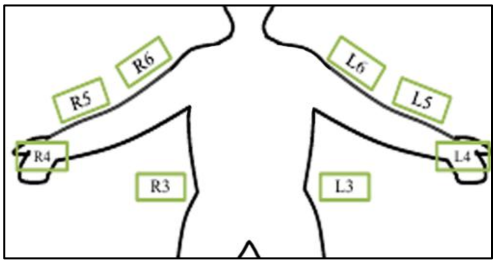


Fig. 5. Middle Segment Points [20][22]

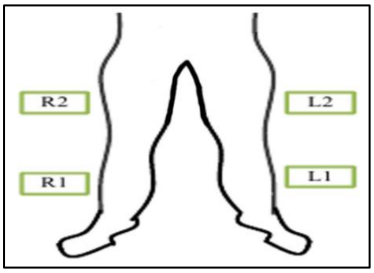
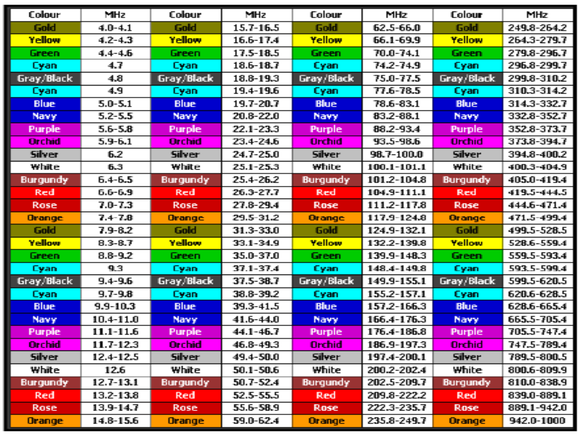


Fig. 6. Lower Segment Points [15]



Colour	MHz	Colour	MHz	Colour	MHz	Colour	MHz
Gold	4.0-4.1	Gold	15.7-16.5	Gold	62.5-66.0	Gold	249.8-264.2
Yellow	4.2-4.3	Yellow	16.6-17.4	Yellow	66.1-69.9	Yellow	264.3-279.7
Green	4.4-4.6	Green	17.5-18.5	Green	70.0-74.1	Green	279.8-296.7
Cyan	4.7	Cyan	18.6-18.7	Cyan	74.2-74.9	Cyan	296.8-299.7
Gray/Black	4.8	Gray/Black	18.8-19.3	Gray/Black	75.0-77.5	Gray/Black	299.8-310.2
Cyan	4.9	Cyan	19.4-19.6	Cyan	77.6-78.5	Cyan	310.3-314.2
Blue	5.0-5.1	Blue	19.7-20.7	Blue	78.6-83.1	Blue	314.3-332.7
Navy	5.2-5.5	Navy	20.8-22.0	Navy	83.2-88.1	Navy	332.8-352.7
Purple	5.6-5.8	Purple	22.1-23.3	Purple	88.2-93.4	Purple	352.8-373.7
Orchid	5.9-6.1	Orchid	23.4-24.6	Orchid	93.5-98.6	Orchid	373.8-394.7
Silver	6.2	Silver	24.7-25.0	Silver	98.7-100.0	Silver	394.8-400.2
White	6.3	White	25.1-25.3	White	100.1-101.1	White	400.3-404.9
Burgundy	6.4-6.5	Burgundy	25.4-26.2	Burgundy	101.2-104.8	Burgundy	405.0-419.4
Red	6.6-6.9	Red	26.3-27.7	Red	104.9-111.1	Red	419.5-444.5
Rose	7.0-7.3	Rose	27.8-29.4	Rose	111.2-117.8	Rose	444.6-471.4
Orange	7.4-7.8	Orange	29.5-31.2	Orange	117.9-124.8	Orange	471.5-499.4
Gold	7.9-8.2	Gold	31.3-33.0	Gold	124.9-132.1	Gold	499.5-528.5
Yellow	8.3-8.7	Yellow	33.1-34.9	Yellow	132.2-139.8	Yellow	528.6-559.4
Green	8.8-9.2	Green	35.0-37.0	Green	139.9-148.3	Green	559.5-593.4
Cyan	9.3	Cyan	37.1-37.4	Cyan	148.4-149.8	Cyan	593.5-599.4
Gray/Black	9.4-9.6	Gray/Black	37.5-38.7	Gray/Black	149.9-155.1	Gray/Black	599.5-620.5
Cyan	9.7-9.8	Cyan	38.8-39.2	Cyan	155.2-157.1	Cyan	620.6-628.5
Blue	9.9-10.3	Blue	39.3-41.5	Blue	157.2-166.3	Blue	628.6-665.4
Navy	10.4-11.0	Navy	41.6-44.0	Navy	166.4-176.3	Navy	665.5-705.4
Purple	11.1-11.6	Purple	44.1-46.7	Purple	176.4-186.8	Purple	705.5-747.4
Orchid	11.7-12.3	Orchid	46.8-49.3	Orchid	186.9-197.3	Orchid	747.5-789.4
Silver	12.4-12.5	Silver	49.4-50.0	Silver	197.4-200.1	Silver	789.5-800.5
White	12.6	White	50.1-50.6	White	200.2-202.4	White	800.6-809.9
Burgundy	12.7-13.1	Burgundy	50.7-52.4	Burgundy	202.5-209.7	Burgundy	810.0-838.9
Red	13.2-13.8	Red	52.5-55.5	Red	209.8-222.2	Red	839.0-889.1
Rose	13.9-14.7	Rose	55.6-58.9	Rose	222.3-235.7	Rose	889.1-942.0
Orange	14.8-15.6	Orange	59.0-62.4	Orange	235.8-249.7	Orange	942.0-1000

Fig. 7. Human Body Radiation Detector Identifies by Colors and Frequencies [26]

Table 1. Total range score [11]

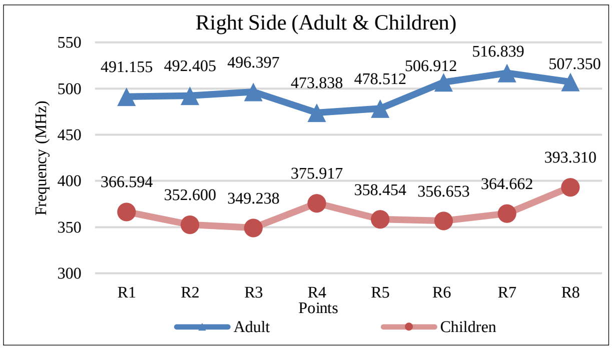
Category	Color	Score
Excellent	Gold, Silver	5
Good	White, Orchid, Cyan	4
Moderate	Rose, Blue, Purple	3
	Orange, Yellow, Navy, Gray, Green	2
Poor	Burgundy	1
	Red	0

Table 2. Physical Health Score [11]

Range of Total Score (%)	Category
>75	Excellent
56 - 75	Good
31 - 55	Moderate
15 – 30	Poor
0 - 14	Bad

4. Results and Discussions

4.1. *Right & Left Side Comparison Between Adults and ChildreFig. 8. Right side of body*



points (Adults & Children)

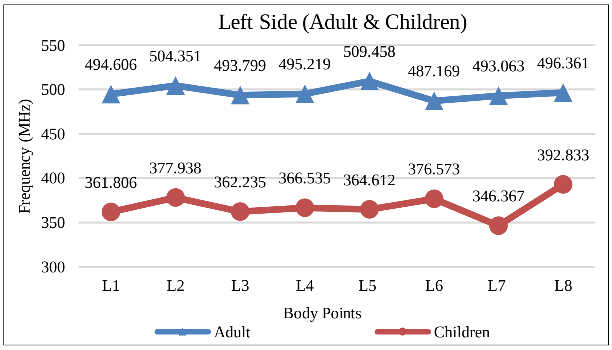


Fig. 9. Left side of body points (Adults & Children)

Fig. 8 and Fig. 9 show that the frequencies on both sides of adults and children are considerably different, with adults having a higher average frequency across all points. This might be because adults' bodies are considerably larger and more mass than children's bodies, which are still in the growing phase [27]. In Fig. 8, there are noticeable patterns in adult and children graphs. The adults' graph shows that there was a slight rise from R1 to R3 and then a significant fall occurred at R4, where it rose again until R7 and then declined again. For the children's graph, the average frequency dropped from R1 to R3, then surged up at R4, after which it gradually declined until R6, and then increased until R8. It seems that the Right-side frequency pattern of adults and children seems to be almost mirroring each other based on Fig. 8.

As seen in Fig. 9, both graphs exhibit nearly identical trend patterns from L1 to L4, where they fluctuate similarly. The graph for adults shows an increase from L4 to L5, but then declines at L6 and gradually climbs after that. The children's graph has a different pattern than the adults' graph where after L4, it drops slightly in L5 and after L5, it fluctuates drastically until L8.

Table 3. Right Side Color Analysis

Group	R1	R2	R3	R4	R5	R6	R7	R8
Adults	491.155	492.405	496.397	473.838	478.512	506.912	516.839	507.350
Children	366.594	352.600	349.238	375.917	358.454	356.653	364.662	393.310

Table 4. Left Side Color Analysis

Group	L1	L2	L3	L4	L5	L6	L7	L8
Adults	494.606	504.351	493.799	495.219	509.458	487.169	493.063	496.361
Children	361.806	377.938	362.235	366.535	364.612	376.573	346.367	392.833

Table 5. Right Side Physical Health Score

Group	Score	Full Score	%	Category
Adults	25	40	62.5%	Good
Children	24	40	60%	Good

Table 6. Left Side Physical Health Score

Group	Score	Full Score	%	Category
Adults	22	40	55	Moderate
Children	26	40	65	Good

All frequencies in Fig. 8 and Fig. 9 are color-coded according to Fig. 7 and tabulated in Table 3 and Table 4. These tables show that adults and children have distinct and scarcely comparable color codes, with adults having more orange and gold color codes than children, who have purple, navy, and pink color codes on both sides. Hence, all of these colors are translated into scores based on Table 1, and the scores are tabulated in Table 5 and Table 6.

In Table 5, it appears that on the right-hand side, adults have a slightly better score than children in health physical score, with a difference of 2.5%. With those scores, both groups are put into the "Good" category in health physical analysis. In Table 6, however, the adult group has a lower score than the children on the left side. With a physical health score that is 10% lower than that of children, adults are classified as "Moderate," while children are classified as "Good." Table 7 combines the scores from Table 5 and Table 6, giving adults a score of 47 and children a score of 50. Adults had a 3.75% lower health score than children, but both groups are categorized as "Good".

Table 7. Total Score & Category of Adults and Children in Health Score

Group	Score	Full Score	%	Category
Adults	47	80	58.75%	Good
Children	50	80	62.50%	Good

4.2. *Analysis of Body Segmentation of Upper, Middle, and Lower part*

This section of analysis of human EMR was analyzed in different segmentations of the upper, middle, and lower segments. Then, in each segment, the adult's and children's average frequencies were compared and classified into color as in Table 1 and scored by following Table 2 classification and Table 3 categorization.

4.2.1 *Analysis of Upper Segment*

Fig. 10 illustrates that the upper segment of adults has a higher frequency than children. It is also observed that the frequency values on both sides of the adults are almost equal, resulting in a practically identical flat line on the line chart. Meanwhile, the children's line graph has a distinct pattern from the adults, where it seemed to have high and low points in comparison to the previous points.

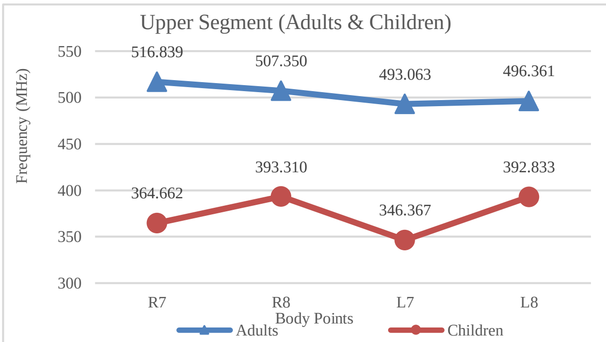


Fig. 10. Upper Segment Frequencies

Table 8. Upper Segment Color Analysis

Group	R7	R8	L7	L8
Adults	516.839	507.350	493.063	496.361
Children	364.662	393.310	346.367	392.833

Table 9. Upper Segment Physical Health Score

Group	Score	Full Score	%	Category
Adults	14	20	70	Good
Children	13	20	65	Good

In Table 8, the frequencies in the upper segment of adults and children are color-coded. As a result, those colors were transformed into scores, which can be shown in Table 9, where adults have a 5% greater score than children. Even though adults have a greater score than youngsters, both groups are in the "Good" category.

4.2.2 *Analysis of Middle Segment*

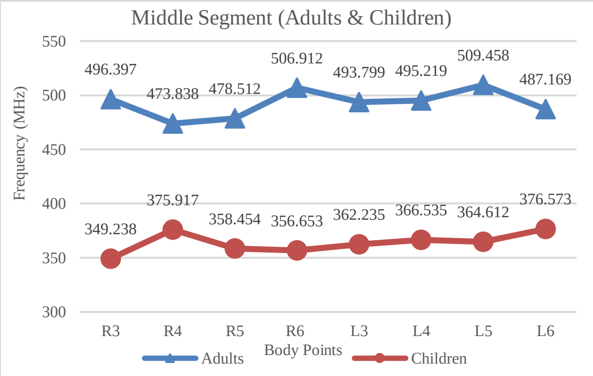


Fig. 11. Middle Segment Frequencies

Table 10. Middle Segment Color Analysis (Right)

Group	R3	R4	R5	R6
Adults	496.397	473.838	478.512	506.912
Children	349.238	375.917	358.454	356.653

Table 11. Middle Segment Color Analysis (Left)

Group	L3	L4	L5	L6
Adults	493.799	495.219	509.458	487.169
Children	362.235	366.535	364.612	376.573

Table 12. Middle Segment Color Analysis

Group	Score	Full Score	Percentage	Category
Adults	22	40	55%	Moderate
Children	25	40	62.5%	Good

The graphs shown in Fig. 11, show that the adults have higher frequencies than the children in all body points of the middle segment. The children's line graph appeared to be flat on its Left Side while the Right side had drastic differences in frequency values at R3 and R4. For the adults, the line graph is shifting and inconsistent throughout all body points in the middle segment. In Table 10 and Table 11, the average frequencies of the middle segment of children and adults are interpreted into color. In Table 12, children have a better score than adults with 7.5% higher which put the children in the "Good" category while the adult put into the "Moderate" category. Thus, children are better than adults in physical health.

4.2.3 *Analysis of Lower Segment*

Fig. 12 shows that adults have greater frequency values for the lower segment than children. The adults' line graph demonstrates an increasing trend with a small increment. The children's line graph shows a decreasing trend in the beginning and afterward, rise slowly. Table 13 shows that adults have three oranges and one gold

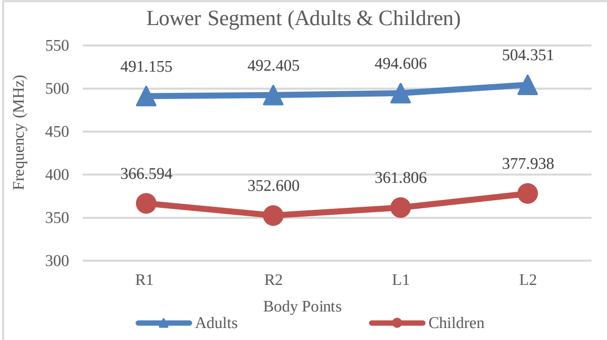


Fig. 12. Lower Segment Frequencies

Table 13. Lower Segment Colour Analysis

Group	R1	R2	L1	L2
Adults	491.155	492.405	494.606	504.351
Children	366.594	352.600	361.806	377.938

Table 14. Lower Segment Physical Health Score

Group	Score	Full Score	%	Category
Adults	11	20	55%	Moderate
Children	12	20	60%	Good

color, whereas children have two purple, one navy, and one pink color. When these colors were turned into a score, it revealed that adults had a 5% lower score than children, as seen in Table 14. Therefore, the children were placed in the "Good" group, while the adults were placed in the "Moderate" category in terms of physical health. Overall, the children have better physical health scores even though they have lower frequency values compared to adults.

5. Conclusion

Finally, in conclusion, this study demonstrated that children and adults differ in terms of EMR frequency and frequency pattern whereas adults have higher average frequency in all points. It also shows that children and adults have unique frequency patterns on both sides. In terms of physical health scores, children have better physical health scores compared to adults overall even in segmentation. Thus, making children healthier than adults [28].

Acknowledgment

The authors would like to thank members School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia for their guidance and support during this project.

References

1. S. Rathus, *Childhood and Adolescence: Voyages in Development*. Cengage Learning, 2011. https://tinyurl.com/yktv46sm
2. R. V Kail and J. C. Cavanaugh, A Life-Span View, Cengage Learning, 2019. https://tinyurl.com/hfdsu425

3. F. Bin Jumain et al., “Mood assessment using human electromagnetic radiation,” Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2017, pp. 251–256, 2017.
4. R. Raveendra, M. Regueiro, and K. George, “Acute Stress Detection and Analysis Using Resonant Field Imaging (RFI) Technique,” 2018 9th IEEE Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2018, pp. 1009–1016, 2018.
5. N. I. Abdullah, O. C. Bakar, A. Omar, Z. H. Murat, and M. I. Selamat, “The screening of human psychological stress using Resonant Field Imaging (RFI) frequency counter,” RFM 2013 - RFM 2013 - 2013 IEEE International RF and Microwave Conference, Proceedings, pp. 382–386, 2013.
6. N. A. A. Nazari, N. M. Fauzi, N. Farahiyah Rosli, S. N. Zakaria, S. Z. A. Jalil, and N. M. Noor, “Physiological studies of human fatigue using human electromagnetic radiation,” Proceedings of the 2017 IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2017, pp. 559–562, 2017.[Retracted Paper]
7. N. I. Abdullah, O. C. Bakar, A. Omar, Z. H. Murat, and M. I. Selamat, “Bioenergy frequency measurement of stressed and non-stressed individuals using Resonant Field Imaging (RFI) frequency counter,” IECBES 2014, Conference Proceedings - 2014 IEEE Conference on Biomedical Engineering and Sciences: "Miri, Where Engineering in Medicine and Biology and Humanity Meet” no. December 2014, pp. 667–671, 2014.
8. R. M. Dougherty, “A Comparison Of Human Energy Fields Among Children, Youth, Adults & Dahn Masters,” Journal of General Virology, vol. 33, no. 1, pp. 61–70, 1976.
9. S. Z. A. Jalil, M. N. Taib, and H. Abdullah Idris, “Investigation of frequency radiations of the human body,” Proceedings of 2010 IEEE EMBS Conference on Biomedical Engineering and Sciences, IECBES 2010, no. December, pp. 377–380, 2010.
10. S. Z. A. Jalil et al., “Investigation of human electromagnetic radiation characteristic for kidney disease patients,” International Journal of Engineering and Technology (UAE), vol. 7, no. 4, pp. 40–43, 2018.
11. S. Z. A. Jalil, M. N. Taib, H. Abdullah, and M. M. Yunus, “Frequency radiation characteristic around the human body,” International Journal of Simulation: Systems, Science and Technology, vol. 12, no. 1, pp. 34–39, 2011.
12. R. S. S. A. Kadir, Z. H. Murat, and N. S. Bin Johari, “Electromagnetic radiation of human body before and after intense exercise,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 17, no. 1, pp. 141–147, 2019.
13. R. S. S. A. Kadir, Z. H. Murat, M. N. Taib, and F. A. Hanapiah, “The initial investigation of electromagnetics radiation for the right hemisphere stroke patient,” 2014 2nd International Conference on Electrical, Electronics and System Engineering, no. November, pp. 108–113, 2014.
14. R. S. S. A. Kadir, Z. H. Murat, M. N. Taib, and F. A. Hanapiah, “The preliminary investigation of electromagnetics radiation for the left hemisphere stroke,” 2014 2nd International Conference on Electrical, Electronics and System Engineering, no. November, pp. 108–113, 2014
15. S. Z. A. Jalil, S. A. Mohd Aris, N. A. Bani, M. N. Muhtazaruddin, and S. Usman, “Segmentation of the human body based on frequency of human electromagnetic radiation,” Indonesian Journal of Electrical Engineering and Computer Science vol. 18, no. 1, pp. 268–275, 2020.
16. H. C. Lee, P. W. Khong, and D. N. Ghista, “Bioenergy based medical diagnostic application based on gas discharge visualization,” Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings, vol. 7 VOLS, pp. 1533–1536, 2005
17. A. Dutta, S. Kour, and P. Jain, “Justifying Biofield (Aura) Studies as Complementary and Alternative Medicine (Cam)”, Annals of the Romanian Society for Cell Biology Journal, vol. 25, no. 4, 2021.
18. I. Ignatov, O. Mosin, and C. Stoyanov, “Biophysical Fields. Color Coronal Spectral Analysis. Registration with Water Spectral Analysis. Biophoton Emission,” An Open Access Journal, vol. 6, pp. 1–23, 2014.
19. R. S. S. A. Kadir, Z. H. Murat, M. N. Taib, and S. Z. A. Jalil, “Investigation of electromagnetics radiation for stroke patients and non-stroke participants,” Proceedings - 2015 6th IEEE Control and System Graduate Research Colloquium, ICSGRC 2015, pp. 130–134, 2016.
20. S. Z. A. Jalil, M. N. Taib, H. Abdullah, and M. Mohd Yunus, “A study of frequency radiation of human torso for gender classification,” 2012 International Conference on Biomedical Engineering, no. February, pp. 342–346, 2012.
21. S. Z. A. Jalil, M. N. Taib, H. A. Idris, and M. M. Yunus, “Classification of human radiation wave on the Upper body segment,” Proceedings - 2013 IEEE 3rd International Conference on System Engineering and Technology, pp. 73–77, 2013.
22. S. Z. A. Jalil, M. N. Taib, H. Abdullah, and M. M. Yunos, “Characteristic of human arm frequency radiation,” Progress in Electromagnetics Research Symposium, no. January, pp. 250–254, 2012.
23. J. Liu, G. Yang, F. Xu, and M. Zhou, “The Differences In Coordination Between Children With ADHD And Healthy Children Based On Two-Way ANOVA Analysis,” 2012 2nd International Conference on Consumer Electronics, Communications and Networks, CECNet 2012 - Proceedings, pp. 2972–2975, 2012.
24. C. W. N. F. Che Wan Fadzal, W. Mansor, Y. K. Lee, S. Mohamad, N. Mohamad, and S. Amirin, “Comparison between characteristics of EEG signal generated from dyslexic and normal children,” 2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences, IECBES 2012, pp. 843–846, 2012.
25. N. Fuad, W. Mansor, and K. Y. Lee, “Wavelet packet analysis of EEG signals from dyslexic children with writing disability,” Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pp. 2890–2893, 2013.
26. M. Rosdi, R. S. S. Ros Shilawani, Z. Hj. Murat, and N. Kamaruzaman, “The comparison of human body Electromagnetic radiation between Down Syndrome and Non Down Syndrome person for brain, chakra and energy field stability score analysis,” Proceedings - 2012 IEEE Control and System Graduate Research Colloquium, ICSGRC 2012, no. Icsgrc, pp. 370–375, 2012.
27. S. Kirchengast, “Gender Differences in Body Composition from Childhood to Old Age: An Evolutionary Point of View,” Journal of Life Sciences, vol. 2, no. 1, pp. 1–10, 2010.
28. B. Bontemps et al., “Children exhibit a more comparable neuromuscular fatigue profile to endurance athletes than untrained adults,” Frontiers in Physiology, vol. 10, no. FEB, pp. 1–11, 2019

Authors Introduction

Dr. Hjh Ros Shilawani S Abdul Kadir

She is working as a Senior Lecturer at the School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She obtaimed her PhD in Electrical Engineering UiTM, Biomedical Engineering, specifically Human Body Electromagnetic Radiation and other research related to human beings.

Dr. Aziati Husna Awang

She works as a Senior Lecturer at School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She received her PhD in Electrical, Engineering at Leeds University, UK. Her interests include THZ Detection, Photonic Sensors Optical Devices Microwave/RF Devices and IoT Technologies

Muhamad Azizularif Mohamad Azizan

He is currently serving as a protege of Tenaga Nasional Berhad (TNB) Klang. He holds a degree in Electrical Engineering majoring in Power from UiTM. His interests are electrical power engineering.

Dr. Suzi Seroja Sarnin

She works as a Senior Lecturer at School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She received her PhD in Electrical, Engineering at Leeds University, UK. Her interests include THZ Detection, Photonic Sensors Optical Devices Microwave/RF Devices and IoT Technologies

Ir. Ts. Dr. Suhaila Subahir

She is working as a Senior Lecturer at the School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She secured PhD in Electrical Engineering from UiTM. Her interests include antenna design, integrated antenna, RF devices and circuits, and electromagnetic radiation analysis

Dr. Roshakimah Mohd Isa

She is working as a Senior Lecturer at the School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She received PhD in Electrical Engineering from UiTM. Her interests include Biomedical Engineering, Brain Computer Interface and Signal Processing

Solar Powered Smart Parcel Box System: Energy Efficient Solution for Modern Deliveries

Wan Norsyafizan W. Muhamad, Amerul Ashraf Zulkepli, Nani Fazlina Naim, Suzi Seroja Sarnin, Ros Shilawani S Abdul Kadir, Md Nor Mat Tan

Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

E-mail: syafizan@uitm.edu.my, suzis045@uitm.edu.my, nanifadzlina@uitm.edu.my, ros885@uitm.edu.my, mohdn829@uitm.edu.my*

Abstract

Issues with existing smart parcel system using conventional power supply contribute to high energy consumption which lead to a significant carbon footprint and environmental pollution. To overcome the issue, this work introduces an innovative solution: the "Energy Efficiency Smart Parcel Box System Using Solar Energy." This system aims to revolutionize the parcel delivery and retrieval process by integrating smart technology and sustainable energy sources to enhance efficiency, security, and environmental sustainability. The proposed system consists of a network of smart parcel lockers strategically placed in residential and commercial areas. Each locker unit is equipped with solar panels to power its operations, QR code authentication, an infrared sensor for detection and automated door closure, and a user-friendly mobile application. Customers can conveniently receive and send parcels through a secure and automated process. This innovative methodology ensures an energy-efficient and eco-friendly operation by utilizing solar energy. Generally, literature and case studies have shown potential energy consumption reductions of up to 100% for certain periods under optimal conditions. However, for continuous and reliable operation, a combination of solar power and energy storage (batteries) is necessary. It also incorporates various advanced technologies, such as QR code scanning, infrared sensing, and GSM communication, to enhance security, automate processes, and provide a seamless user experience. The system's ability to operate independently using solar power makes it a cost-effective and sustainable solution for modern parcel delivery needs. This innovation aligns with the ongoing global efforts to create a greener and more sustainable future, providing a promising solution for the challenges faced in modern logistics and parcel management.

Keywords: smart parcel, solar power, energy efficiency, QR code, delivery system.

1. Introduction

Online shopping has indeed experienced significant growth, particularly during the pandemic, as it offers convenience and safety by allowing people to get their necessities delivered to their homes. However, the current or conventional parcel delivery system, while efficient in many ways, does have several limitations and challenges such as customer presence requirements, parcel theft, rescheduling delays, inefficient redelivery attempts, and lack of flexibility [1]. One of the main issues is leaving parcels unattended at doorsteps which leads to parcel theft. According to Penn Elcom's Global Package Theft Report 2022, more than one billion packages or roughly one in ten consumers suffered parcel theft or loss over the course of a year [2]. The US (13%), Australia (14%) and Canada (17% of the population) had the greatest rates of package loss, respectively, according to. One-third of consumers reported having had a package stolen or lost at least once in their lifetime, and all the regions surveyed saw an increase in parcel theft or loss between 2021 and 2022. The largest increase, however, was in the UK, which went from 7% between May 2020 and April 2021 to 12% for the same period the following year [3].

To address these problems, the proposed solution aims to introduce a new delivery modality that offers customers an alternative choice, reducing the need for customer presence and providing a more secure and flexible parcel delivery experience. By offering a range of delivery options that cater to different customer preferences, the goal is to improve customer satisfaction, enhance delivery efficiency, and minimize the impact of package theft on the overall online shopping experience [4]. The proposed solution involves implementing a smart parcel system using QR codes, an Arduino Uno GSM module, and mobile phones. Each courier or post is assigned a unique QR code, which is scanned by a QR reader installed in the mailbox. The scanned information is then transmitted via the GSM module to the authorized recipient's mobile phone, providing real-time notification of the parcel or mail arrival.

Implementing a smart parcel system using QR codes, an Arduino Uno GSM module, and mobile phones involves several steps and components such as hardware setup, QR code generation, database and authentication, parcel arrival and QR code scanning, data transmission and notification and finally recipient confirmation [5], [6]. Hardware setup consists of few modules which are the Arduino uno which is a microcontroller board that will be the brain of the smart parcel system. It will handle data processing, communication, and interaction with other components [7]. Following that, a QR code reader module, such as a camera or a dedicated QR scanner, will be connected to the Arduino Uno. It will read the QR codes placed on the packages. While the GSM Module which is an Arduino-compatible GSM module will be used for communication. It allows the system to send and receive data via mobile networks, enabling real-time notifications to the recipient's mobile phone. Finally, the mailbox will be modified to accommodate the QR code reader and Arduino Uno. The QR code reader should be positioned in a way that it can easily scan the QR codes on the packages when they are placed inside the mailbox. By following this methodology approach, a smart parcel system using QR codes, an Arduino Uno GSM module, and mobile phones can be successfully implemented, providing real-time parcel arrival notifications and enhancing the security and efficiency of the delivery process.

In the literature, there are few works that have been carried out by researchers in order to solve the delivery problem [8], [9], [10], [11], [12]. In Malaysia, Pos Laju Kiosk and Ezibox have been developed to secure the parcel delivery process. Furthermore, customers did not have to wait for the next delivery attempt or go to the post office to pick up their package when they were not at home. To pick up their package, they only need to travel to the nearby Pos Laju Kiosk or EziBox rather than the post office. In Malaysia, Pos Laju Kiosk and Ezibox are self-service parcel collection points developed by Pos Malaysia Berhad. The kiosk provides convenient pickup points for customers to collect their parcels when they are not at home during delivery attempts.

A smart packaging box was presented by Chaturvedi et al [13] that alerts the seller and the buyer when the package carrying the goods has been tampered with. This smart device consists of a circuit with a microprocessor, sensors, and the Global System for Mobile Communications. The work in [14] suggests a low-cost smart parcel box system with improved security that will be deployed at individual homes. All of the designed system's processes were controlled by an Arduino Mega 2560 in this system. The system will start when couriers message the user via applications with the tracking number for the package in order to obtain the password. When the message delivered by the courier matches the message that the user has specified, a password will be given for security reasons. Following that, couriers can insert the package into the smart parcel box by entering the specified password.

A mobile application and a smart personal delivery box make up DroParcel as presented by Alghfeli et al [15]. By scanning the QR-Code or bar code that is printed with the tracking number on the package's shipping label, users can gain access to the smart box. For the purpose of receiving parcel posts from the postal service, a smart box prototype has been designed as proposed in this work [16]. This system integrates IoT and solar energy technologies to offer services to a recipient in the interest of sustainable development. One of the main functions of this system is that when mail or package postings are dropped into the box, notifications or alerts are sent to the recipient using a mobile application called LINE Notify. A ParcelRestBox device equipped with an ESP8266 NodeMCU V3 microprocessor, an infrared sensor, and an Android mobile application [17]. The project deployment follows the Mobile Application Development Life Cycle (MADLC) approach as a general direction. The recipient interacts with the ParcelRestBox device using the ParcelRestBox smartphone application. The proposed system will track every package delivery, and send notifications to the ParcelRestBox mobile application.

The proposed “Energy Efficiency Smart Parcel Box System Using Solar Energy"aims to modernize parcel management by introducing an Energy Efficiency Smart Parcel Box System powered by solar energy. The core methods involve the strategic placement of smart lockers, harnessing solar power, incorporating intelligent technology, and enhancing security to achieve a more efficient, secure, and environmentally sustainable parcel management process.

2. Methodology

2.1 *System Architecture*

Fig. 1 shows the System Architecture of Energy Efficiency Smart Parcel Box System Using Solar Energy. Basically, the system will be initialized when the solar panel supply the electrical energy to the system. Before giving permission to open the door, the QR code scanner will start scanning the QR code on the parcel box. Next, door will be opened when the information on the QR code is tally with the information in the system. When the door is open, the parcel will be inserted into the parcel box. At the same time, the infrared sensor will detect the person in charge for delivery process. When the person in charge for delivery no longer detected, the door will be closed after few seconds. After that, GSM module will send the notification in SMS form to notify the user on the parcel delivery status. LCD screen will display all information given by the Arduino Mega 2560.

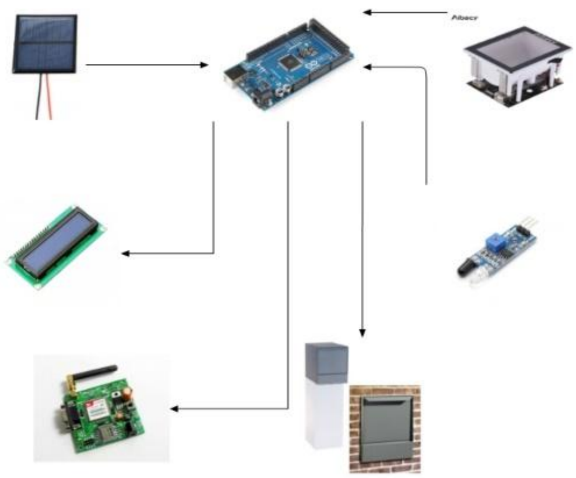


Fig.1 System architecture

2.2 *Flowchart*

Fig. 2 shows the flowchart of the Energy Efficiency Smart Parcel Box System (EESPBS) using Solar Energy. The system starts when the system is supplied with the electrical energy source from the sun. It is called as solar energy. As the PV module absorb the sunlight, it will convert the heat energy to electrical energy. So, that is how the system will be supplied with solar energy. Then, the QR scanner installed in the EESPBS which is function to scan the QR code on the parcel box will scan the QR code and send the data to Arduino Mega 2560. QR code is always in matrix form so the software inside the QR scanner will converts the dots inside the code into numbers or character strings. Then the data will be sent to the Arduino Mega 2560. Data received from the QR scanner will be analyzed and if the QR code address information is exactly the same with the QR data for current address in the Arduino Mega 2560. Then, Arduino Mega 2560 will be granted access to open the door to allow the parcel to be inserted into the box. There is no timer set for the door to be open or closed because IR sensor will detect the postman in front of the parcel box. If the postman is still standing in front of the parcel box, it will detect the obstacle in front of it. So, the data will be sent to Arduino Mega 2560 to execute an instruction for the sliding door of smart parcel box to keep open. When the postman is no longer standing in front of the parcel box, IR sensor will send the information to the Arduino Mega 2560. Hence, the sliding door of smart parcel box will be closed. At the same time, GSM module will be sending notification through SMS to the recipients of the parcel.

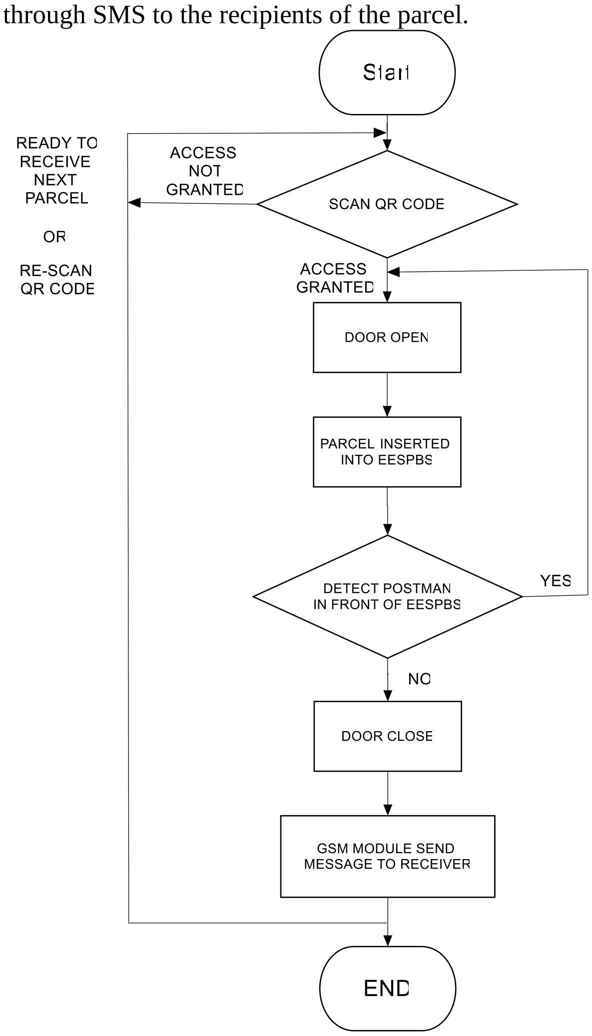


Fig. 2. Flowchart of the system

3. Results

Arduino system act as a brain for the smart parcel box. So, all operation that has been set to this box will not function if the Arduino system is not installed in this system or for the worst case is damage. That also will cause the system failed to operate. Fig. 3 shows the developed prototype of the Energy Efficiency Smart Parcel Box System using Solar Energy.

Fig. 3. Prototype of the Energy Efficiency Smart Parcel Box System

All solar cells are connected in parallel in order to generate 1 ampere of current. 1 ampere is the ideal amount of current that is needed to start up the whole system. As the system is switched on, all components used in the smart parcel box will automatically turn on. First, the LCD display operated well by displaying "Welcome, Please Scan QR Here" as shown in Fig. 4.

Fig. 4. LCD display

Later, the security of the smart parcel box have been checked by scanning two different QR code by using the QR scanner that have been attached to the smart parcel box. The first QR code is the correct QR code that will allow the door to open while the other QR code showed a wrong destination address which cannot be used to open the door of smart parcel box. The system used in this EESPBS is very unique because one QR code is specified only for one address. As shown in Fig. 5, the LCD displaying "Access Granted, Door Will Open". That indicate that the QR code that have been scanned by the scanner is the correct QR code which match the house address. Also, in the same figure, the LCD displayed "Access Denied, Wrong Address". One QR code is for one specific address. The reason why the system cannot grant permission for the postman to send the parcel box with the wrong QR code attached on it because that parcel belongs to another address.

Fig. 5. LCD display

Next is on how the door operated in the system. Access to open the door is depend on the system. As the system grant access to open the door, the door will slide to the right and ready to receive parcel. As shown in Fig. 6, the door is opened and waiting for the postman to deliver the parcel. The LCD will display "Opening, Please Wait" when the door is sliding. An infrared sensor also had been installed in the system. To make sure that the door always open when the delivery attempt occurs, this sensor will detect the present of postman in front of it. As the infrared sensor detect obstacle in front of him, it will cause the door to stay open until the postman leave the smart parcel box. When he left, there is no obstacle detected by the infrared sensor. So, the system will now close the door and indicating that the delivery attempt has been complete. The LCD will display "Closing, Please Careful" and Delivery Attempt Is Success", as in Fig. 6.

Fig. 6. LCD display

Following that, after the delivery attempt completed by the postman, the owner of the smart parcel box will receive a message via SMS. GSM module will send a message to the box owner that the parcel has been delivered successfully. Fig. 7 shows the message sent by the GSM module to the owner of the smart parcel box.

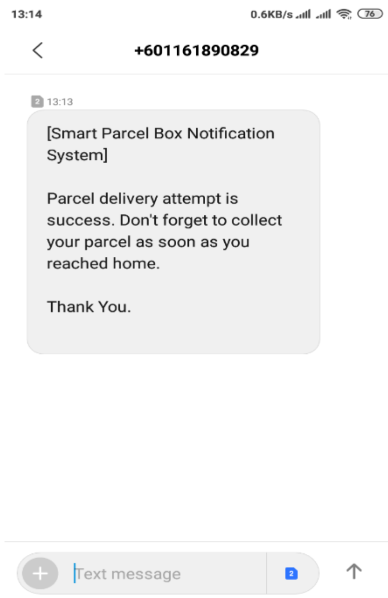


Fig. 7. Notification to user

As this smart parcel box use solar energy generated by small photovoltaic solar cell. It can reduce the amount of energy used from the direct electrical energy supplied by Tenaga Nasional Berhad (TNB) Malaysia. The electrical power used by this smart parcel box is fully comes from solar energy. Even though the energy used by smart parcel box is small, but the electrical energy used from TNB supply will also be considered as cost. So, the house that already installed PV module on their roof will have a lot of advantage than other house that did not use the photovoltaic module. Table 1 shows the total cost that need to be paid per kWh based on the domestic payment value from TNB.

Table 1 Total cost per kWh

Tariff Category	Unit	Current Rate
Tariff A – Domestic Tariff		
For the first 200kWh per month (1-200kWh)	Sen/kWh	0.218
For the next 100kWh per month (201-300kWh)	Sen/kWh	0.334
For the next 300kWh per month (301-600kWh)	Sen/kWh	0.516
For the next 300kWh per month (601-900kWh)	Sen/kWh	0.546
For the next kWh per month (901kWh onwards)	Sen/kWh	0.571
The minimum monthly charge is RM3.00		

4. Conclusions

This project presents an Energy Efficiency Smart Parcel Box System using Solar Energy which functioned to receive the parcel when there is no one at home. All modules and components have been integrated into the system to control the activity and process of this project. As for the security, the QR scanner just accept only 1 QR code which is the QR code that was set for the current address. On the safety part, there will be no other parcel from another address that can be inserted into the parcel box. Next, GSM module that have been integrated together with this system gave a message to the user of smart parcel box. The message will notify user on receiving of parcel. As the message was received by the user which means that the delivery attempt at the user house is a success. So, the user was not worried about the condition of their parcel because of the security and notification system that has been provided to this developed system. In terms of energy saving, this proposed system used the solar cell as an electrical supply to the Arduino system. Thus, it will give a lot of advantage to the user because solar energy is one of the renewable energies which will give benefit to user in reducing energy consumption. The total energy use in kWh will determine the total cost that user need to pay to the energy consumer. So, by using solar energy as energy source in activating the system, the user will not be charged by TNB because the user did not use the electrical energy provided by the TNB to supply energy to the system of smart parcel box.

For future improvements, the size of the door for this prototype should be bigger because it is too small. Only items that have size up to medium size of parcel packaging could be inserted into this parcel box. This is also due to the limited area that this door can open. Other than using a sliding door, it is also recommended for the smart parcel box to use rolling door to give more space to insert a larger type of parcel. Solar energy generated by this PV module also did not stored. So, when it is cloudy, it will be less efficient. Hence, it will affect the performance smart parcel box. Thus, for the future purpose this smart parcel box can store energy to overcome this kind of weather problem.

Acknowledgement

This study was funded by the School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA.

References

1. Jung-Fa Tsai, Hong Ngoc Ngo, Zhen-Hua Che, "Last-mile delivery during COVID-19: A systematic review of parcel locker adoption and consumer experience", Acta Psychologica, Volume 249, 2024
2. https://www.crresearch.com/blog/2020-package-theft-statistics-report/
3. "Addressing the issue of parcel theft", last accessed 10 May 2023."[Online].Available:https://www.parcelandpostaltechnologyinternational.com/features/addressing-the-issue-of-parcel-theft.html/
4. Mangiaracina R., Perego A., Seghezzi A., Tumino A, "Innovative solutions to increase last-mile delivery efficiency in B2C e-commerce: a literature review", (2019) International Journal of Physical Distribution and Logistics Management, 49 (9), pp. 901 - 920, DOI: 10.1108/IJPDLM-02-2019-0048
5. J. Z. Ooi and C. C. Tan, "Smart Modular Parcel Locker System using Internet of Things (IoT)," 2021 IEEE 11th International Conference on System Engineering and Technology (ICSET), Shah Alam, Malaysia, 2021, pp. 66-71, doi: 10.1109/ICSET53708.2021.9612542.https://ieeexplore.ieee.org/document/9612542
6. Yuk Ming Tang, Ka Yin Chau, Duo Xu, Xiaoyun Liu, "Consumer perceptions to support IoT based smart parcel locker logistics in China," Journal of Retailing and Consumer Services, Volume 62, 2021,102659, ISSN 0969-6989, https://doi.org/10.1016/j.jretconser.2021.102659.
7. S. Abdolinezhad, M. Schappacher and A. Sikora, "Secure Wireless Architecture for Communications in a Parcel Delivery System," 2020 IEEE 5th International Symposium on Smart and Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS), Dortmund, Germany, 2020, pp. 1-6, doi: 10.1109/IDAACS-SWS50031.2020.9297086.
8. A. A. Abdulkareem Alawsi, B. H. Jasim and S. M. Raafat, "Design and Implementation of a Global Smart Box for Quadcopter Delivery System," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp. 1408-1415, doi: 10.1109/ICISS49785.2020.9316012
9. N. Mahesh, P. Revathi, M. I. Riswan Ahamed, S. Sowndharyan and A. Vignesh, "Development of Automatic Postal Parcel Collecter with Pincode based Segregation," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 805-808, doi: 10.1109/ICOSEC54921.2022.9952085.
10. T. Gundu, "Smart Locker System Acceptance for Rural Last-Mile Delivery," 2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC), Kimberley, South Africa, 2020, pp. 1-7, doi: 10.1109/IMITEC50163.2020.9334107.
11. Bohao Ma, Yiik Diew Wong, Chee-Chong Teo, Parcel self-collection for urban last-mile deliveries: A review and research agenda with a dual operations-consumer perspective, Transportation Research Interdisciplinary Perspectives, Volume 16, 2022, 100719, ISSN 2590-1982, https://doi.org/10.1016/j.trip.2022.100719.
12. Hideyama S., Phung-Duc T., Okada Y, "Queueing Analysis of Home Delivery Services with Parcel Lockers", (2019) Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11688 LNCS, pp. 351 - 368, DOI: 10.1007/978-3-030-27181-7_21
13. Chaturvedi S.P., Mukherjee R., Yadav A, "GPS based Novel Approach for Secure Delivery of Online Purchased Items", (2022) INDICON 2022 - 2022 IEEE 19th India Council International Conference, DOI: 10.1109/INDICON56171.2022.10040092
14. Mohd Rusli A.L.B., Muhamad W.N.W., Sarnin S.S., Meor Hamzah M.M.A., "A Low-Cost Smart Parcel Box System with Enhanced Security", (2022) Proceedings of International Conference on Artificial Life and Robotics, pp. 1011 - 1018,
15. Alghfeli M., Alnuaimi M., Alsebaiha N., Alnuaimi S., Pradeep B., Kulkarni P., "DroParcel: Smart System for Secure Parcel Delivery", (2022) IEEE International Conference on Consumer Electronics - Berlin, ICCE-Berlin, 2022-September, DOI: 10.1109/ICCE-Berlin56473.2022.9937128
16. J. Kaewsrisuphawong, V. Waelun, J. P. N. Ayuthaya, S. Paengkanya and T. Daengsi, "Development of A Smart Box Prototype for Mail and Parcel Posts Using IoT and Solar Energy," 2022 5th International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 2022, pp. 77-81, doi: 10.1109/ICOIACT55506.2022.9972195.
17. M. Mokhsin, M. Z. M. Ludin, A. I. H. Suhaimi, A. S. Zainol, M. H. Mohd Som and H. A. Halim, "ParcelRestBox: IoT-Based Parcel Receiving Box System Design for Smart City in Malaysia," 2021 IEEE International Conference on Computing (ICOCO), Kuala Lumpur, Malaysia, 2021, pp. 180-185, doi: 10.1109/ICOCO53166.2021.9673588.

Authors Introduction

Dr Wan Norsyafizan W.Muhamad

She received her Bachelor in Electrical Engineering from the Universiti Malaya, Malaysia in 2002. She completed her Master of Electrical Engineering from Universiti Malaya, Malaysia in 2009. She obtained her Ph. D in 2017 from University of Newcastle, Australia. Currently, she is a Senior Lecturer in Faculty of Electrical Engineering at Universiti Teknologi MARA, Shah Alam, Malaysia. Her current research interests are in the area of wireless communication (Physical and MAC cross layer optimization).

Ir. Ts. Dr Nani Fadzlina Naim

She is working as a Senior Lecturer at School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She received PhD in Electrical, Electronic and System Engineering UKM, Bangi. Her interests include Optical sensor, Fiber Bragg Grating (FBG) based sensor, Optical Network.

Ir. Ts. Dr Suzi Seroja Sarnin

She is working as a Senior Lecturer at School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She received PhD in Wireless Communication UiTM, Master of Science in Microelectronic and Degree in Electrical Engineering UTM. Her interests include Wireless Communication and IoT.

Ts. Dr. Ros Shilawani S Abdul Kadir

She is working as a Senior Lecture at UiTM Shah Alam. she received PhD in Electrical Engineering in 2018. Her research interest includes Biomedical Engineering, specifically Human Body Electromagnetic Radiation and other research related to human beings.

Mohd Nor Md Tan

He is working as a Senior Lecturer at School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. He received Master in Radio Frequency Communication Engineering, Bradford University, Bradford, UK. His research themes include Antenna Design Technology.

Desktop-Based Expiry Date Application for Retailers Inventory Management

Nur Hazwani Ahmad Halil, Suzi Seroja Sarnin, Nani Fadzlina Naim, Azlina Idris, Wan Norsyafizan W Muhamad, Darmawaty Mohd Ali, Mohd Nor Md Tan, Mohd Rizal Dohad, Ros Shilawani S Abdul Kadir, Nur Syahirah Saipul
Universiti Teknologi MARA (UiTM) Shah Alam, 40450 Shah Alam, Selangor, Malaysia.

Zarina Baharudin Zamani
Universiti Teknikal Melaka, Durian Tunggal, Malaysia

E-mail: suzis045@uitm.edu.my, nanifadzlina@uitm.edu.my, azlina831@uitm.edu.my, syafizan@uitm.edu.my, darma504@uitm.edu.my, mohdn829@uitm.edu.my, ros885@uitm.edu.my, syahirah.saipul25@gmail.com, zarina@utem.edu.my

Abstract

Inventory management is a complex process that requires constant evaluation and control. It is essential for retailers to perform inspections regularly to ensure that all goods are available for sale and in good condition to meet customer expectations. While many retailers are still dependent on manual paper-based method, it poses a high risk of human error and causes negative impact on inventory management and eventually the business. It is not only labor-extensive, but also time-consuming. This project is proposed to automate inventory management process by developing a real-time inventory monitoring system and notify user on low stock items and items close to expiration. The desktop-based inventory management application is equipped with a PC and a handheld scanner and involve three main inventory processes; scan incoming inventory, scan outgoing inventory during customer checkout and allow users to manage stock counts and update data if necessary. The desktop application is developed using Windows Forms form as user interface and SQL Server as the database engine which enables the users to monitor stocks in real-time. As a result, users can use the desktop application to monitor inventory levels and receive notifications when items need to be restocked or nearing their expiration date. As different item categories may need different reminder durations based on expiry dates, users entering the correct password are able to update the minimum quantity and minimum duration for each item category for notification purposes. With this system, issues related to inventory can be resolved quickly based on real-time data monitoring, less time spent for repetitive works while providing a better customer experience.

Keywords: inventory management; desktop application; expiry data tracking.

1. Introduction

Inventory management has been widely studied as part of Industry 4.0 movement while helping retailers to manage businesses better. Inventory management is crucial for any retail industry be it huge organizations or small businesses. Retailers are expected to sell only good products and ensure products remain available for customers. A well-established organization has a team of inventory planners dedicated in managing and controlling inventory. However, small businesses do not have the luxury of establishing a team and dedicate only one person in managing stock incoming and stock checking manually at the end of the day. In fact, many retailers are still dependent on manual paper-based methods to inspect their inventory, which poses a high risk to inventory management. This is due to retailers often failing to notice low stock and expired items, which can lead to a negative customer shopping experience. While retailers often perform inventory checking periodically, this process is time-consuming when resources can be directed to a more prioritized task. As there is a possibility of human error in conducting inventory inspection, the consequences include food waste and eventually lesser revenues for retailers. This is due to unsold items having passed the expiry date without retailers' knowledge. In the European Union, nearly 57 million tons of food waste (127 kg/inhabitant) are generated annually with estimated market value at 130 billion euros [1]. 17 percent of food produced globally is wasted at the retail, food service, and consumer stages [2]. A lot of parties are accountable to address food waste issues, but the primary entity is the retailer. Furthermore, food waste is recognized as a setback in achieving Sustainable Development Goal (SDG) as it is mentioned in the United Nations 2030 Agenda for SDG where target 12.3 states "…to halve per capita global food waste by 2030 at the retail and consumer levels and reduce food losses along production and supply chains". The achievement of SDG 12 can be enhanced by the implementation of circular economy strategies based on the reduction of food waste [3]. With up to 60% of supermarket sales deriving from the perishable category [4], food retailers face an ongoing challenge in managing sell-by dates effectively and this matter can certainly be avoided with a good inventory management system.

In order to address the inventory issues discussed, this project is proposed to improve inventory management. With this project, retailers will be able to use the system to monitor inventory and reduce workload by automating repetitive tasks with less risk of mishandling inventory inspection. Real-time monitoring feature supports Industry 4.0 move by making the system agile, easily accessible and provide a summarized dashboard that is extremely beneficial for the user. Apart from improving operational efficiency, this project enables retailers to make informed business decisions, such as selling items nearing expiration at a discounted price, which ultimately reduces food waste.

1.1. *Review of Related Studies*

Many studies have focused on developing applications to automate inventory tracking. In [5], the author designed a web application for inventory management using full-stack framework RedwoodJS and TypeScript. Although extensive experiments were conducted, the possibility of an unstable internet connection was not considered. This is a high-risk factor that could prevent users from using the system, as web-based applications are heavily reliant on the internet. In [6], author conducts performance analysis of inventory management by developing a mobile PC-based stock management using Java for real-time inventory management and based on customers' orders. However, the author does not include product conditions or expiry date tracking and it works solely based on orders. Authors in [7] developed a desktop-based web application for production industries in managing inventory and keeping records. This paper presented a development using Model-View-Controller (MVC) architecture for administrators use. However, the scope of this project is limited to the production industry and does not include retailers.

There have been several significant studies on inventory expiry date tracking. In [8], authors focus on detection and recognition of expiry date on food packaging using deep neural network (DNN) approach where image processing takes place to record the expiry date printed. However, this paper does not discuss inventory management and data storage. In [9] and [10], author experimented on integrating IoT into food inventory tracking system where Arduino UNO microcontroller is used as the 'brain', load cell, MQTT broker, desktop application and Android application are used to display real-time data. The system simulated is for inventory tracking purpose only through weight sensing. Despite this, expiry date monitoring is not included in the mobile application. The application is suitable for inventory monitoring, but not when there is a large amount of data stored on the server, especially if the application is meant for retailers. In [11], authors studied on packaged food expiry tracking to reduce food waste. A web application is designed for users to enter products and the expiry dates. The data is stored in MongoDB as database and user will receive an SMS to notify any products close to expiring. The same concern arises when considering the condition where Internet condition is unstable when using a web application.

Based on the review, this paper shall focus on improving inventory management. To automate the inventory management process, a desktop-based application is proposed with SQL Server as the database. Desktop application is selected so user can always use the system despite unstable Internet connection. An alternative of local data storage is proposed, and data will be stored once connection to SQL Server is established again. Furthermore, a desktop application has the benefit of lower exposure to security risks, more robust and providing consistent user experience. With just one desktop application, users can perform many activities that include scanning incoming inventory and outgoing items by installing a scanner to the PC while monitoring inventory levels and receiving notifications of low stocks and expiring items. This is not only helpful for inventory inspection but also provides solution and ease of business decisions for retailers.

2. Methodology

2.1. *System Design*

Fig. 1 illustrates the proposed overall system design. Users can utilize the desktop application at an PCs after software deployment and connection to SQL Server is established. The application can be used whether during incoming inventory scan-in, at the cashier counter during customers checkouts or if any users require to monitor the stock levels. The system is equipped with a scanner for scanning processes

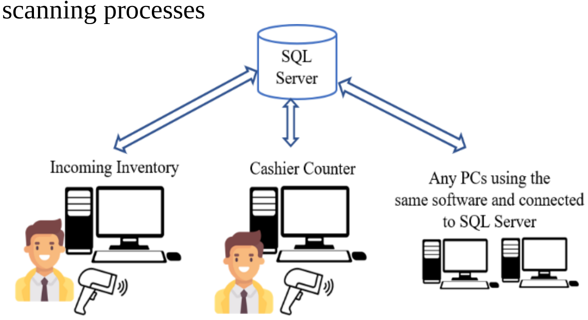


Fig.1 System Design.

The system must be capable of below processes (Fig. 1):

i. Scan-In:
   Users scan incoming inventory and enter all required data such as quantity, price, and expiry date. All data will be stored in SQL Server.
ii. Scan-Out:
   Users scan outgoing inventory at the cashier counter and system can calculate total price and quantity of each item during checkout. SQL will deduct the input quantity from inventory data.
iii. Notification:
   User will be notified through the application once low stocks detected and items nearing expiration based on items categories such as canned food and milk. The notification shall stop appearing once the user acknowledge and entered the correct password.
iv. Inventory Management:
   Users can monitor and update stock levels and data through the desktop application. As notifications are given based on categories, users can customize their notifications to fit their needs by specifying the minimum stock level, and the minimum duration until expiry date to receive notifications.

2.2. *Inventory System Flowchart*

Fig. 2 illustrates the sequence of the proposed system. Before any users can utilize the system, the user must install a scanner to the PC and connect to the Internet for server connection to SQL Server. Firstly, connection to SQL Server must be established with specified user login details and password. The system can be categorized into three processes. The first process is scan-in during incoming inventory. After connection is established, the user must be able to scan in item barcode for incoming stocks and enter required data such as item name, item category, quantity, expiry date, price. An alternative is needed in the case of unstable Internet connection that affects data storage activity. To cater for unsuccessful data logging to SQL Server, data is stored locally in .csv

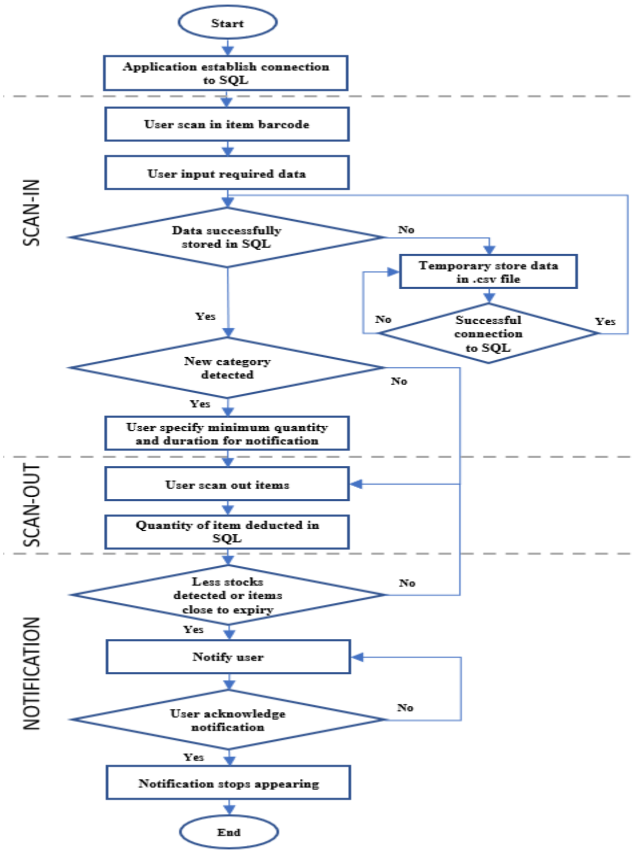


Fig.2 Overall Flowchart of Inventory System.

file and stored in SQL Server once connection is established again as the system is designed to continuously try to reconnect to database.

Secondly, scan-out process during checkout. Users can continue using the system at the checkout counter where the application can display prices and automatically deduct quantity in the database. Thirdly, notification for the users. When low stock or items close to expiry date are detected, an alert will continue to display on the application until the user acts. User must acknowledge for the notification to disappear. A display of inventory summary is helpful for inventory analysis and enables retailers to make business decisions.

2.3. *Overall Software Architecture*

This project uses three-tier architecture as shown in Fig. 3. The closest to client is called the presentation tier. This tier contains Windows Forms form as an interface between the user and the system, to display data and get input data from user. This tier is the most crucial as it plays the role of communicating with users through the interface. The interface must be planned and designed properly to ensure a great view of system as it mirrors how well-structured a system is. Despite data transfer can go through two-tier architecture from presentation tier to data tier, a business layer in three-tier architecture is vital in ensuring input and output data from data tier are validated. This layer prevents the presentation tier to have direct access to data tier, where an additional gate to verify input and output data are correct before proceeding to access data tier. This can avoid assertion errors from database as well as output to presentation layer. The data tier is where information is stored and retrieved from a database of file system. The information is then passed back to the business tier for processing and eventually back to the user. The three-tier architecture in this project consists of presentation layer using Windows Forms form programmed with Microsoft Visual Studio in .NET environment and C# language while data tier uses SQL Server.

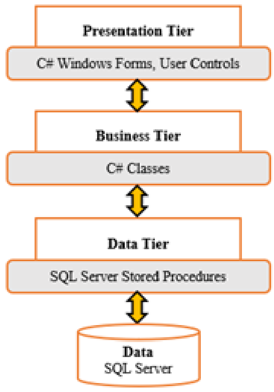


Fig.3 Three-Tier Architecture.

2.4. *Hardware and Software Setup*

2.4.1. *Hardware*

This project requires a handheld scanner for items barcode scanning and one PC for desktop application.

2.4.2. *SQL Server*

For storing and retrieving data, SQL Server is used as the database engine. This project is tested with a server and specifically created login ID and password. As a validation gate, stored procedures are created for data insertion, data update and constant checking of inventory status. Fig. 4 illustrates the software flowchart for notification purposes if user chose to enable notification.

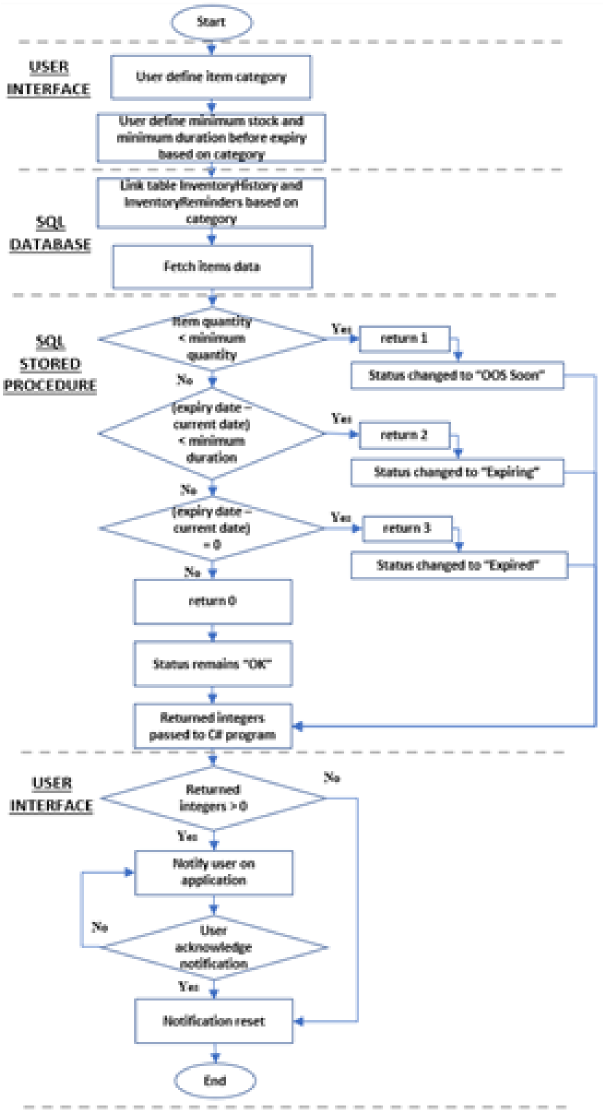


Fig.4 Software Flowchart for Notification.

2.4.3. *Microsoft Visual Studio*

Microsoft Visual Studio acts as the brain in this project as it is widely used for programming environment and the most popular framework among programmers. Some advantages include its capability to support multiple programming languages, Intelli-Sense feature that helps programmers to detect incomplete portions of the code, cross-platform support on Windows, Linux, or Mac Systems and most importantly the ease of use for multi-projects management. As for programming language, this project uses C# in .NET 4.0 framework. Fig. 5 shows the flowchart of how .NET C# works. As for interface, a Windows Forms form is created to display out data from SQL Server. The interfaces are for users scan incoming items, scan outgoing items, show summary of inventory and alert user.

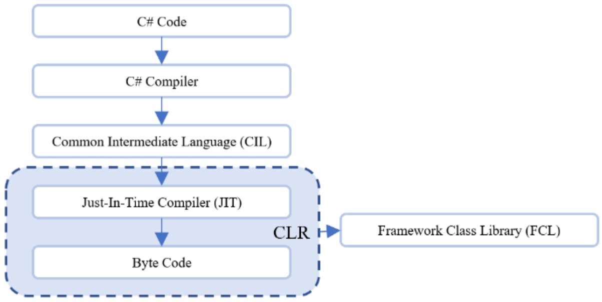


Fig.5 NET C# Structure.

3. Results

Through one desktop application, three tabs are created for users namely "Scan In", "Scan Out" and "Summary" to fulfil the retailer's needs.

3.1. *Scan In*

Fig. 6 shows the interface of scan in process during incoming inventory. Once the scanner is connected, the user can scan barcode and enter required data where all fields are not nullable.

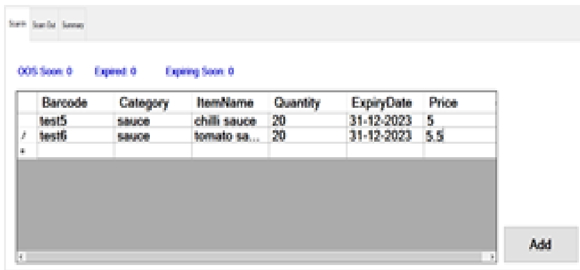


Fig.6 Scan-In Process

Two possibilities are expected when the user clicks the Add button. Fig. 7 shows the alert popup once items are stored successfully. Meanwhile, Fig. 7 shows a popup when the category entered is not found in history. User needs to define the minimum quantity and minimum duration for notification purposes. Otherwise, data will not be stored. In the case where Internet connection is unstable, data are alternatively stored in a .csv file until connection is established again. Fig. 7 shows the popup when application sensed connection to SQL is interrupted.

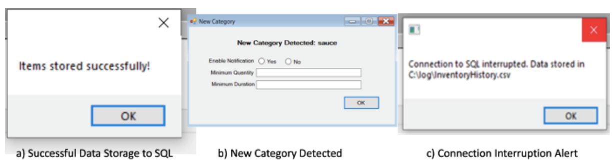


Fig.7 Scan in Possibilities.

3.2. *Scan Out*

The scan out process occurs during customers checkout at the counter. In this process, the important parameters are quantity and price. Fig. 8 shows the interface during scan out and item quantity in database will automatically be deducted once button Done is clicked.

Fig.8 Scan-Out Process.

3.3. *Notification*

The most crucial output of this project is the notification to the user. In each of the tabs created, a one-liner status is displayed so users will always notice the alert. Three categories of alert are “OOS Soon” indicating low stock, “Expired” for expired items and “Expiring Soon” for items nearing expiration. Fig. 9 shows the one-liner status displayed at the top and Fig. 10 shows a popup alert that displays at 10 a.m., 3 p.m. and 10 p.m. These timeframes are chosen based on the usual store’s operating hours.

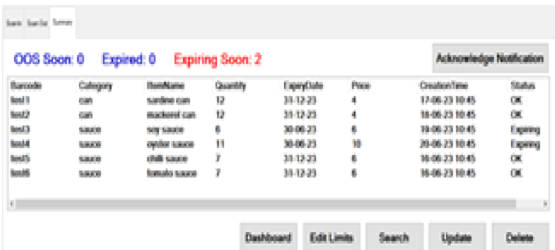


Fig.9 Summary View.

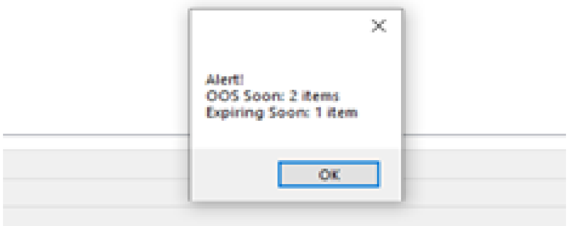


Fig.10 Notification Alert.

To ensure user acknowledge the notification and avoid the constant popup, user may enter the correct password once button “Acknowledge Notification” shown in Fig. 10 is clicked. Fig. 11 shows the password prompt box when any administrators activities initiated.

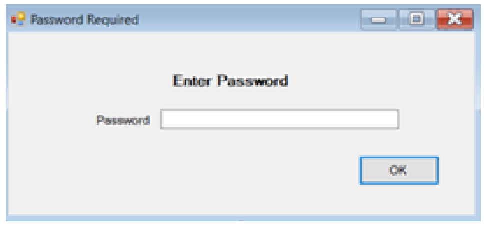


Fig.11 Password Prompt.

3.4. *Administrator Activities*

The same password prompt of Fig. 12 will pop up and correct password will allow the action to proceed. Meanwhile, the button “Edit Limits” is meant for user to edit the low stock and nearing expiration notification based on category. Fig. 12 shows the display when user clicks on “Edit Limits” button. Application will fetch existing categories from SQL and displayed in combo box. Users can decide to enable notification for the item categories and specify minimum quantity and duration until expiration for notification purposes. Considering the case where items remain idle in the inventory system, an alert will pop up for user to decide whether to delete the item from history or let it remain in the system. As for “Dashboard” button, user is capable to view items based on the overall inventory status. This will give the user an overall view of current stock levels and thus help the user to decide on next sales actions. Fig. 12 shows the dashboard view.

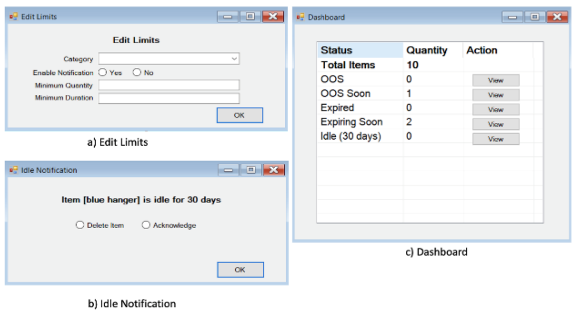


Fig.12 Administrator Activities.

This section has presented all the results gathered and discussion for each part of the system. Apart from monitoring quantities and expiry dates, the application provides retailers an overall view of inventory stocks with added features. Furthermore, a system connected to SQL allows users to monitor data in real-time.

4. Conclusion

In this project, a desktop-based application for inventory management is proposed. The inventory system aims to automate manual inventory management process, provide real-time data accessible in many hosts and notify users on low stock items and items close to expiration. The efficiency is evident as the application is tested and demonstrated all the scopes mentioned; to manage incoming and outgoing inventory while monitoring stock levels. For future developments, it is suggested that the dashboard view should have a user-friendly display such as bar charts summary display as the development in this project has limitation in terms of the .NET framework. Secondly, notification to users' mobile phone would be extremely helpful to alert users more prominently. Apart from that, a shelf based IoT system can be integrated into the system to have double validation between quantity at the shelf in comparison to quantity in database.

Acknowledgements

The authors would convey their profound appreciation and gratitude to the College of Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia for permitting us to perform this study. Special thanks to those who contributed to this project directly and indirectly.

References

1. Eurostat, “Food waste and food waste prevention – estimates” October 2022.
2. UNEP, "Planetary Action: climate, nature, chemicals & pollution," 2021.
3. United Nations, "The 2030 Agenda for Sustainable Development. Sustainable Development Goals," 2022.
4. B. Weber, S. Herrlein and G. Hodge, "PlanetRetail : The Challenge of Food Waste," 2022.
5. X. X. Ong, "Inventory management web application," Nanyang Technological University, Singapore, 2023.
6. H. K. Krishnappa and S. Abbigeri, "Performance analysis of Store Inventory Management (SIM) an Enterprise Application developed using Java," International Research Journal of Engineering and Technology (IRJET), vol. 6, no. 5, pp. 6725-6728, 2019.
7. P. G. Mehar, H. Umredkar, D. Vaidya, R. Somkuwar, A. Borkar, A. Pardhi, "Development of Inventory Management Systemin Production Industry," International Research Journal of Modernization in Engineering Technology and Science, vol. 5, no. 4, pp. 2528-2531, 2023.
8. L. Gong, M. Yu, W. Duan, Y. Xujiong, K. Gudmundsson and M. Swainson, "A Novel Camera Based Approach for Automatic Expiry Date Detection and Recognition on Food Packages," IFIP International Federation for Information Processing, pp. 133-142, 2018.
9. S. P. L. Narayan, E. Kavinkartik and E. Prabhu, "IoT Based Food Inventory Tracking System," CCIS, pp. 41-52, 2019.
10. A. S. Shelake, P. V. Sonawane, S. S. Ghumare, P. S. Deshmukh and P. R. Pingale, "IoT Based Food Inventory Tracking System for Domestic and Commercial Kitchens," International Journal of Scientific Research and Engineering Development, vol. 3, no. 1, 2020.
11. S. Bansal, R. Khokar and S. Chandran, "Best Before: An Expiry Date Checker Website to Track the Expiry of Perishable Items," in Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2022, New Delhi, 2022.

Authors Introduction

Nur Hazwani Ahmad Halil

She is working as a software engineer in a multinational company in Selangor, Malaysia. She graduated with MSc. in Telecommunication & Information Engineering from UiTM Shah Alam and Bachelor (Hons.) Electrical & Electronic Engineering from Universiti Teknologi PETRONAS. Her interests include embedded system, information engineering and software development.

Ir. Ts. Dr Suzi Seroja Sarnin

She is working as a Senior Lecturer at School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She received PhD in Wireless Communication UiTM, Master of Science in Microelectronic and Degree in Electrical Engineering UTM. Her interests include Wireless Communication and IoT.

Ir. Ts. Dr Nani Fadzlina Naim

She is working as a Senior Lecturer at School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She received PhD in Electrical, Electronic and System Engineering UKM, Bangi. Her interests include Optical sensor, Fiber Bragg Grating (FBG) based sensor, Optical Network.

Prof. Ts Dr Azlina Idris

She is working as a Professor in School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She received PhD in Wireless Communication. Her research interest includes Wireless Communications, Signal Processing, Multiple Access, Space Time Coding and Coding Theory, Internet of Things (IoT), 5G systems, Multiple input Multiple output (MIMO).

Ir. Ts. Dr Wan Norsyafizan W Muhamad

She is working as a Senior Lecturer at School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. Her interests include Wireless Communication.

Assoc. Prof. Ir. Ts. Dr Darmawaty Mohd Ali

She is working as an Associate Professor in School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She received her PhD of Wireless Communication, University of Malaya. Her research interest includes Wireless Access Technology, QoS Provisioning in Wireless Network.

Mohd Nor Md Tan

He is working as a Senior Lecturer at School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. He received Master in Radio Frequency Communication Engineering, Bradford University, Bradford, UK. His research themes include Antenna Design Technology.

Mohd Rizal Dohad

He is working as a lecturer at the School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. He received a Master of Science in Telecommunication and Information Engineering at UiTM and Bachelor of Engineering (Electrical) with Honors at UTM. Directly engaged in research projects based on the Internet of Things and Electronic Engineering.

Ts. Dr. Hjh Ros Shilawani S. Abdul Kadir

She is working as Senior Lecturer at School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. Her research interests include Biomedical Engineering, with a particular focus on human body electromagnetic radiation and other human-related studies.

Nur Syahirah Saipul

She holds a Bachelor's degree in Electronic Engineering and a Diploma in Electrical & Electronic Engineering from UiTM. Currently, she is pursuing an MSc in Electrical Engineering at UiTM Shah Alam. Her areas of interest include communication systems and Internet of Things project.

Ts. Zarina Baharudin Zamani

She is working as a Senior Lecturer at Fakulti Kejuruteraan Elektronik dan Kejuruteraan komputer, Universiti Teknikal Melaka, Durian Tunggal, Malaysia. She received Master in Microelectronic, UKM. Her interests include VLSI, IC Design, PCB.

Automatic Metal Debris Collection Robot for Laboratory Safety: A Review

Sophia Fahima binti Hapizan, Ja'aris Samsudin
Robotics and Automation in Electronic Engineering, Mara-Japan Industrial Institute, 43700 Beranang, Malaysia

Heshalini Rajagopal
Department of Electrical & Electronic Engineering, Mila University, No. 1, Persiaran MIU, 71800 Putra Nilai, Negeri Sembilan Darul Khusus, Malaysia

Takao Ito
Graduate School of Advanced Science and Engineering, Hiroshima University, Japan

Muhammad Amirul Aiman Asri
Department of Electrical Engineering, Universiti Malaya, Malaysia

Anees ul Husnain
Faculty of Engineering, The Islamia University of Bahawalpur, 63100 Pakistan

Email: sophiafahimahapizan22@gmail.com, norrimamokhtar@um.edu.my, heshalini.rajagopal@mila.edu.my

Abstract

Laboratory safety is paramount, especially when managing hazardous metal debris. This review explores advancements in robotics for waste management, emphasizing autonomous sorting and collection systems driven by AI and machine learning. These technologies enhance the precision and adaptability of robots, enabling accurate detection, classification, and handling of metal debris. By integrating advanced sensors and real-time decision-making, such systems improve resource management and safety in laboratory environments. Challenges such as power efficiency and scalability are also discussed, highlighting future opportunities for optimizing robotic solutions in critical waste management applications.

Keywords: Laboratory Safety, Metal Debris, Robotics, Waste Management, Machine Learning, Sensors, Resource Management.

1. Introduction

Laboratories are environments where precision and safety are paramount whether in industrial or research settings such as addressed by Y. Yang et al. [1]. However, the presence of metal debris generated during experiments, fabrication processes or equipment handling can pose significant risks. Such debris may result in physical injuries to personnel, contamination of sensitive experiments, and potential damage to expensive equipment that results in costly repairs or interruptions to critical research. Maintaining laboratory safety is not only about compliance with standards but also creating an environment where experiments can proceed without hazards.

Often, traditional methods of metal waste management often rely on manual handling such as brooms, or manual vacuum systems. While these methods are straightforward, they are labor-extensive, time-consuming, and susceptible to human error, particularly in larger or more complex facilities. In high-tech laboratories, the need for innovative solutions is crucial to address these challenges whereas traditional methods often fall short of meeting the rigorous standards required for operational efficiency and safety. Fig. 1 shows some metal dust after work either in an industrial site or research laboratory, when swept with a broom.

The integration of automation and robotics in laboratory safety has emerged as a revolutionary solution to overcome these challenges. Robots designed for metal debris collection combine advanced technologies such as efficient power systems, smart sensors, and autonomous navigation. These robots are engineered to reduce energy consumption, making them cost-effective and environmentally sustainable. Power-efficient designs, such as low-power processors and optimized motor systems, ensure that these robots can operate for extended periods without frequent recharging, making them ideal for continuous or large-scale applications.

Safety and sustainability are driving forces behind the adoption of these automated solutions. The ability of robots to autonomously detect, collect, and safely dispose of metal debris eliminates the need for manual intervention, thereby reducing the risks of injuries. Furthermore, these robots minimize the contamination risks, especially in laboratories handling sensitive materials or conducting critical experiments. Their consistent performance ensures that no debris is overlooked, enhancing the overall safety and cleanliness of the workspace.

Fig.1 Metal Work Processes [2].

Robots for metal debris collection are equipped with cutting edge technologies:

- Sensors for Metal Detection: Advanced inductive or magnetic sensors allow the robots to identify metal debris accurately, even in mixed material environments. These sensors are designed to work efficiently while consuming minimal power.
- Magnetic Systems for Collection: Electromagnets or permanent magnets are used to collect metal debris such as in Fig. 2. Power-efficient electromagnets can be selectively activated, ensuring energy is used only when necessary.

Fig.2 12V Electromagnet that Can Attract Ferrous Metals Based on Current Input.

- AI and Autonomous Navigation: Artificial Intelligence enables robots to map laboratory layouts, navigate obstacles, and optimize cleaning paths. Machine learning algorithms enhance their ability to adapt to changing environments or debris patterns, ensuring thorough and efficient operations.

The adoption of such automated systems is especially critical in laboratories where human safety and operational precision are non-negotiable. For example, in facilities working with hazardous materials or intricate instrumentation, the introduction of robots not only safeguards personnel but also ensures that research activities are conducted without disruption. This review explores the advancements in robotics and automation technologies for waste management, particularly focusing on their application in laboratory safety. It examines the technological components that enable efficient and intelligent metal debris collection, including power-efficient designs, sensor technologies, and AI-driven navigation. By providing insights into these innovations, this review aims to shed light on how laboratories can enhance safety, improve resource management, and reduce operational risks through the adoption of automated solutions.

2. Key Research Themes

2.1 *Metal Detection and Collection*

The accurate detection and efficient collection of metal debris are essential for ensuring laboratory safety. Numerous studies have explored the use of sensors, such as inductive proximity sensors, metal detectors, and vision systems to identify metallic objects with high precision.

These technologies are particularly suited for laboratory environments, where debris may vary in size and composition. Research has further emphasized magnetic collection systems, noting their ability to attract and retain metallic fragments of various sizes with high efficiency. For instance, Gundupalli [3] highlights the strength of modern magnetic systems and their role in automating waste collection. Such technologies enable safe handling of sharp and hazardous debris, reducing risks for laboratory personnel.

2.2 *Navigation and Obstacle Avoidance*

Autonomous navigation is a critical component for robots operating in confined and cluttered environments like laboratories. Studies have demonstrated the effectiveness of AI-based techniques, such as machine learning algorithms and Simultaneous Localization and Mapping (SLAM), in enabling robots to map their surroundings, detect obstacles, and plan optimal paths. For example, P.Valsalan et al. [2] showed that efficiency of a camera-based system robot from Fig. 3 that takes an image of the room (area of interest) then identifies the metallic scraps: the robot calibrates the coordinate locations of all the scraps. With Scale Invariant Function (SIFT) and Nearest Neighbor Search (NNS) algorithms the robot is able to calculate and provide coordinate position of scrap locations. This is particularly relevant for laboratory applications, where precision in movement is crucial to avoid disrupting sensitive experiments or equipment.

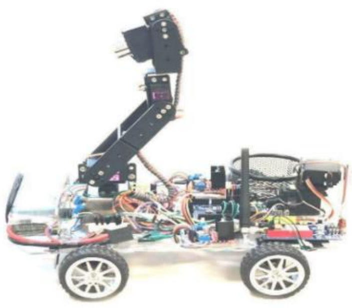


Fig.3 Valsalan's Robot Design for Metallic Scrap Collection Robot [2].

Another example of AI-based techniques robots that use path planning and obstacle avoidance in dynamic environments for cleaning robots is shown by S. Khanna and S. Srivastava [4] that addresses limitations caused by traditional cleaning robots that often rely on predetermined paths and struggle to adapt to changes in environment. To overcome this, they presented a design using advanced path planning algorithms and neural network models for path optimization, and a performance evaluation. Such an approach is important to enhance the adaptability and efficiency of cleaning robots, enabling them to dynamically respond to real-time environmental changes, avoid obstacles effectively, and optimize cleaning paths. This not only improves operational performance but also ensures that robots can function seamlessly in unpredictable and complex environments like laboratories, where maintaining safety and precision is critical.

2.3 *Design and Structural Considerations*

The physical design of a debris collection robot significantly impacts its performance in specialized environments. Compact and modular designs are often highlighted in the literature for their ability to navigate narrow lab spaces and adapt to various tasks. Storage compartments for collected debris are also emphasized, ensuring efficient handling and disposal processes. Case studies, such as those focusing on robots designed for hospitals or manufacturing, provide valuable insights into how design principles can be tailored for laboratory safety applications such as from 5. H. Alsahafi and M. Almaleky [5]. These examples demonstrate the adaptability of robotic systems for different operational needs.

2.4 *Efficiency and Sustainability*

Efficiency and sustainability are key metrics for evaluating the performance of automated cleaning robots. Research highlights the importance of optimizing speed, accuracy, and energy consumption to enhance the robot's effectiveness while minimizing operational costs. Furthermore, the environmental benefits of automated systems are discussed in terms of recycling and waste segregation. For instance, A. Jha et al. (2019) [6] reviews the contribution of autonomous garbage collector robots to resource conservation and pollution reduction. By integrating energy-efficient components and focusing on sustainable practices, these systems align with broader goals of environmental responsibility.

3. Applications and Case Studies

3.1 *Existing Implementations*

Robots designed for cleaning and debris collection are already being deployed in industrial and laboratory settings, showcasing their effectiveness in improving safety and efficiency. For instance, automated cleaning roots like HR-Recycler system have been adapted for waste sorting, including metallic components in industrial facilities A.Axenopoulus et al. [7]. The system integrates hybrid human-robot collaboration to ensure accurate separation of hazardous waste, demonstrating their potential for laboratory applications where safety is paramount. Manual, expensive, hazardous and time-consuming tasks of materials pre-processing will be substituted by automatic robotic-based procedures before the materials enter the shredding machine.

In laboratory settings, robots equipped with basic cleaning tools and sensors have been utilized for tasks such as biohazard material handling and spill cleanup such as introduced by H.Prendinger et al. [8]. While these robots may not specifically target metal debris, their success highlights the feasibility of adapting such systems for metal collection. Furthermore, studies have demonstrated the application of magnetic collection systems integrated into robots explained by J.J.Abbott et al. [9], which have proven effective in scrapyards and recycling facilities and can be scaled down for lab use.

3.2 *Innovative Prototypes*

Recent advancements have introduced prototypes tailored for metal debris collection in hazardous and confined environments. For example, S. Khanna and S. Srivastava [4] developed an AI-driven cleaning robot with advanced path planning algorithms and obstacle avoidance capabilities. Their design demonstrates how intelligent systems can optimize cleaning operations in dynamic and unpredictable environments.

Another notable prototype is the Metallic Scrap Collection Robot by P. Valsalan et al. [2], which uses electromagnets and inductive sensors to detect and collect ferrous metals. This system has been trialed in industrial contexts and offers insight into how such designs could be adapted for laboratories. Its compact design and efficient energy management system make it particularly suitable for spaces with limited mobility.

Additionally, research projects have explored robots with modular designs, allowing them to adapt to specific laboratory conditions. These prototypes often include features like interchangeable tools for handling different types of debris and Internet of Things (IoT) systems for remote monitoring and control. Such innovations underline the growing versatility of automated solutions in addressing safety and operational challenges in laboratory environments.

By examining these existing implementations and emerging prototypes, this review highlights the potential for further advancements in robotic solutions for metal debris collection, paving the way for safer and more efficient laboratory operations.

3.3 *Design Suggestion*

The proposed design for an automatic metal debris collection as shown in Fig. 4 is developed to enhance laboratory safety by automating detection, collection, and management of hazardous metallic waste. This system integrates advanced components such as proximity inductive sensors, electromagnets, camera module paired with microcontroller to enable efficient and reliable operations.

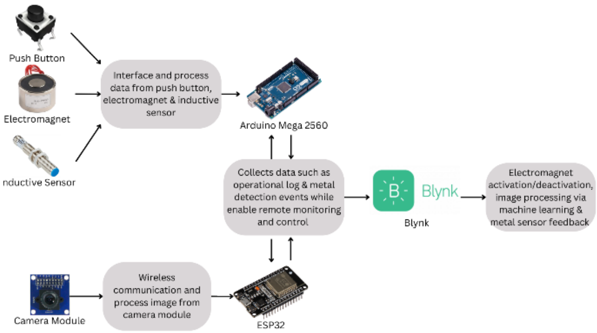


Fig.4 Proposed Design for Metal Debris Collection Robot.

The design aims to address the limitations of traditional manual debris handling methods. By incorporating such features such as autonomous navigation, image processing for precise metal identification, and IoT capabilities for remote control and monitoring, this robot ensures a seamless and effective approach to waste management in laboratory settings.

Moreover, the proposed system is tailored to operate in dynamic environments, leveraging modern robotics technologies and sustainable power solutions to optimize energy usage while maintaining performance.

4. Challenges and Future Directions

4.1 *Current Limitations*

Despite significant advancements, several challenges persist in the development and implementation of automated metal debris collection robots:

- Material Differentiation: One of the major limitations is the difficulty in accurately differentiating between metallic and non-metallic materials, especially when debris consists of mixed or coated substances. Current systems often rely on magnetic detection, which may fail to distinguish between various material types.
- Navigation in Dynamic Environments: Laboratory environments are often crowded and dynamic, with constantly changing layouts and obstacles. Ensuring reliable navigation and obstacle avoidance in such conditions remains a significant technical challenge for robotic systems.
- Maintenance and Durability: Robots operating in harsh laboratory conditions, such as exposure to corrosive chemicals or abrasive materials, face issues with durability and maintenance. Regular wear and tear of components like sensors and electromagnets can reduce operational efficiency over time.

4.2 *Research Gaps*

Several areas require further exploration to improve the effectiveness of automated metal debris collection robots:

- Multi-functional Robots: There is a need for robots capable of handling both metallic and non-metallic waste to provide comprehensive cleaning solutions. Integrating advanced sensors and adaptive tools could enable robots to address this challenge.
- Energy Efficiency: While energy-efficient designs are emerging, optimizing power consumption, especially in portable robots with limited battery capacity remains an area of active research.
- Scalability for Laboratory Use: Many existing designs are tailored for industrial applications. Adapting these systems to the compact and intricate nature of laboratory environments requires further innovation.

4.3 *Emerging Trends*

The field is witnessing exciting trends that promise to address current limitations and expand the capabilities of robotic systems for laboratory safety:

- IoT Integration: The incorporation of IoT technology allows for real-time monitoring, remote control, and data-driven optimization of robot operations. Such systems enable better oversight and improved decision-making in dynamic lab environments.
- AI and Machine Learning Advancements: The use of AI for enhanced material recognition, path planning, and adaptive learning is a transformative trend. Machine learning algorithms can enable robots to become more efficient and autonomous over time, adapting to specific lab conditions.
- Miniaturization of Robots: The development of compact and lightweight robotic systems is particularly important for laboratory settings. Miniaturization designs enable robots to navigate narrow spaces, access hard-to-reach areas, and operate with greater precision.

By addressing these challenges and leveraging emerging trends, future robotic systems can significantly enhance laboratory safety and efficiency, setting a benchmark for automated waste management solutions in critical environments.

5. Conclusion

This review highlights the advancements and challenges in the development of automated robots for metal debris collection in laboratory environments. The reviewed literature emphasizes significant progress in key areas, such as material detection using advanced sensors, navigation through dynamic environments powered by AI-based techniques, and the integration of magnetic systems for efficient debris collection. Additionally, trends like IoT-enabled monitoring, machine learning for enhanced autonomy, and compact robotic designs are crucial to revolutionizing laboratory safety protocols. The importance of developing robust, efficient and sustainable robotic solutions cannot be overstated, particularly in laboratory settings where precision and safety are utmost. Automated systems not only mitigate risks posed by hazardous metal debris but also improve operational efficiency and reduce manual intervention. Such advancements add to creating safer and more sustainable working environments for researchers and personnel.

References

1. Y.Yang, G. Reniers, G. Chen, F. Goerlandt, "A Bibliometric Review of Laboratory Safety in Universities," Safety Science, vol. 120, pp. 14-24, Dec 2019.
2. P. Valsalan, N. U. Hasan, I. Baig, "Metallic Scrap Collection Robot with Efficient Trajectory," International Journal of Recent Technology and Engineering (IJRTE), vol. 8, issue 5, pp 2-6, Jan 2020.
3. S. P. Gundupalli, S. Hait, A. Thakur, "A Review on Automated Sorting of Source-separated Municipal Solid Waste for Recycling," Waste Management, vol. 60, pp 56-74, Feb 2017.
4. S. Khanna, S. Srivastava, "Path Planning and Obstacle Avoidance in Dynamic Environments for Cleaning Robots," Quarterly Journal of Emerging Technologies and Innovations, vol. 8, June 2023.
5. H. Alsahafi and M. Almaleky, "Design and Implementation of Metallic Waste Collection Robot," in ASEE 2014 Zone I Conference, University of Bridgeport, Bridgeport, CT, USA, 2014, pp. 1-6.
6. A. Jha, A. Singh, R. Kerketta, D. Prasad, K. Neelam, V. Nath, "Development of Autonomous Garbage Collector Robot," Proceedings of the Third International Conference on Microelectronics, Computing and Communications Systems, vol. 556, pp 567-576, May 2019.
7. A. Axenopoulos, G. Th. Papadopoulos, D. Giakoumis, I. Kostavelis, A. Papadimitriou, S. Sillaurren, "A Hybrid Human-Robot Collaborative Environment for Recycling Electrical and Electronic Equipment," in 2019 IEEE SmartWorld, Leicester, UK, 2020.
8. H. Prendinger, N. Alvarez, A. Sanchez-Ruiz, M. Cavazza, J. Catarino, J. Oliveira et al., "Intelligent Biohazard Training Based on Real-Time Task Recognition," ACM Transactions on Interactive Intelligent Systems (TiiS), vol. 6, issue 3, pp. 1-32, Sept 2016.
9. J. J. Abbot, E. Diller, A. J. Petruska, "Magnetic Methods in Robotics," Annual Review of Control, Robotics and Autonomous Systems, vol. 3, pp. 57-90, May 2020.

Authors Introduction

Ms. Sophia Fahima Binti Hapizan

She is currently pursuing study in Electronics Engineering at Mara-Japan Industrial Institute, MJII.

Mr. Ja'aris Samsudin

He is the head of department of Robotic and Authomation at Mara-Japan Industrial Institute, MJII. His research interests include Microcontrollers and Pneumatic & Hydraulics.

Dr. Heshalini Rajagopal

She received her PhD and Master's degree from the Department of Electrical Engineering, University of Malaya, Malaysia in 2021 and 2016, respectively. She is an Assistant Professor in the School of Engineering and Computing, Mila University. Her research interests include image processing, artificial intelligence and machine learning.

Dr. Takao Ito

He is Professor of Management of Technology (MoT) in Graduate School of Engineering at Hiroshima University. He is serving concurrently as Professor of Harbin Institute of Technology (Weihai) China. He has published numerous papers in refereed journals and proceedings, particularly in the area of management science, and computer science.

Mr. Muhammad Amirul Aiman Bin Asri

He received the B.E (Electrical) from the Department of Electrical Engineering, University of Malaya, Malaysia in 2020 Currently, he is a Research Assistant in University of Malaya, Malaysia. His research interest includes image processing, artificial intelligence and machine learning.

Dr. Anees ul Husnain

Anees received his Doctoral in Electrical Engineering at University of Malaya, Malaysia. He holds a master's degree in computer engineering from UET Taxila, Pakistan. He is working on autonomous path generation of UAVs to monitor fugitive emissions. Currently, he is an Assistant Professor at Faculty of Engineering, The lslamia University of Bahawalpur, Pakistan.

Smart Solar LED Street Light with ESP32 Camera Module

Nurul Farhana Zailani, Suzi Seroja Sarnin, Mohd Nor Mad Tan, Nani Fadzlina Naim, Mohd Rizal Dohad, Ros Shilawani S. Abdul Kadir, Nur Syahirah Saipul

School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA (UiTM) Shah Alam, 40450 Shah Alam, Selangor, Malaysia.

Zarina Baharudin Zamani

Fakulti Kejuruteraan Elektronik dan Kejuruteraan komputer, Universiti Teknikal Melaka, Durian Tunggal, Malaysia

E-mail: suzis045@uitm.edu.my, mohdn829@uitm.edu.my, nanifadzlina@uitm.edu.my, syahirah.saipul25@gmail.my, zarina@utem.edu.my

Abstract

In 2016, there were 19,299 reports of street crime nationwide, according to the police. The most cases were in Selangor, Kuala Lumpur, Johor, and Perak with an increase in cases over 2013 in Kelantan, Negeri Sembilan and Perlis. The objective of this study was to determine how smart streetlights can adapt to the environment, which helps people today with their ability to observe their surroundings clearly while avoiding the existence of places where criminals can hide. This project is designed based on LDR sensor to turn ON the LED depends on light intensity while PIR sensor to control the brightness and microcontroller used is Arduino Uno. Furthermore, this project will also introduce the method to generate energy by using solar power that will store all the voltage in the battery. To ensure the safety of people at the surrounding, ESP32 camera module and emergency button are provided, and this will help to monitor the place and avoid any crimes that could possibly happened. The outcome from this project can be seen by the LDR sensor will turn ON the LED if the intensity of the light is less than 400 while the PIR sensor will increase the brightness of LED to 100% if there is motion detected but if there is no motion detected the brightness will remain 30%. Moreover, the push button will work as emergency button and when user press, it will turn on the buzzer and on the same time send the notification to the person in-charged. On the other hand, the ESP32 cam will display the live video through blynk apps. Automation is intended to reduce manpower with the help of intelligent systems, since the supply of electricity is limited for various of reasons, power consumption is always a top priority.

Keywords: LDR sensor, PIR sensor, solar energy power, ESP32 camera module, Arduino Uno.

1. Introduction

The world may have been exposed to environmental issues like climate change due to the design of the earlier conventional street light systems, which may have neglected carbon dioxide (CO_2) emissions. According to studies, LED lights generate significantly less heat than incandescent bulbs. In contrast to incandescent bulbs, which release 90% of their energy as heat, LEDs emit very little heat. LEDs also exert some degree of control over the light they produce. This suggests that LEDs do not need reflectors and diffusers to capture light. Since the light never exits the fixture, other lighting designs may waste 50% of it [1].

One of the technologies that enables the city to become a smart city is energy-saving street light systems. It is also one of the routes that gets the globe ready for a futuristic one. A study demonstrates that it is possible to produce electricity using solar energy. The PV cells on a solar panel capture the energy from the sun's rays as it shines on the panel. In reaction to an internal electrical field within the cell, this energy generates electrical charges that move, which results in the flow of electricity.

Public safety has been the main concern to the government. Many criminal activities, such as robbery, murder, and chain snatching occur in open spaces like the streets. There is minimum safety for women to stroll along in the street at night. Thus, to avoid any crime from happening, an emergency button is provided at each streetlight [2]. This show that security sensor that include in streetlight can ensure the safety of the people of the surrounding by using nearby lamp posts as "gateways," the networked streetlights can instantly communicate the pertinent, real-time photos back to the control system, enabling law enforcement officials in responding to any occurrences right away. Smarter and more extensive surveillance is improved public safety by implementing smart street lighting systems [3].

1.1. *Problem Statement*

Nowadays, a significant amount of energy is used by streetlights, which undoubtedly results in energy waste. Most nations use standard street lighting, which simply turns on and off depending on the timer that had been set, and which often uses fluorescent bulbs that use more energy. This really cause a lot of energy wastage as there is some place that not fully occupied like the main city street and sometimes in certain period the road will be empty. For example, the standard streetlight will alight all night along the highway although there is less route user during that time, and this will cause power consumption is relatively high.

In addition, the problem that led to this project is because of shortage of workers. Some street is using the workers to turn on and off the streetlight manually. In this case if the workers have lack of responsibilities which make the streetlight to alight all day, and this consumes large amount of electricity. This is why smart streetlight is useful as it is using LDR sensor to turns on and off based on the light intensity.

Moreover, by using standard streetlight is more expensive and can cause fortune. When purchasing equipment, a major factor to consider is the quality of the street lighting. Mostly the streetlight requires frequent maintenance and replacement and if the cost is too high it is difficult for the maintenance fee. This is the reason why this project is using LEDs as it consumes substantially less energy than incandescent bulbs because diode light is far more efficient, power-wise, than filament light. Compared to incandescent lights, LED bulbs consume more than 75 percent less energy. Besides, LED can continue to shine even with weak or reduced current flowing through it, LED is more efficient.

Furthermore, the sudden increase in crime that affect the passer-by especially during nighttime. Most manual streetlight do not include safety equipment, and this is dangerous especially at the place where there is not crowded as it can lead to crime. For example, if user have been followed by unfamiliar people it is hard for them to ask for help and this can cause major problem and make them feel unsafe to walk in the street especially for woman. Thus, crime prevention through this system is necessary as it can avoid any serious problem in the future.

1.2. *Background of Study*

The invention of artificial lights has made everyday life so much easier. Almost every home, workspace, school, and street are now well lit even during the night-time making it so much safer for the public to wander around. Ultimately, the need for electricity worldwide has increased exponentially as it contributes to 19% of the worldwide electricity consumption [4],[5] and accounts for nearly 5% of global CO2 emission [6]. Considering these data, it is now more important than ever to protect the limited resources while being energy and cost efficient by implementing a more self-sufficient system that do not harm the environment. Amidst the process of modernizing the lighting system at public spaces, a smart pathway and garden lighting system can be opted where it can be turned on and off on its own as well as adjusting the brightness according to the surrounding. This way, the system will be more effective and energy efficient. In fact, the idea of street lighting control system has appeared a long time ago due to the hassle of managing large number of streetlamps. A new technology was necessary so that the lamps could be grouped in smaller numbers and controlled more proficiently in terms of switching on and off along with maintenance. The reformation of street lighting system was done by designing and utilizing the feeder pillars and power cabinet. This allowed each light section to be turned on manually which were later outfitted the controllers that replaces human involvement by serving as an automation for turning on and off the lights.

A few papers studied have already provide more advanced solution to reform the street lighting system [7],[8],[9].An author from previous study have suggested that an Ethernet-based communication interfaced to a multiple-phase digital control driving system to implement a smart light system with solar energy power system. Though this system is energy efficient, it is also very expensive to be installed and maintained.

1.3. *Objectives*

Multifunctional gadgets have been made possible by modern technology and computers are now more potent, portable, and quick than ever. Technology has advanced, expedited, and enhanced our lives along with all these developments. Thanks to technological improvements people are looking for more sensible methods of doing things than manual and mechanical process. The objective of this project is:

i. To design a smart light system which targets the energy saving and autonomous operation on economical affordable for the sidewalk.
ii. To implement an intelligent light system with sensors that can ensure the safety of user.

2. Methodology

This chapter will go over the block diagram and flowchart of this project. The process of the smart solar LED streetlight with an ESP32 camera will be thoroughly analyzed in this chapter. The block diagram in figure 2.0 described how this project functions based on the ESP 32 and Arduino Uno microcontroller and the notification received through a Blynk App. On the other hand, there is also a schematic diagram for the project for a smart solar LED streetlight with an ESP32 camera.

2.1. *Block Diagram*

Fig. 1 shows that how all the components are connected in this system. LDR, PIR, ESP 32 camera module, push button and solar power will act as the inputs of the system. Meanwhile, LEDs, buzzer, notification, and live video are the output of the system.

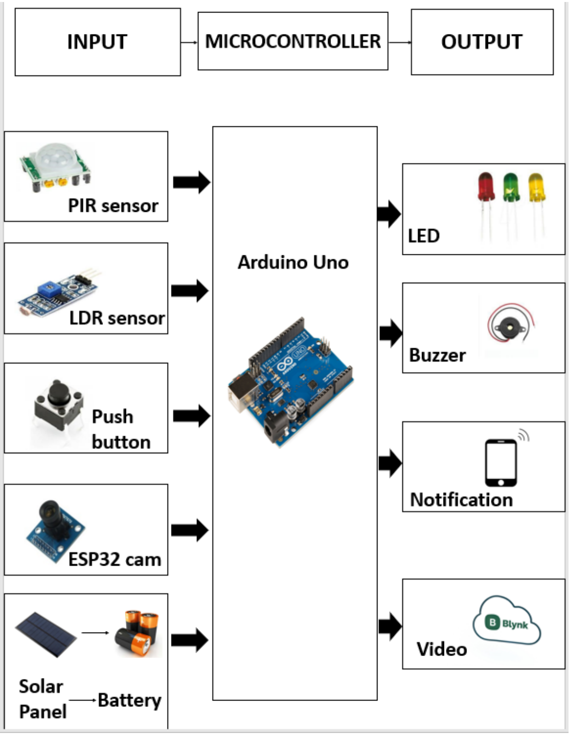


Fig.1 Block Diagram of the System.

2.2. *Flowchart*

Fig. 2 shows that, the system starts when solar power generates electricity during daytime, and it stored in battery. The power gained from the solar will be used to supply the lighting system. There will be two system which is the system to turn on the LED and the system when user press the push button. When, the LDR sensor detect whether the intensity of light of the surrounding is below than 400. If this condition is satisfied then the streetlight will turn on and if not, the light will be off. Then, when the PIR sensor detects any motion within the parameter of the garden it will increase the brightness of the LED to 100% but if there is no motion detected the brightness of the LED will remain 30%. While, for the push button, if the user press, it will turn on the buzzer and on the same time sending the notification, “Emergency Alert” to the person in-charged through Blynk apps. In addition to this, camera module will send live video through Blynk apps to the person in-charge.

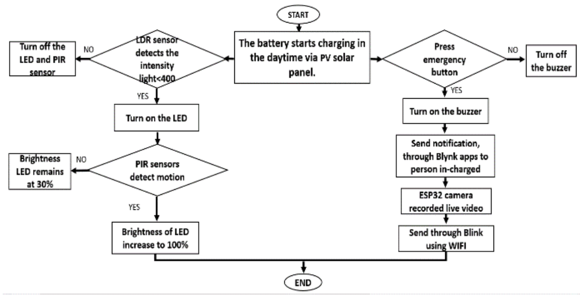


Fig.2 Flowchart of Smart Solar LED Street Light with ESP32 Camera Module.

2.3. *Schematic Diagram*

Fig. 3 shows that the schematic diagram of this project. This shows the details of the connection for Arduino Uno with other components.

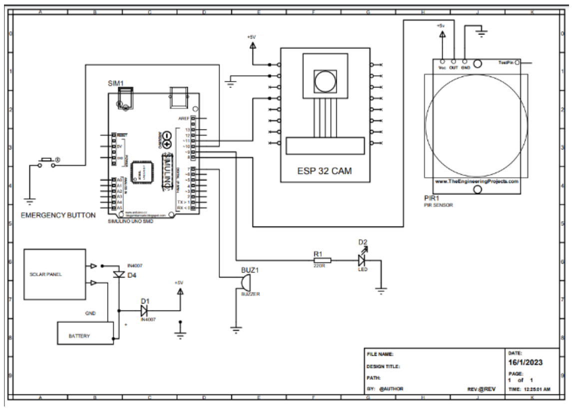


Fig.3 Schematic Diagram.

2.4. *Method Description*

Fig. 4 shows that, a photoresistor, also known as a light-dependent resistor, is a light-sensitive electrical component. This project utilized a digital kind of LDR, where the value of "1" indicates light and the value of "0" indicates darkness. The resistance alters as light strikes it. The resistance of the LDR can vary by order of magnitude, with the resistance decreasing as the light intensity rises. To save energy, we used LDR in this project to switch on and off the LED.

Fig.4 LDR Sensor.

Fig. 5 shows that, an application involving thermal sensing, a passive infrared (PIR) sensor is employed. Which utilized frequently in applications for autonomous lighting, motion detection alarms, and security alarms. PIR sensors are employed in this system to identify human presences, although they can only do so when a person is moving. By recognizing the human in a stationary posture, the Grid-EYE sensor solves the PIR sensor's drawback.

Fig. 5 PIR sensor

Fig. 6 shows that, when an electric current flow through a PN-junction LED in the forward direction, it creates light. White was chosen as the colour. This output correlates with the motion sensor and LDR sensor. When the environment is dark and the LDR determines the detection, the LED will turn on. While the LED will brighten more if motion is detected, this prevents energy from being squandered.

Fig.6 LED.

Fig. 7 shows that, OV7670 which is the most affordable camera module for the Arduino. Users will connect to Arduino, and it will send data when buzzer have been pressed and notification through Blynk. This is used for image processing so that Blynk apps can communicate the recorded video.

Fig.7 ESP32 camera module.

Fig. 8 show that, a push-button, that can be referred to as a button, is a straightforward switch mechanism used to run a machine or process. Buttons are composed of durable materials and are typically made of metal or plastic. It is usually flat or curved to accommodate a human finger or hand, making the surface easy to push against or compress.

Fig.8 Push button.

Fig. 9 show that, a sounding device that can transform audio signals into sound signals is a buzzer. The effect is alarm. To protect the user, the smoke sensor activates the buzzer when it senses smoke.

Fig.9 Buzzer.

These cells are made from a single crystal, as the name would imply. In comparison to polycrystalline panels, monocrystalline panels have higher efficiency. Efficacy is at 18%. Electricity is produced during the day by a high-efficiency monocrystalline solar panel shown in Fig. 10 and is stored in a battery.

Fig.10 Solar panel.

3. Results and Discussion

The result was inferred from the fact that the project is to avoid from the issue of streetlight, and this is for sensors connected to streetlight. The system works on the principle of detecting light by using sensor technology then sending data to the microcontroller so that the data can be processed to alight the streetlight. Beside this help in saving the energy by using solar power to generate power. This project is generally used by the authorities responsible for resolving the issue. Moreover, in this project show that the power consumes from smart light system with solar power is less compared to manual light system. According to this, it helps in minimize power wastage and this method also help to reduce electricity cost while implement an innovative way to generate electricity. The table below show the summarize result for this project (Table 1).

Table 1. Summarize from Result

Day	**LDR sensor**	**Condition of LED**	**PIR sensor**	**Condition of LED**
Morning	Light intensity more than 400	LED turn OFF	PIR sensor is turning off	-
Night	Light intensity less than 400	LED turn ON 30%	If there is motion detected	LED increase brightness to 100%
			If there is no motion detected	LED remain brightness of 30%

Fig. 11 shows that, solar panels provide electricity during the day, which is then stored in the rechargeable battery. The LDR sensor commands the Arduino controller from night to daylight. When LDR sensor detect light intensity more than 400, LED will turn off. When morning comes, LDR will instruct Arduino to turn off the streetlight. Typically, streetlights run on electricity that is stored in the battery. In Fig. 12 shows that, when LDR sensor detect light intensity less than 400, LED will turn on and if there is no motion beneath the streetlight, the programmed is carried out, and the LED is turned on to 30% of its maximum intensity. While in Fig. 13 shows that LED remain the brightness on 30% if there is no motion detected and Fig. 14 shows that, the brightness of the LED is increases as motion sensor turns on when a human or vehicle approaches a nearby streetlight, and then instructs the Arduino to increase the brightness to 100%.

When the timer expires and no movement is seen, the intensity progressively decreases to 30%. Furthermore, Fig. 15 shows that, when user press the push button it will turn on the buzzer and send a notification of “Emergency Alert” through Blynk apps to the person in-charge. This help to notified people at the surrounding of any danger that have been occurred. Besides, Fig. 16 shows that, ESP32 will record the video live and this can be view through Blynk apps that connected to the phone of the person in-charge.

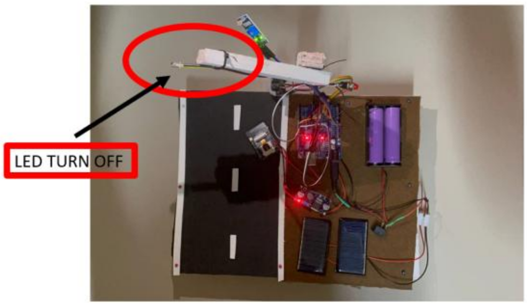


Fig.11 LED turn off.

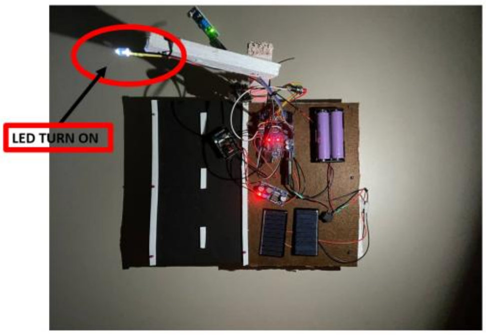


Fig.12 LED turn on.

Fig.13 Brightness of LED remain 30%.

Fig.14 Brightness of LED increases to 100%.

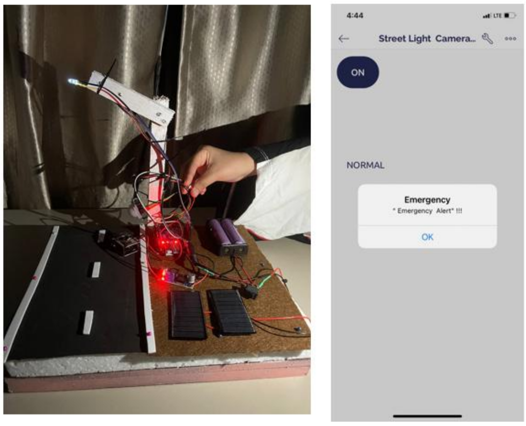


Fig.15 Notification through Blynk apps when press the push button.

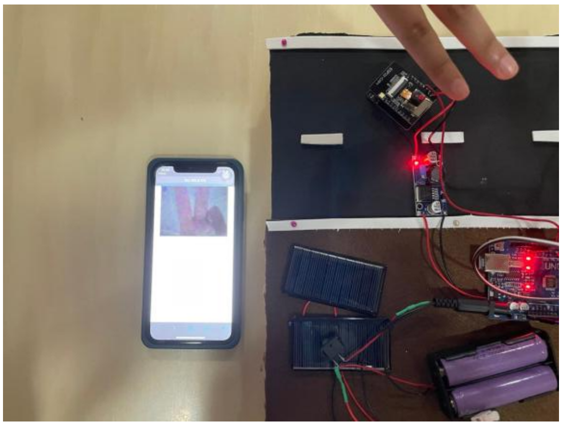


Fig.16 Blynk apps to view the recorded video.

4. Conclusion

Streetlights are crucial for the operation of smart cities. The internet of things introduces the idea of smart lighting with solar electricity. Numerous sensors can be used to automate life's operations. The resources are having trouble keeping up with the rising demand for electricity. People require new processes to reduce the cost of electricity and discover new methods of generating it. In this research project, the student proposes an energy-efficient technique for automating solar-powered street lighting. Streetlights can be powered by the electricity that solar power have generate. In the future, the system can be upgraded with light levels will change in response to motion detection. By finishing the project, students will learn more about the function and applications of the entire component, including how to find out more about the ESP32 camera module or other sensors.

This smart streetlight project not only helps in rural areas but also beneficial in urban areas too. The usage of renewable resources is helpful and useful as the world develop and need more electricity to do so. With the help of this technology, other features may also include such as intelligent parking for vehicles, which is beneficial for driverless vehicles. This technology has a promising future because it would not only save energy but also lessen natural disasters and even crime. Example, the security team can respond quickly with GPS positioning live feed and online voice call at the step when the user presses the push button, for instance, since this will direct link to the command centre for public emergency. With the help of this smart project, can also assess the speed of the car, identify the license plate, and identify accidents.

Acknowledgements

The authors would convey their profound appreciation and gratitude to the College of Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia for permitting us to perform this study. Special thanks to those who contributed to this project directly and indirectly.

References

1. S.N.A.Kazmi,"How street lighting upgrades can have a positive impact on our climate," 2020.
2. S. Singh, C. Shekhar, and A. Vohra, "Advanced footstep power generation system to charge e-vehicles,"2020, International Journal Of Engineering Research & Technology (IJERT).
3. Z. Zhang, F. Luan, and S. Cheng, "How smart street lighting can improve city's safety,"2017.
4. Segers, K. Steenhaut, and A. Touhafi, "Everything you need to know about smart street lighting," vol. 14, no. 2,pp. 1918–1949, 2014. [Online serial].
5. Segers, K. Steenhaut, and A. Touhafi, "inteliLIGHT® StreetLight Control softwar," vol. 14, no. 2,pp. 1918–1949, 2014. [Online serial].
6. UK Prime Minister Boris Johnson brings forward end date for new conventional petrol and diesel vehicles to 2030.
7. Y. L. Bo He, "Iot for energy efficient green highway lighting system," vol. 158, pp. 102575, 2020.
8. K.V.M.S.K.P.P.J. Dr.A. Senthil Kumar, "A study on IoT based smart street light systems," vol. 8, no. 7, 2020.
9. L. Li, K. Lian, J. Fu, P. Zhu, Z. Hu, and C. Guo, "Intelligent street light controller with security system,"vol.8,2020.

Authors Introduction

Nurul Farhana Zailani

She is currently serving as a Protege trainee at Telekom Malaysia. She holds a degree in Electronic Engineering and Diploma in Electric & Electronic Engineering from UiTM. Her interests encompass Communication.

Assoc. Prof. Ir. Ts. Dr Suzi Seroja Sarnin

She is working as a Senior Lecturer at School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She received PhD in Wireless Communication UiTM, Master of Science in Microelectronic and Degree in Electrical Engineering UTM. Her interests include Wireless Communication and IoT.

Mohd Nor Md Tan

He is working as a Senior Lecturer at School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. He received Master in Radio Frequency Communication Engineering, Bradford University, Bradford, UK. His research themes include Antenna Design Technology.

Ir. Ts. Dr Nani Fadzlina Naim

She is working as a Senior Lecturer at School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. She received PhD in Electrical, Electronic and System Engineering UKM, Bangi. Her interests include Optical sensor, Fiber Bragg Grating (FBG) based sensor, Optical Network.

Mohd Rizal Dohad

He is working as a lecturer at the School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. He received a Master of Science in Telecommunication and Information Engineering at UiTM and Bachelor of Engineering (Electrical) with Honors at UTM. Directly engaged in research projects based on the Internet of Things and Electronic Engineering.

Ts. Dr. Hjh Ros Shilawani S. Abdul Kadir

She is working as Senior Lecturer at School of Electrical Engineering, College of Engineering, UiTM Shah Alam, Selangor, Malaysia. Her research interests include Biomedical Engineering, with a particular focus on human body electromagnetic radiation and other human-related studies.

Nur Syahirah Saipul

She holds a Bachelor's degree in Electronic Engineering and a Diploma in Electrical & Electronic Engineering from UiTM. Currently, she is pursuing an MSc in Electrical Engineering at UiTM Shah Alam. Her areas of interest include communication systems and embedded systems.

Ts. Zarina Baharudin Zamani

She is working as a Senior Lecturer at Fakulti Kejuruteraan Elektronik dan Kejuruteraan komputer, Universiti Teknikal Melaka, Durian Tunggal, Malaysia. She received Master in Microelectronic, UKM. Her interests include VLSI, IC Design, PCB.

Efficient Weed Detection in Agricultural Landscapes using DeepLabV3+ and MobileNetV3

Renuka Devi Rajagopal, Manthena Rishit Varma

School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India

Heshalini Rajagopal

Department of Electrical & Electronic Engineering, Mila University, No. 1, Persiaran MIU, 71800 Putra Nilai, Negeri Sembilan Darul Khusus, Malaysia

Email: renukadevi.r@vit.ac.in, m.rishitvarma@gmail.com

Abstract

Weed detection is a crucial task in precision agriculture, significantly impacting crop yields and reducing the dependency on herbicides. Effective weed management enhances agricultural productivity by ensuring that crops receive adequate nutrients, water, and sunlight, which weeds would otherwise consume. Traditional weed control methods are labor-intensive and often rely heavily on chemical herbicides, which can have detrimental environmental effects. This paper presents a deep learning approach for weed detection, utilizing the DeepLabv3+ model with a MobileNetv3 backbone. This study underscores the potential of integrating advanced deep learning techniques into agricultural practices, paving the way for more sustainable and efficient weed management strategies.

Keywords: Deep Learning, Semantic Segmentation, Agriculture, Weed Detection, Image Processing, Computer Vision

1. Introduction

Weeds are a significant challenge in agriculture, competing with crops for essential resources such as nutrients, water, light, and space. This competition can drastically reduce crop yields, thereby impacting food production and the overall agricultural economy. Weeds can also serve as hosts for pests and diseases, further threatening crop health and productivity. The presence of weeds often necessitates increased labor and financial input for their control, making weed management a critical concern for farmers globally. Traditionally, weed management has relied heavily on chemical herbicides and manual labor. While effective, these methods come with several drawbacks. Chemical herbicides can have harmful environmental impacts, contaminating soil and water sources and affecting non-target plant and animal species. Moreover, over-reliance on herbicides can lead to the development of resistant weed species, complicating control efforts. Manual weeding, on the other hand, is labor-intensive and time-consuming, making it impractical for large-scale farming operations. To address these challenges, precision agriculture has emerged as a promising approach. Precision agriculture aims to optimize field-level management regarding crop farming. One of the key aspects of precision agriculture is the development of advanced technologies for weed detection and control. Accurate and timely identification of weeds is crucial for effective weed management, allowing for targeted interventions that minimize resource use and environmental impact. This paper presents an innovative solution for weed detection using deep learning, leveraging the capabilities of using DeepLabv3+ with a MobileNetv3 backbone. DeepLabv3+ is a state-of-the-art deep learning model designed especially for semantic segmentation, which involves classifying each pixel in an image into a predefined category. In this context, it is used to distinguish between crops and weeds within agricultural images. MobileNetv3, on the other hand, is a lightweight and efficient neural network architecture optimized for mobile and embedded vision applications. Its integration with DeepLabv3+ enhances the model's efficiency, making it suitable for real-time applications in the field where computational resources may be limited. The proposed system works by capturing images of the agricultural field using cameras mounted on drones or ground-based vehicles. These images are then processed by the DeepLabv3+ model, which uses MobileNetv3 as its backbone to perform real-time weed detection. The model analyses the images, segmenting the weeds from the crops with high accuracy. The integration of DeepLabv3+ and MobileNetv3 presents a powerful tool for precision agriculture, offering an efficient and accurate method for weed detection. This approach not only enhances the sustainability of weed management practices but also supports the broader goals of precision agriculture by optimizing resource use and minimizing environmental impact. According to the various papers referred and studied, there are many ways of implementation like i) a method for weed detection using low-level features like color and area to distinguish weeds from crops using Random Forest classifier (used in the CWFI dataset), ResNet-50, Inception-v3 models, ConvNets, SVM, AdaBoost, Random Forest models [1]. ii) an algorithm which detects weeds by segmenting green plants, using median and morphological filters, and area-based thresholding by using binary classification using image-processing techniques in MATLAB [2]. iii) a deep neural networks process crop images to detect weeds, which are then eliminated using a robotic arm with a weed cutter which uses YOLOv5 and CNN models [3]. iv) using Erosion and Dilation approach with Raspberry Pi where weeds are detected by converting images to binary, applying erosion and dilation, and counting white pixels exceeding a threshold [4]. v) a model which uses TensorFlow Lite where weeds are identified using TFL Classify app, which processes images from ESP32 AI Cam and directs herbicide spraying [5]. vi) a method utilizing YOLOv5, for accurately identifying and localizing weeds in agricultural fields [6]. vii) another method which works on YOLOv4, a state-of-the-art object detection algorithm which accurately identifies and classifies weeds in agricultural fields [7]. viii) with help of Yolov5, weed detection is done where the modified model replaces the 3x3 convolution in Yolov5's backbone with multi-head self-attention (MHSA) to improve weed detection accuracy [8]. ix) a method which proposes usage of YOLOv3 object detection algorithm for accurately identifying weed crops by applying CNN [9]. x) an approach where three deep learning models R-CNN, YOLOv3, and CenterNet are used for weed detection and measuring weed growth by calculating length and breadth [10]. The major challenges observed here are: (i) capturing image without any movements, (ii) good intensity level must exist, iii) there must be a standard height maintained between the crop top and the camera.

2. The Model's Architecture

DeepLabv3+ with MobileNetV3 is a powerful and efficient model for weed detection, combining the segmentation capabilities of DeepLabv3+ with the lightweight architecture of MobileNetV3. This combination enables accurate weed identification with reduced computational resources, making it suitable for real-time applications in agriculture. Its use of Atrous convolution and depth wise separable convolutions ensures high performance and fast inference times.

2.1. *MobileNet V3*

From Fig. 1, we can see the architecture of MobileNetV3, which can serve as a backbone network for models like DeepLabV3+ used for tasks such as weed detection. The architecture of MobileNetV3 consists of multiple layers. First comes the NL,Dvis layer where this layer represents a non-linear, depth wise convolutional layer. In this layer, the input image is processed using depth wise separable convolutions, which apply a single convolutional filter per input channel. This approach is computationally efficient and helps reduce the number of parameters. The "NL" likely stands for "non-linear," indicating that a non-linear activation function (e.g., ReLu) is applied after the depth wise convolution. Next comes the 3x3 block which represents a standard convolutional layer with 3x3 kernels. It takes the output feature maps from the previous layer and applies a set of 3x3 convolutional filters to extract higher-level features. This layer helps capture spatial and structural information in the input image. Next comes the Pool layer which down samples the spatial dimensions of the feature maps. It applies a sliding window operation, selecting the maximum or average value within each window. This reduces computational complexity and introduces translation invariance. The pooled features capture essential information while reducing spatial resolution. Pooling creates a spatial hierarchy, enabling detection of complex patterns. Then we have the FC, Relu and FC, Hard. These blocks represent fully connected (FC) layers with different activation functions. The "FC, Relu" branch uses the ReLU (Rectified Linear Unit) activation function, which introduces non-linearity by setting negative values to zero. The "FC, Hard" branch likely uses a different activation function, such as the hard swish or hard sigmoid, which can potentially improve performance for certain tasks. As shown in Fig. 1, the symbol X likely indicates a point where the two branches (FC, Relu and FC, Hard) are combined or merged. This could be done through an element-wise operation like addition or concatenation, allowing the network to leverage features from both branches. The "NL,|X|" component at the end likely represents a non-linear activation function applied to the merged features from the two branches (FC, Relu and FC, Hard).

In the context of weed detection using DeepLabV3+ with MobileNetV3 as the backbone, the input images would go as: the NL,Dvis layer extracts low-level features like edges, textures, and shapes from the input image using depth wise separable convolutions and non-linear activations. Then the 3x3 layer further processes these low-level features and captures spatial and structural information. The pooling layer down samples the feature maps, reducing their spatial dimensions while preserving essential information. The fully connected layers (FC, Relu and FC, Hard) capture higher-level semantic information crucial for tasks like object detection and segmentation. The merged features from the two branches are then fed into the main model which is chosen as per requirement. In our approach we use the main model as DeepLabV3+ to which the MobileNetV3 model acts as a backbone or foundation model to it.

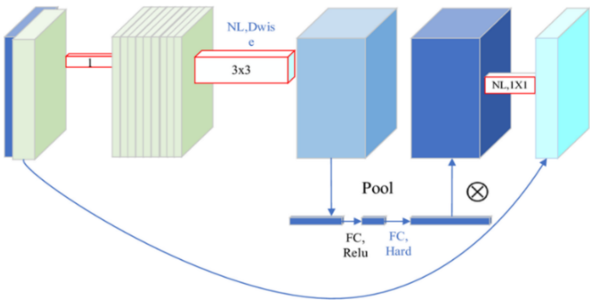


Fig. 1 Architecture of MobileNetV3

2.2. ***DeepLabv3+***

From Fig. 2, we can see the architecture of DeepLabV3+, with a focus on the ASPP (Atrous Spatial Pyramid Pooling) layer. In DeepLabV3+ also we have number of layers which perform unique tasks at each level. The architecture of this DeepLabV3+ comprises of 2 main layers namely Encode and Decoder.

Encoder Layer

Encoder layer is one which again consists of 4 major levels namely 1x1 Conv, 3x3 with different rates of 4,8,12, Global Pooling Layer, ASPP (Atrous Spatial Pyramid Pooling) layer. The Encoder layer also consists of the main backbone model used for DeepLabV3+ model. In the Fig. 2 we can see that a dedicated "Deep Convolution Neural Network" block is present which contains the backbone model and, in our case, it is the MobilNetV3 model. The outputs of the backbone model are sent to the next 4 layers of DeepLabV3+ model. This all happens in the Encoder Layer itself. Now comes the 1x1 Conv Layer which applies 1x1 convolutional filters to the input segmented image, projecting the feature maps to a lower-dimensional space, which helps reduce computational complexity. Next comes the 3x3 layer each with a rate of 4,8,12 individually. These are parallel 3x3 dilated convolutional layers with different dilation rates (4, 8, and 12). Dilated convolutions introduce sparse sampling patterns, allowing the model to capture multi-scale information and expand the effective receptive field without increasing the number of parameters. Now the Global Pooling layer applies global average pooling or global max pooling to the feature maps, capturing global context information from the entire input image. This global context can help the model differentiate between different regions based on their spatial distribution and relationships. The last layer in the Encoder layer of DeepLabV3+ is the ASPP (Atrous Spatial Pyramid Pooling) layer. This is a crucial component of DeepLabV3+ that effectively combines multi-scale information. The ASPP module applies parallel dilated convolutions with different dilation rates (e.g., 6, 12, 18) and also includes an image-level feature pooled globally. By capturing information at multiple scales, ASPP enables the model to accurately segment objects of various sizes and shapes. The features from the parallel dilated convolutions and the global pooling are concatenated and processed by a 1x1 convolutional layer, effectively combining the multi-scale information. This multi-scale representation is essential for weed detection, as weeds can have varying sizes, shapes, and spatial distributions within the input image.

Decoder Layer

The second layer in DeepLabV3+ is the Decoder layer which in turn has 5 important layers in it. Those are 1x1 Conv, Up sample by 4 layer, 3x3 Conv, Concatenate and again Up sample by 4 layers after which we get the final output image. Each layer has its own abilities and those are as follows: The 1x1 Conv layer combines and processes the multi-scale features from the Encoder outputs. Next the Up sample by 4 layer performs an operation which up samples the low-resolution feature maps from the Encoder, increasing their spatial resolution. This is necessary for producing a high-resolution segmentation map at the end of the process. After that the 3x3 Conv (Convolutional) layer processes the up sampled features which come from the previous layer refining the spatial information. Then the Concatenate layer performs concatenation of up sampled features from the Decoder with the corresponding low-level features from the Encoder. This helps the model leverage both high-level semantic information (from the Encoder) and low-level details (from the early layers of the Encoder) for accurate weed segmentation. At last, the Up sample by 4 again performs up sampling operation which further increases the spatial resolution of the feature maps, preparing them for the final segmentation map of the weeds.

The output segmentation map assigns a class label (weed or non-weed) to each pixel in the input image, enabling precise localization and identification of weeds. By using segmented images as input to the MobileNetV3 backbone, the model can leverage the pre-computed segmentation information to focus on extracting relevant features for weed detection. The multi-scale and global context information captured by the ASPP layer, combined with the low-level details from the early layers, enables the model to accurately segment and localize weeds, even when they appear at different scales and spatial distributions within the input image.

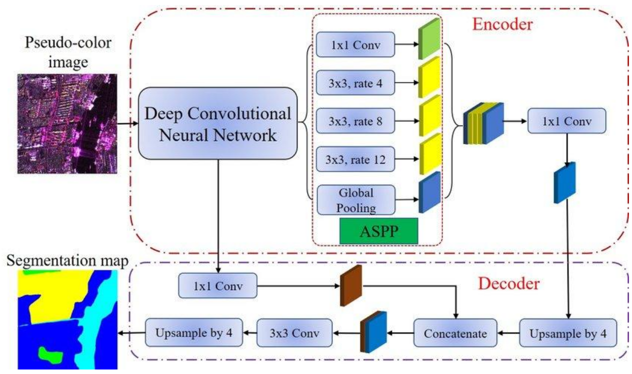


Fig. 2: Architecture of DeepLabV3+ model

3. Methodology

As shown in Fig. 3, initially we set up a camera which captures live aerial view of crops in the field with the help of OpenCV. Now the captured images are pre-processed and are sent to our trained model i.e. the DeepLabV3+ model with MobileNetV3 backbone. Inside the model, the input image is sent to the Encoder layer of the DeepLabV3+ model. In that Encoder layer, initially the images are sent to the MobileNetV3 model and the outputs from that model are then sent to other layers present in the Encoder layer. Now the remaining layers perform their operations and pass the results to the Decoder layer of DeepLabV3+ model. Sending into it, the sublayers in Decoder layer perform their operations and finally provides the high quality semantic segmented images of weeds present in the original images. Now convert the model's output to the binary mask and if needed for good visualization, apply color map to the binary segments. After that overly these color segmented maps with the original frame captured. Finally display the overlaid image at that instant on the screen. Now these segmented maps of weeds are further used for many applications in the field of precision agriculture.

Algorithm: Weed Detection using DeepLabV3+ with MobileNetV3
Input: Live Video from Camera
Output: Segmented Maps of Weeds

Step 1: Load necessary libraries (like OpenCV, TensorFlow)
Step 2: Load the custom trained DeepLabV3+ model with MobileNetV3 backbone.
Step 3: Continuously capture frames from the live video feed through OpenCV.
Step 4: Preprocess the captured frame (formatting, resizing etc.,)
Step 5: Pass the pre-processed frame to the loaded model
Step 6: Obtain the segmentation map as output.
Step 7: Convert the model's output to binary output And apply a color map to the mask to visualize the weed.
Step 8: Overlay the segmentation mask with the original frame.
Step 9: Display the frame with the overlaid weed detection mask.
Step 10: Release the camera and close all OpenCV windows.

DeepLabv3+ with MobileNetV3 is particularly well-suited for scalability in large-scale weed detection applications due to its lightweight architecture and efficient performance.
MobileNetV3 is designed to be lightweight, making it suitable for deployment on a wide range of devices, from high-end servers to edge devices like drones and field robots. This allows the model to be scaled across various platforms without significant modifications. The efficient computation and reduced resource requirements mean it can process large volumes of data quickly, which is essential for large-scale agricultural operations. Additionally, its compatibility with mobile and embedded systems ensures that real-time weed detection can be performed directly in the field, reducing the need for constant connectivity and enabling more autonomous operations.

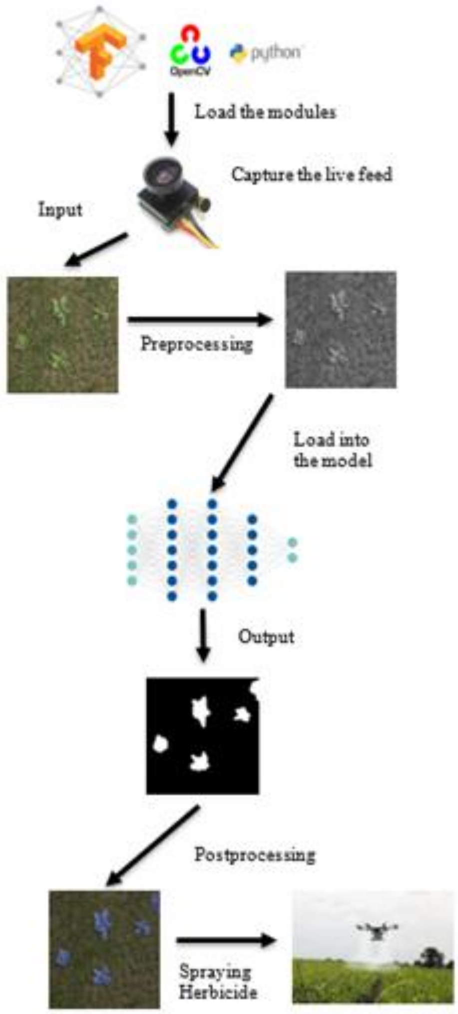


Fig 3: Pictorial representation of Algorithm

The segmented images of weeds from this model can be utilized in various applications across agriculture and related fields. In precision agriculture, these images can control robotic weeders for targeted weed removal and enable variable rate application of herbicides, thereby reducing chemical use and costs. In crop management, they can aid in yield prediction and field mapping to understand weed distribution and plan strategies accordingly. For research and development, the data can support weed ecology studies and crop breeding programs. Smart farming solutions can integrate these images into decision support systems and drone technologies for efficient weed monitoring. Additionally, they are valuable in environmental monitoring for biodiversity studies and managing invasive species. In education, they can create training materials and simulation models for agricultural training. The data also enhances machine learning models for better weed detection and segmentation accuracy and facilitates big data analysis to uncover patterns for future weed management practices. Commercially, these images can support agricultural consulting services and the development of customized weed management solutions,

leading to improved crop productivity and environmental sustainability.
Overall, the combination of DeepLabv3+ with MobileNetV3 offers a compelling mix of performance, efficiency, and scalability, making it superior to many other models for weed detection tasks. This makes it an ideal choice for extensive agricultural applications, ensuring high accuracy and real-time processing capabilities while being resource-efficient and versatile enough to be used across a broad spectrum of devices and platforms.

4. Experimental Results

The dataset used for this weed detection research consists of 1,400 high-resolution RGB images, captured from various agricultural fields to ensure a comprehensive representation of different weed species and environmental conditions. These images were sourced from both publicly available agricultural image repositories and direct field data collection using a drone-mounted camera. To facilitate the training, validation, and testing of the weed detection model, the dataset was divided into three subsets: 1,160 images were allocated for training, 196 images for validation, and the remaining 44 images for testing. This division ensures that the model is trained on a diverse range of images and can be rigorously evaluated to confirm its effectiveness in accurately detecting weeds in various scenarios. The outputs of test images are shown in the Fig. 4 and Fig. 5:

Fig. 4: Sample RGB images of test data

From Fig. 4, we can see the original RGB test samples which are randomly taken from the test dataset i.e. on which the model is applied, and Fig. 5 shows the predicted masks of the weed in binary format which are generated by the DeepLabV3+ model. Now these outputs can be further analyzed to get the coordinates or areas of the weed mask and send or store that information for further applications in precision agriculture.

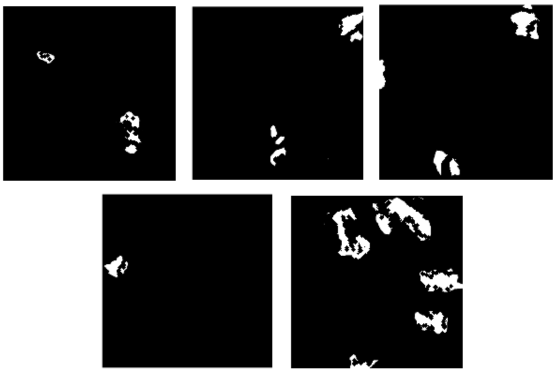


Fig. 5: Binary masks for the test images

Fig. 6 illustrates the relationship between the number of epochs and the corresponding accuracy achieved at each epoch during the training of the DeepLabV3+ model for weed detection. The training process of DeepLabV3+ involves optimizing the model's parameters to minimize the loss function, which measures the discrepancy between the predicted labels and the ground truth. As the number of epochs increases, the model iteratively adjusts its parameters to improve its performance on the training dataset. The overall accuracy achieved by the DeepLabV3+ model is around 84%, which is a significant performance metric for real-time weed detection applications. This high level of accuracy indicates that the model effectively distinguishes between weed and non-weed areas in the images. On other side Fig. 7 shows the graph between number of epochs vs the loss obtained at each point. The loss starts from 0.6 and gradually decreases to 0.06 which is considered to be minimal in real time. This overall metrics suggest that the model is performing well on weed images.

Weed detection using DeepLabv3+ with MobileNetV3 stands out when compared to other models due to its balance of accuracy, efficiency, and speed. While models like U-Net or traditional DeepLabv3 might offer high accuracy, they often require significant computational resources and processing time, making them less suitable for real-time applications in resource-constrained environments. In contrast, MobileNetV3 is specifically optimized for mobile and embedded devices, offering a lightweight yet powerful backbone that significantly reduces the model's computational requirements without sacrificing much accuracy. DeepLabv3+ enhances this further with its advanced segmentation capabilities, such as atrous spatial pyramid pooling and an improved decoder, which capture fine details and boundaries of weeds more effectively.
Compared to heavier models like ResNet or Inception used in segmentation tasks, DeepLabv3+ with MobileNetV3 is faster and more efficient, enabling deployment on edge devices in the field. This ensures quick processing and decision-making, which is critical

in precision agriculture. Additionally, the reduced computational overhead translates to lower energy consumption and cost, making it a more sustainable and practical choice for widespread agricultural use. Overall, the combination of DeepLabv3+ with MobileNetV3 offers a compelling mix of performance, efficiency, and scalability, making it superior to many other models for weed detection tasks.

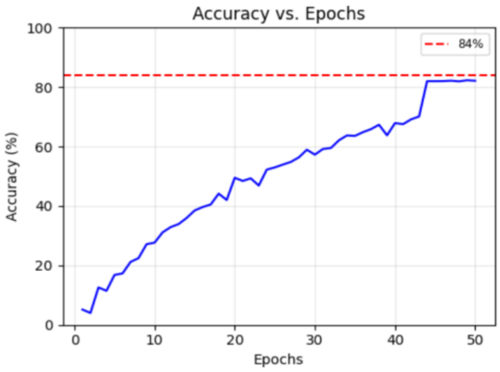


Fig. 6: Graphical analysis between Accuracy and Number of Epochs

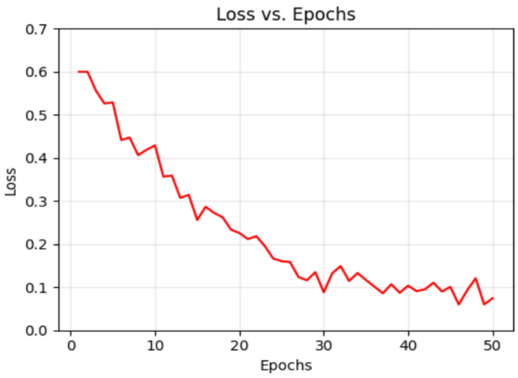


Fig. 7: Graphical analysis between Loss obtained and Number of Epochs

5. Conclusion

This research presents an innovative deep learning approach for weed detection using the DeepLabv3+ model with a MobileNetv3 backbone, offering a significant advancement in precision agriculture. The integration of these architectures provides a robust, efficient, and lightweight solution for real-time weed detection, addressing the challenges posed by traditional weed management methods. The DeepLabv3+ model's strength in semantic segmentation, combined with MobileNetv3's efficiency, enables accurate and fast weed identification, crucial for optimizing agricultural productivity and sustainability. Our methodology effectively leverages real-time image processing to distinguish between crops and weeds, facilitating targeted weed control interventions. This precision reduces the reliance on chemical herbicides, thereby mitigating their environmental impact and promoting sustainable farming practices. By harnessing advanced deep learning techniques, our approach enhances the efficiency of weed management strategies, contributing to better resource utilization and higher crop yields. The segmented weed images produced by our model can be utilized in various applications within the agricultural sector. These include guiding robotic weeders for precise weed removal, optimizing herbicide application rates, aiding in yield prediction, and supporting field mapping. Additionally, the data generated can be valuable for research in weed ecology, crop management, and machine learning model enhancement. Overall, the study underscores the potential of integrating advanced deep learning models into agricultural practices. The proposed DeepLabv3+ with MobileNetv3 backbone model not only provides an effective solution for weed detection with good accuracy but also paves the way for future innovations in smart farming. This approach exemplifies the move towards more sustainable and efficient agricultural practices, ensuring that modern farming can meet the growing demands for food production while preserving environmental integrity.

References

1. P. Khaire, V. Attar and S. Kalamkar, "A Comprehensive Survey of Weed Detection and Classification Datasets for Precision Agriculture," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-5, doi:10.1109/ICCCNT56998.2023.10306880.
2. A. J. Irías Tejeda and R. Castro Castro, "Algorithm of Weed Detection in Crops by Computational Vision," 2019 International Conference on Electronics,Communications and Computers (CONIELECOMP), Cholula, Mexico, 2019, pp. 124- 128, doi: 10.1109/CONIELECOMP.2019.8673182.
3. S. Tummapudi, S. S. Sadhu, S. N. Simhadri, S. N. T. Damarla and M. Bhukya, "Deep Learning Based Weed Detection and Elimination in Agriculture," 2023 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2023, pp. 147-151, doi: 10.1109/ICICT57646.2023.10134186.
4. R. Aravind, M. Daman and B. S. Kariyappa, "Design and development of automatic weed detection and smart herbicide sprayer robot," 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS), Trivandrum, India, 2015, pp. 257-261, doi: 10.1109/RAICS.2015.7488424.
5. T. B. M. I, J. J. Paul, B. Beulah and J. Joanna, "IoT Based Weed Detection and Removal in Precision Agriculture," 2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 2023, pp. 1-4, doi:10.1109/ICAECA56562.2023.10200751.
6. J. Prabavadhi and K. S, "Mobile Based Deep Learning Application for Weed and Medicinal Plant Detection Using YOLOV5," 2023 International Conference on System, Computation, Automation and Networking

(ICSCAN), Puducherry, India, 2023, pp. 1-5, doi:10.1109/ICSCAN58655.2023.10395553.
7. N. Khalid, H. Elkhiri, E. Oumaima, N. ElFahsi, Z. F. Zahra and K. Abdellatif, "Revolutionizing Weed Detection in Agriculture through the Integration of IoT, Big Data, and Deep Learning with Robotic Technology," 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Tenerife, Canary Islands, Spain, 2023, pp. 1-6, doi: 10.1109/ICECCME57830.2023.10252598.
8. Z. Yu et al., "Weed detection based on improved Yolov5," 2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 2023, pp. 1936-1940, doi: 10.1109/ICSP58490.2023.10248865.
9. Y. Dandekar, K. Shinde, J. Gangan, S. Firdausi and S.Bharne, "Weed Plant Detection from Agricultural Field Images using YOLOv3 Algorithm," 2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA, Pune, India, 2022, pp. 1-4, doi: 10.1109/ICCUBEA54992.2022.10011010.
10. S. N, M. Sundaram, R. Ranjan and Abhishek, "Weedspedia: Deep Learning-Based Approach for Weed Detection using R-CNN, YoloV3 and Centernet," 2023 International Conference on Quantum Technologies, Communications, Computing, Hardware and Embedded Systems Security (iQ-CCHESS), Kottayam, India, 2023, pp. 1-5, doi: 10.1109/iQ-CCHESS56596.2023.10391389.

Authors Introduction

Dr. Renuka Devi Rajagopal

Dr. R. Renuka Devi is an Associate Professor in the School of Computer Science and Engineering, VIT Chennai, India. Her research interests include Cyber-Physical Systems, Block Chain Technology, Data Mining and Machine learning in the field of Health Care.

Mr. Manthena Rishit Varma

He is currently pursuing Bachelor's Degree in Technology in Computer Science and Engineering with specialisation in Artificial Intelligence and Machine Learning at Vellore Institute of Technology in India.

Dr. Heshalini Rajagopal

She received her PhD and Master's degree from the Department of Electrical Engineering, University of Malaya, Malaysia in 2021 and 2016, respectively. She is an Assistant Professor in the School of Engineering and Computing, Mila University. Her research interests include image processing, artificial intelligence and machine learning.

AI-Based Weed Detection Algorithm using YOLOv8

Renuka Devi Rajagopal, Rethvik Menon C, T S PradeepKumar
School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India

Heshalini Rajagopal
Department of Electrical & Electronic Engineering, Mila University, No. 1, Persiaran MIU, 71800 Putra Nilai, Negeri Sembilan Darul Khusus, Malaysia

Email: renukadevi.r@vit.ac.in, rethvikmenon.c2021@vitstudent.ac.in, heshalini.rajagopal@mila.edu.my

Abstract

The development of a country relies heavily on agricultural produce and its related sectors. However, farmers face significant challenges due to the uncontrolled growth of weeds, which reduces their yield. Weed detection is a key step in the removal process, and advances in technology, such as the YOLOv8 model, have simplified this task. YOLOv8 offers improved weed and crop detection, with enhancements of 1.3% and 1.17% in mAP50 and mAP50-95, respectively, over the previous YOLOv5 model. This allows farmers to efficiently identify and eliminate weeds, leading to higher productivity and better crop yields, ultimately supporting the agricultural growth of the country.

Keywords: Agriculture, Artificial Intelligence, Weed Detection, YOLOv8

1. Introduction

Weeds are the primary reason for the reduced production of agricultural produce on farms. It is estimated that over 43% of crop losses in farms are caused by weeds which is a huge amount of loss. It is therefore imperative to get rid of the weeds in an agricultural farm. The most significant and basic step in this process of weed elimination from an agricultural farm involves the process of weed detection. The classical and traditional process of weed removal is cumbersome and can sometimes result in the elimination of useful crops. Hence it is very crucial to ensure a reliable and efficient method to detect and eliminate these crops. The traditional image processing techniques involve the analysis of the morphological features of the weed to ensure the detection of the weeds [1]. This process involves a large amount of time. It can be forced to undergo bias which makes it ineffective to implement thereby creating a need to search for a more efficient method to ensure a sustainable process for weed detection. A combination of deep learning algorithms have been used by authors of [2] have utilized the ResNeXt feature extraction along with a Faster R-CNN model to ensure efficient weed detection. The introduction of YOLOv8 by Ultralytics offers greater performance than its predecessors like YOLOv5 and its associated models [3].

2. Literature Review

Faisal Ahmed et al. studied the separation of weeds from crops in digital photos with the help of support vector machines (SVM) and Bayesian classifiers. It was observed that the SVM classifier performs better than the Bayesian classifier. A weed-detecting device powered by solar panels was utilized. The machine consisted of a camera installed at the base for continuous image capture, and it was mobile to help in the movement between crop rows [4]. The paper [5] by Liu et al. (2021) displays a deep learning-based system for weed detection in crops based on Unmanned Aerial Vehicle imagery. The authors employed a modified version of the Faster R-CNN model to detect and classify weeds in the images. The system had successfully achieved high accuracy in weed detection. The authors of the paper [6] employed techniques of CNN and Res-Net 50 algorithms on a dataset to efficiently classify weeds and crops. The results show that the Res-Net 50 algorithm outperforms the CNN model with an output of 84.6% and 90% accuracy on two different datasets of cucumber and onion respectively. The U-Net-MobileNetV2 architecture is utilized in [7] to achieve accurate and efficient detection of weeds while reducing computational costs. The methodology proposed in this study has illustrated a 96% accuracy. The experiment conducted in paper [8] uses the ADAM optimizer on a four-class weed dataset showing a 96.58% accuracy in classifying and categorizing weeds in a field. In the paper [9] the authors have employed the technique of using CNN which will thereby help to demonstrate the effectiveness of deep learning models to help in efficient weed detection and classification. This research in [10] represents an integrated method for transforming farming by combining robots, the Internet of Things (IoT), and environmental protection. The Machine learning

Convolutional Neural Networks (CNNs) algorithm has been combined for focused and efficient treatment.

3. Methodology

The process of employing the YOLOv8 model to efficiently classify weeds and crops includes various steps. The primary step involves isolating and collecting a dataset that can be suitably engineered to help in the ease of training. The dataset was received from an open-source tool online [11]. The dataset was further annotated to help in the efficient classification of the weed and crop to improve the training metrics while employing the YOLOv8 model. Before training the model, certain elementary steps were involved which included importing the Ultralytics and YOLO libraries. The training included using the ADAM optimizer and training it for an epoch size of 100 with a worker size of 4 and a comprehensive batch size of 16 and included a freezing time of 10 ms to ensure an efficient training of the dataset under the YOLOv8 model. The trained model was allowed to be evaluated by passing it under a validation and testing dataset to verifyand isolate the results produced by themodel. The evaluation metrics of Precision, Recall, and IoU have been generated to produce an analysis of the model which has been trained on the given weed dataset as shown in Eq. [1], [2] and [3], respectively.

$$Precision = TP/(TP + FP) \quad [1]$$

$$Recall = TP/(TP + FN) \quad [2]$$

IoU =(Area of Intersection)/(Area of Union) [3]

After training and testing this model suitable validation and testing dataset the results and the findings were reported to justify the utility of this model to ensure a clear classification of weed and crop dataset. Fig. 1 shows the training of the model under YOLOv8 for 100 epochs while Fig. 2 shows the visual representation of the entire run history of the training model.

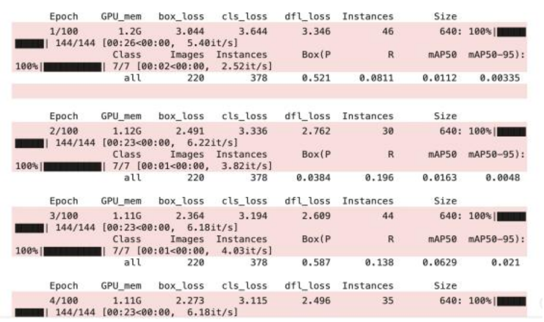


Fig. 1. Training of the model under YOLOv8 for 100 epochs

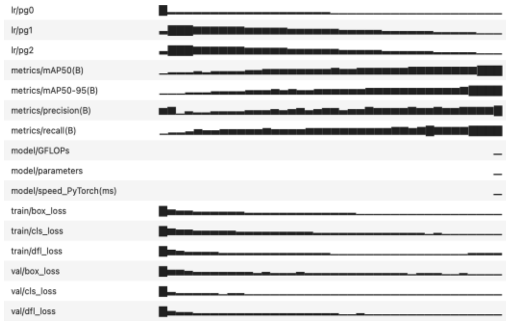


Fig. 2. Visual representation of the entire run history of the training model

The flowchart in Fig. 3 gives a lucid description of the processes involved in the training and running of this model to efficiently classify the weedsand crops in the given dataset.

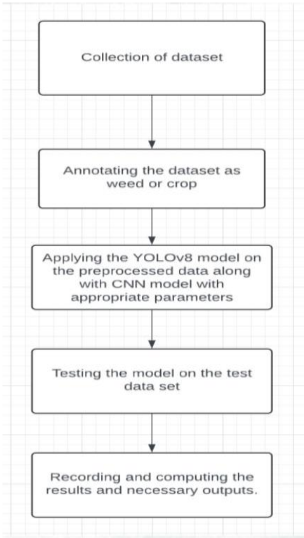


Fig. 3. Flow of processes of the training and testing of the YOLOv8 model.

4. Results

After employing the YOLOv8 model on the given dataset. The classification has occurred successfully indicating a successful construction and testing of the model. Fig. 4 shows a clear and distinct classification between weeds and crops where 0- crops and 1- indicates the crops. The image in Fig. 5 signifies the precision-confidence curve at different thresholds. Fig. 6 establishes the graphical representation of the recall-confidence curve. The table shown in Fig. 7 establishes the cumulative values of all the metrics taken for this classification.

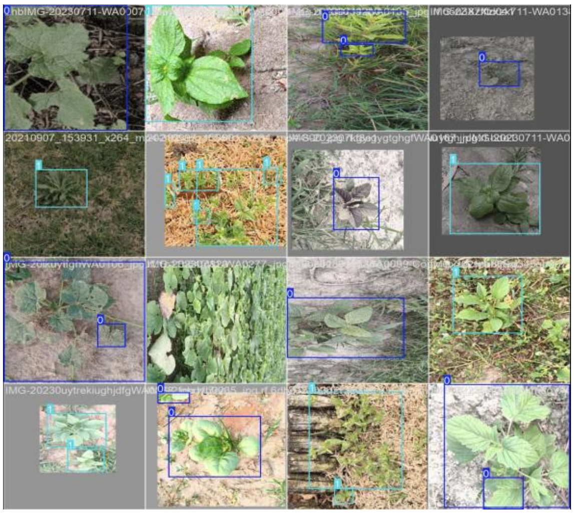


Fig. 4. Classification between weeds and crops where 0- crops and 1- indicates the crops.

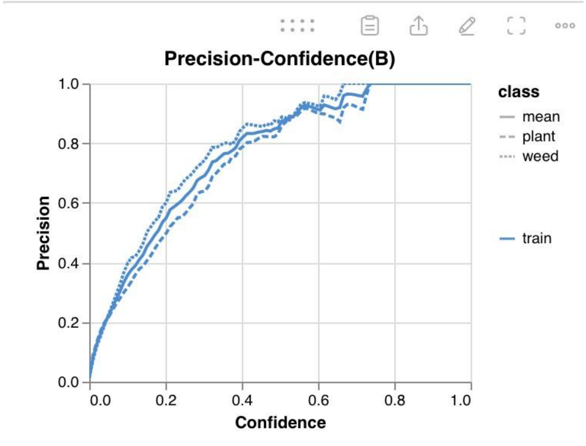


Fig. 5. Precision-confidence curve at different thresholds.

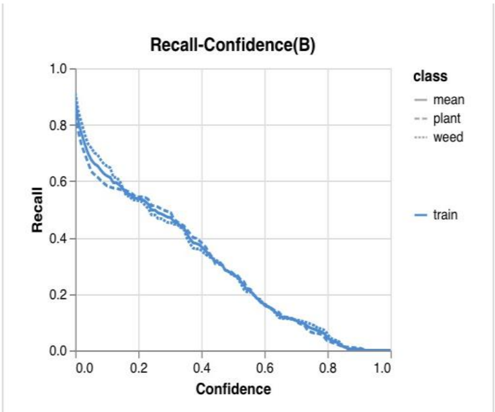


Fig. 6. Recall-confidence curve.

S.no	Metric	Value
1	Precision	0.9895
2	F1-score	0.9483
3	mAP-50	0.9564

Fig. 7. Cumulative values of all the metrics taken for this classification.

5. Conclusion:

The objective of conducting this experiment using the YOLOv8 on the given weed dataset was to ensure an efficient and computationally less intensive method to classify and precise classification rate of 0.9895 which indicates a highly accurate and successful classification. The graphical representation of precision and confidence indicates that precision values almost attain a near-perfect maximum while plotting the confidence at different thresholds which shows a significant growth at every threshold. Therefore, we can conclude that the above-mentioned YOLOv8 model is a very clear and imperative model for classifying weeds and crops thereby acting as a suitable means to help in the primary step of weed classification and detection.

References

1. Rai, N.; Zhang, Y.; Ram, B.G.; Schumacher, L.; Yellavajjala, R.K.; Bajwa, S.; Sun, X. Applications of deep learning in precision weed management: A review. Comput. Electron. Agric. 2023, 206, 107698.
2. Mu, Y.; Feng, R.; Ni, R.; Li, J.; Luo, T.; Liu, T.; Li, X.; Gong, H.; Guo, Y.; Sun, Y.; et al. A faster R-CNN-based model for the identification of weed seedling. Agronomy 2022, 12,2867.
3. Hussain, M. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. Machines 2023, 1, 677.
4. FaisalAhmed, A.S.M.HussainBari, EmamHussain, Performance analysis of Support Vector Machine and Bayesian Classifier for Crop and Weed classification from digital Images", World Applied Sciences Journal, 2019 .
5. Etienne, A., Ahmad, A., Aggarwal, V., & Saraswat, D. (2021). Deep learning-based object detection system for identifying weeds using uas imagery. Remote Sensing, 13(24), 5182.
6. S. Badhan, K. Desai, M. Dsilva,R. Sonkusare and S. Weakey, "Real-Time Weed Detection using Machine Learning and Stereo-Vision," 2021 6th International Conference for Convergence in Technology (I2CT),

Maharashtra, India, 2021, pp. 1-5, doi: 10.1 09/I2CT51068.2021.9417 989.
7. M. F. Qureshi, F. Amin, Z. Mushtaq, M. Ali, A. A. Haris and A. Y. Rana, "Real-Time Weed Segmentation in Tobacco Crops Utilizing Deep Learning on a Jetson Nano," 2024 International Conference on Engineering & Computing Technologies (ICECT), Islamabad, Pakistan, 2024, pp. 1-6, doi: 10.1 09/ICECT61618.2024.10581350.
8. S. Samala et al., "Optimizing Crop Management: Customized CNN for Autonomous Weed Identification in Farming," 2024 International Conference on Integrated Circuits Communication Systems (ICICACS), Raichur, India, 2024, pp. 1-5, doi: 10.1 09/ICICACS60521.2024.1 0499041.
9. A. Giradkar, R. Adpawar, R. Agrawal, C. Dhule and N. C. Morris, "Design of Autonomous Weed Elimination using Maching Learning Techniques," 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2023, pp. 102-106, doi:10.1 09/ICSCSS57650.2023.10169371.
10. R. Raman, P. R. Parvathy, P. Sapra, V. B. Sonule and S. Murugan, "Robotic Weed Control and Biodiversity Preservation: IoT Solutions for Sustainable Farming," 2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2023, pp. 1 2-1 7, doi:10.1 09/ICECA58529.2023.10395531.
11. Chakrabarty, R., Sardar, T., Koley, S., Das, S., Sengupta, I., Ali, A., & Chaudhuri, S. S. (2023, December). Weed Segmentation: A Novel Dataset, Benchmark and YOLO Baseline. In 2023 7th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech) (pp. 1-6). IEEE.

Authors Introduction

Dr. Renuka Devi Rajagopal

She is an Associate Professor in the School of Computer Science and Engineering, VIT Chennai, India. Her research interests include Cyber-Physical Systems, Block Chain Technology, Data Mining and Machine learning in the field of Health Care.

Mr. Rethvik Menon

He is currently pursuing B.Tech in the field of Computer Science from the Vellore Institute of Technology, Chennai.

Dr. T S Pradeep Kumar

He received his PhD from VIT University, Vellore, India. He is currently a Professor with School of Computer Science and Engineering, VIT Chennai, India. His research interests include Power modeling of sensors, energy efficiency of wireless networks, Vehicular Adhoc Networks (VANETs), E-learning strategies of higher education in Open source computing.

Dr. Heshalini Rajagopal

She received her PhD and Master's degree from the Department of Electrical Engineering, University of Malaya, Malaysia in 2021 and 2016, respectively. She is an Assistant Professor in the School of Engineering and Computing, Mila University. Her research interests include image processing, artificial intelligence and machine learning.

Novel Gender And Age- Based Detection Technique for Facial Recognition System

Pratham Gupta
School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India

Amutha S
School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India

Dhanush R
School of Electronics and Engineering, Vellore Institute of Technology Chennai, India

Heshalini Rajagopal
Department of Electrical & Electronic Engineering, Mila University, No. 1, Persiaran MIU, 71800 Putra Nilai, Negeri Sembilan Darul Khusus, Malaysia

Email: mail.guptapratham@gmail.com, dhanush.r@vit.ac.in, amutha.s@vit.ac.in, heshalini.rajagopal@mila.edu.my

Abstract

This paper introduces gender and age-based classification approaches in facial recognition systems, addressing challenges posed by diverse demographic characteristics. The model learns typical facial features and identifies deviations from them by employing unsupervised detection methods using autoencoders, enhancing robustness and generalization across populations. Ethical considerations are discussed, emphasizing the importance of fairness and bias mitigation in facial recognition. Experimental results demonstrate the method's effectiveness in handling biases in traditional supervised approaches. This research contributes a novel technique while highlighting the ethical implications of facial recognition. Gender and age-based detection methods improve system reliability in real-world scenarios with diverse demographics. These findings are relevant to researchers, developers, and policymakers navigating the intersection between facial recognition and ethical AI, promoting more responsible and inclusive technology.

Keywords: Facial recognition, Supervised learning, Age and gender classification, Neural network, Demographic characteristics, Generalization, Facial images

1. Introduction

Gender and age-based detection methods for facial images have garnered significant attention in computer vision research due to their wide spectrum of applications ranging between security and healthcare. In the real world, the precise identification of facial images based on gender and age presents unique challenges, including variations in lighting, facial expression, and pose. Such challenges were addressed over the years using diverse methodologies, with a notable shift towards deep learning concepts such as convolutional neural networks (CNNs). This introduction sets the stage for understanding the evolution of gender and age detection algorithms, highlighting key studies and advancements in the field.

Recent advancements in deep learning have revolutionized the field of demographic classification, particularly in age and gender estimation from facial images. Levi and Hassner's groundbreaking work in 2015 [1] marked a pivotal moment by demonstrating the efficacy of deep Convolutional Neural Networks (CNNs) in achieving high accuracy in demographic predictions. Their innovative use of CNNs showcased the potential of hierarchical feature extraction for capturing subtle patterns in facial data, setting a foundational benchmark that inspired subsequent research in the field. Building on Levi and Hassner's foundational work [1], Xing et al. [2] expanded the scope of demographic classification by integrating race classification into their model through multi-task learning. This approach allowed the model to handle multiple demographic attributes simultaneously, improving its robustness and versatility. Multi-task learning leverages shared representations, which enhances the model's capability to generalize across different tasks. Eidinger, Enbar, and Hassner [3] tackled the challenges in age and gender classification by using dropout-SVMs and robust face alignment techniques. Dropout-SVMs helped in improving model generalization by preventing overfitting, while robust face alignment techniques ensured accurate localization of facial features under various conditions. Kelliher's [4] comparative study highlighted the superiority of CNNs over traditional SVMs for complex visual tasks such as age and gender classification, systematically demonstrating how deep learning models outperform traditional machine learning approaches in handling the intricacies of visual data. Lee et al. [5] focused on optimizing feature extraction techniques in gender classification, introducing novel approaches such as improved pre-processing steps, advanced network architectures, and refined training methods, which significantly enhanced prediction accuracy. Liu et al. [6] proposed a CNN-based model specifically designed for automatic age classification, showcasing robust performance across various age groups and demographic variations, highlighting the adaptability and scalability of deep learning techniques. Sundararajan and Woodard's [7] research enhances biometric systems by integrating deep learning, specifically improving age and gender classification accuracy using convolutional neural networks (CNNs). Their approach tackles challenges like facial variability, crucial for reliable authentication and identification in practical applications. CNNs excel in learning facial features despite lighting, expressions, and aging, providing robustness. This advancement promises more secure access control, reliable identity verification in finance, and effective law enforcement applications. Overall, their study signifies a significant stride in leveraging CNNs to bolster biometric system reliability and security, addressing critical limitations and enhancing real-world deployment efficacy. Hamdi and Moussaoui's [8] comparative analysis rigorously examined the efficacy of deep learning models versus traditional methods in age, gender, and ethnicity identification. Their study meticulously evaluated various approaches, highlighting deep learning's superiority in accuracy and robustness. By contrasting these models, they elucidated nuanced strengths and limitations, offering valuable insights for biometric system developers and researchers. The findings underscore deep learning's capability to handle complex data patterns and variability inherent in demographic attributes like age, gender, and ethnicity. Sharma et al. [9] showed better performance in age and gender classification through a pioneering approach that integrated advanced training techniques and utilized extensive datasets. By harnessing these strategies, Sharma et al. significantly enhanced classification accuracy and robustness, surpassing previous benchmarks. Their achievement underscores the importance of methodological innovation and dataset scale in advancing biometric systems, offering insights into effective strategies for improving age and gender classification accuracy in practical applications. Wang et al. [10] developed a sophisticated age estimation model that focused on capturing intricate ageing patterns, particularly in scenarios with limited visual cues. Their model aimed to enhance precision in demographic analysis by addressing challenges such as subtle facial changes over time. By leveraging advanced algorithms and robust data analysis techniques, Wang et al. achieved significant improvements in age estimation accuracy, contributing to more reliable and nuanced demographic assessments in various practical contexts. Meinedo and Trancoso [11] emphasized age and gender classification's pivotal role in forensic applications, particularly in contexts requiring precise demographic analysis like identifying child abuse material. Their focus underscores the importance of accurate biometric techniques in sensitive legal and investigative settings. By advancing methods for age and gender classification, Meinedo and Trancoso contribute to enhancing forensic tools' efficacy, aiding law enforcement in addressing critical issues related to child protection and criminal investigations. Saxena, Singh, and Singh [12] stressed the importance of robust demographic detection for enhancing security and social services, underscoring how accurate demographic predictions play a crucial role in improving safety and service delivery across various domains. Shaker and Al-Khalidi [13] demonstrated robust gender and age detection suitable for surveillance and access control systems, emphasizing the practical utility of accurate demographic classification in enhancing security and operational efficiency. Nada et al. [14] developed a solution for reliable demographic verification from profile photos using advanced image processing techniques, crucial for maintaining accuracy in online platforms. Ghosh and Bandyopadhyay [15] proposed an integrated method for gender classification and age detection, emphasizing feature extraction, selection, and classification techniques. Their approach aimed to enhance accuracy and efficiency across varied datasets, offering a robust framework for biometric applications requiring reliable demographic analysis. Brandao [16] focused on mitigating age bias in pedestrian detection algorithms to enhance fairness and inclusivity in demographic prediction models. Their work aimed to reduce disparities in algorithmic accuracy across different age groups, contributing to more equitable outcomes in automated systems relying on demographic analysis. Kumbhar and Shingare [17] emphasized the importance of comparative studies in deep learning algorithms for age and gender classification, providing a detailed evaluation of various models to guide future improvements. Wang, Ali, and Angelov [18] introduced an approach for gender and age classification crucial for detecting anomalous human behavior, contributing to the development of systems capable of identifying and responding to unusual activities. Karahan et al. [19] engineered machine learning algorithms specifically to elevate real-time accuracy in age and gender classification tasks. Their research underscores the significance of achieving robust performance under real-world conditions, enhancing the reliability and applicability of biometric systems in practical settings such as security and surveillance. Zaman et al. [20] expanded into multimodal biometric systems, enhancing voice-based demographic predictions by leveraging deep learning's versatility to improve accuracy across multiple modalities, highlighting the benefits of integrating multimodal data for comprehensive demographic predictions.

This paper introduces a gender- and age-based classification method in facial recognition systems. Traditional approaches often rely on supervised learning, which may encounter challenges in handling diverse demographic characteristics. Our proposed technique employs unsupervised detection to identify deviations from the norm, providing a more robust solution. We utilize autoencoders, a type of neural network well-suited to learn the inherent features of facial images. This research not only contributes a novel detection technique to the field but also emphasizes the importance of ethical AI in facial recognition. The application the novel detection techniques using gender and age classification enhances the reliability of facial recognition systems, making them more suitable for real-world scenarios where diverse demographic characteristics are prevalent.

2. Methodology

2.1. Dataset Description

Zhang .et. al [21] [22] aggregated large-scale facial image datasets with over 20,000 face images with annotations of age, gender, and ethnicity.

Here's a breakdown of the key components:

i. Images: The dataset contains over 20,000 RGB images of faces. These images vary in resolution and quality but generally provide a diverse set of facial characteristics.
ii. Age Annotation: Each face image is annotated with the subject's age. This annotation spans from 0 to 116 years old, making it one of the few datasets to cover such a broad age range.
iii. Gender Annotation: Alongside age, each image is also annotated with the subject's gender, indicating whether the face belongs to a male or female.
iv. Ethnicity Annotation: The dataset includes annotations for the ethnicity of each subject. This provides additional information for researchers interested in studying facial recognition across different ethnic groups.
v. Variability: The dataset captures various factors influencing facial appearance, including pose, expression, illumination, ethnicity, and age. This variability makes it suitable for training and evaluating algorithms robust to real-world conditions.
vi. Research Applications: UTKFace dataset is commonly used for research in age estimation, gender classification, and ethnicity recognition. Additionally, it can be utilized for tasks like facial recognition, facial expression analysis, and age progression/regression.
vii. The UTKFace dataset has been widely used in the computer vision and machine learning communities due to its large size, diversity, and comprehensive annotations, making it a valuable resource for researchers working on facial analysis tasks.

2.2. Data Preprocessing

Data augmentation techniques involve artificially increasing the diversity of the training dataset by applying transformations such as rotation, flipping, or scaling to the input images.

2.3. CNN for Multi Output Classification Model (CNN_MOCM)

The CNN_MOCM that we have made, constructs a convolutional neural network (CNN) with multiple convolutional and fully connected layers for gender and age prediction, a proposed method could involve the following enhancements:

· Multi-Task Learning (MTL): MTL involves jointly training the model for multiple related tasks, in this case, gender and age prediction. By sharing information between tasks during training, MTL can lead to improved generalization and performance compared to training separate models for each task.

- Attention Mechanisms: Attention mechanisms enable the model to dynamically focus on relevant regions or features of the input data. In the context of facial recognition, attention mechanisms can help the model prioritize important facial features for gender and age prediction, such as eyes, nose, or mouth regions.
- Transfer Learning: Transfer learning involves leveraging knowledge learned from pre-training on a large dataset (e.g., ImageNet) and fine-tuning the model on the specific gender and age prediction task. By initializing the model with pre-trained weights, transfer learning can accelerate the training process and boost performance, especially when the target task has limited training data.
- Regularization: The issue of overfitting could be avoided by adding a penalty term to the loss function. Thereby bypassing the use of large weights in the model. By penalizing complex models, regularization encourages simpler and more generalizable representations, leading to improved performance on unseen data.
- Hyperparameter Tuning: Hyperparameter tuning involves systematically searching for the optimal values of hyperparameters (e.g., learning rate, dropout rate) that govern the training process. By fine-tuning these hyperparameters, we can optimize the model's performance and convergence speed, leading to better results on the gender and age prediction task.

In summary, incorporating these enhancements into the CNN architecture can significantly improve the accuracy, robustness, and interpretability of gender and age prediction in facial recognition systems, ultimately enhancing their practical utility in real-world applications. The flowchart of model used, and the process is shown in Fig. 1.

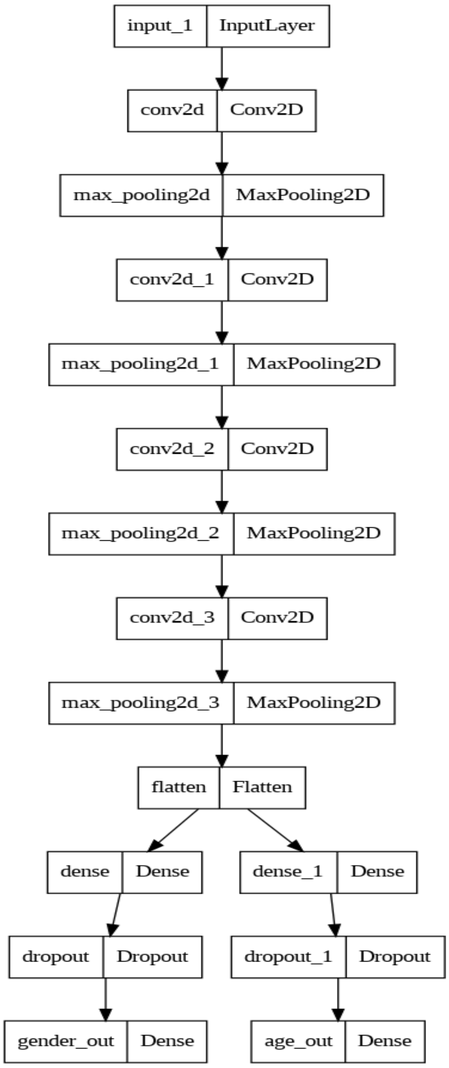


Fig. 1. The flowchart of CNN_MOCM

2.4. Materials

Large datasets of labelled facial images are used to train machine learning models for gender and age detection. CNNs algorithm analyse the visual characteristics of the face, such as facial landmarks, textures, and shapes, to capture important information for gender and age classification. Various software libraries and frameworks provide tools and functionalities for building gender and age detection systems, such as OpenCV, TensorFlow, PyTorch, and sci-kit-learn. These libraries offer pre-trained models, as well as APIs for training custom models and deploying them in applications. Depending on the complexity of the machine learning models used for gender and age detection, training and inference may require significant computational resources. High-performance GPUs (Graphics Processing Units) are often utilized to accelerate training and inference tasks, enabling faster processing of large volumes of facial images. Annotated datasets with labels for gender (male/female) and age groups are essential materials for training gender and age detection models. These annotations provide ground truth labels that the models learn to predict during the training process.

By leveraging these materials, researchers and developers can build accurate and robust gender and age detection systems that can be applied in various real-world applications, such as demographic analysis, targeted advertising, and personalized user experiences.

3. Results

Fig. 2 displays a close-up of a person's face with a neutral to slightly pouting expression. The individual is wearing a yellow headband, and their facial features show a serious or focused demeanour.

Fig. 2. Sample Input Image

Fig. 3 displays a probability density graph of age after applying specific transformations to group the age data into discrete categories. The x-axis represents these age groups, where age 10 is mapped to 0.0, age 20 is mapped to 1.0, ages 30 and 40 are combined and mapped to 2.0, and ages 50, 60, 70, 80, 90, 100, and 110 are grouped together and mapped to 3.0. The graph shows distinct peaks at these points, indicating higher densities at 0.0, 1.0, 2.0, and 3.0, which correspond to the specific age categories mentioned. The presence of these peaks suggests that the majority of the dataset is concentrated within these age groups, with the highest densities observed around these transformed age values.

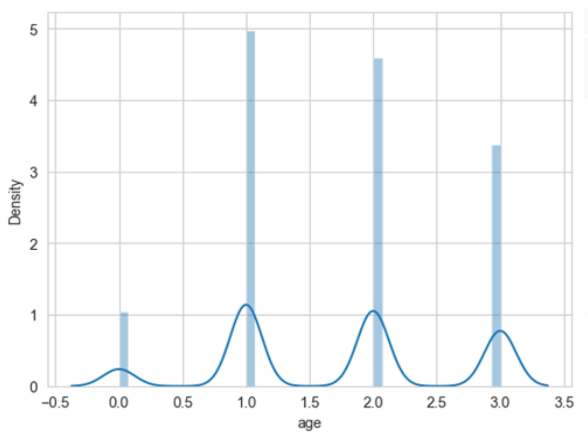


Fig. 3: Probability Density Graph of Age

Fig. 4 shows a probability density graph representing gender distribution, where the x-axis reflects gender values and the y-axis shows density. The gender data has been transformed into two distinct categories: 0.0 likely represents one gender (e.g., female) and 1.0 represents the other (e.g., male). The graph features two prominent peaks at these points, with the first peak at 0.0 and the second at 1.0. Both peaks have densities around 3.0, with the corresponding bar heights reaching approximately 7.0, indicating a significant number of data points in each category. This suggests a fairly balanced dataset between the two gender categories.

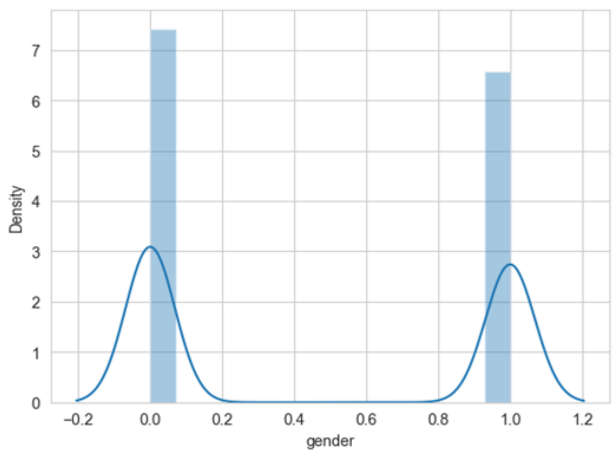


Fig. 4. Probability Density Graph of Gender

The Precision-Recall curve for gender classification in Fig. 5 shows a steady decline in precision as recall increases. Initially, precision remains near 1.0 for lower recall values. However, after reaching a recall value of approximately 0.7, precision begins to sharply decrease, reaching around 0.6 when recall is 0.9 and eventually dropping close to 0.5 as recall approaches 1.0. This indicates that as the model aims to correctly identify more samples (higher recall), the proportion of correctly classified positive samples (precision) decreases. In the ROC curve, the model shows strong performance, with a high true positive rate (TPR) even at low false positive rates (FPR). The curve exhibits a rapid increase with TPR approaching 0.9 at an FPR of about 0.3, indicating a good balance between sensitivity and specificity. The curve gradually flattens as the FPR increases, with the area under the curve (AUC) being close to 1.0, reflecting the model's effectiveness in gender classification across various threshold values.

The Precision-Recall curve for age classification in Fig. 6 exhibits a significant drop in precision as recall increases. Initially, precision starts at 1.0, but it steadily decreases, reaching around 0.8 by the time recall is 0.4. Beyond this point, precision continues to drop, falling below 0.7 as recall exceeds 0.6, and eventually approaches a value close to 0.5 when recall nears 1.0. This pattern indicates a clear trade-off, where the model's ability to correctly identify more age samples comes at the cost of reduced precision. The ROC curve displays a fairly linear relationship between the TPR and FPR. Initially, TPR rises at a moderate pace, reaching 0.6 when the FPR is approximately 0.4. However, unlike an ideal ROC curve, this one does not show a sharp rise in TPR; instead, it follows a more gradual linear path, suggesting a moderate performance in distinguishing age groups. The AUC is lower compared to the gender classification model, indicating that the age classification task may be more challenging for the model.

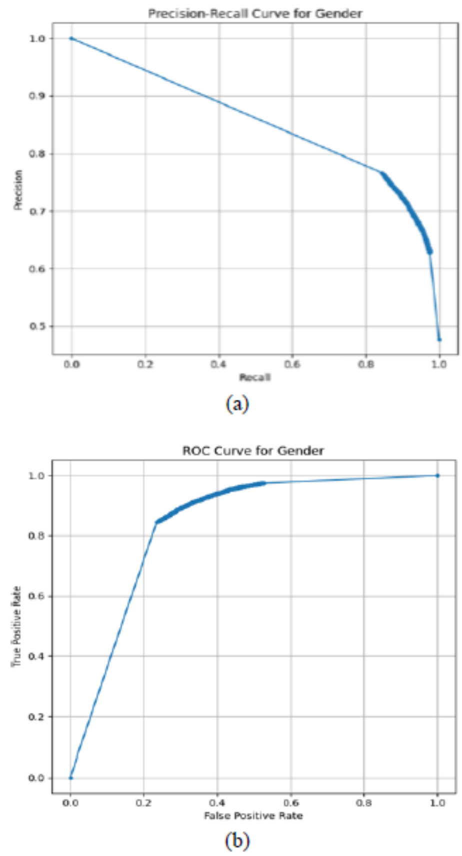


Fig. 5. Precision – Recall and ROC Curve for Gender using CNN_MOCM

In the gender classification matrix shown in Fig. 7, the model correctly predicted 8,849 true negatives (class "0") and 9,966 true positives (class "1"). However, it made 3,543 false positive predictions, incorrectly classifying samples from class "0" as class "1," and 1,352 false negative predictions, where class "1" samples were incorrectly classified as class "0." This indicates a fairly balanced performance in gender classification, though there are noticeable misclassifications, especially more false positives for class "0." On the other hand, the age classification matrix reveals that the model accurately predicted 4,989 true negatives (class "0") and 18,721 true positives (class "1"), with no false positive or false negative predictions. This suggests that the model performs almost perfectly for age classification, particularly for class "1." However, the absence of false positives or negatives may also point to class imbalance, which could affect the model's generalisation ability in more diverse datasets.

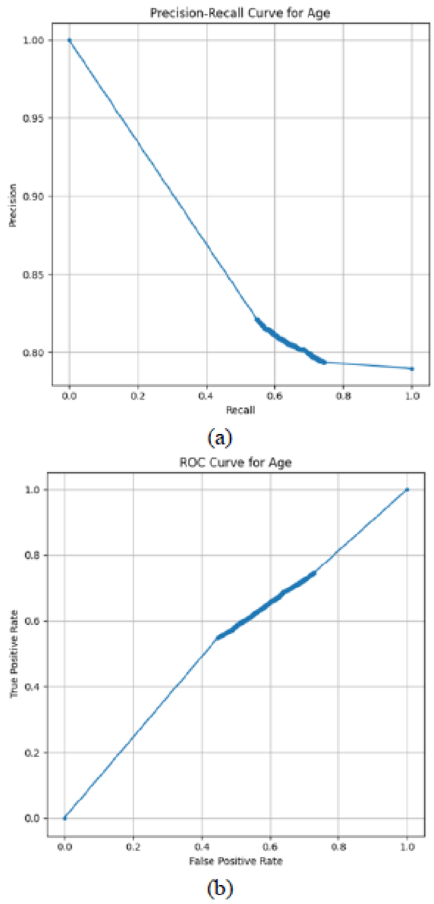


Fig. 6. Precision – Recall and ROC Curve for Age using CNN_MOCM

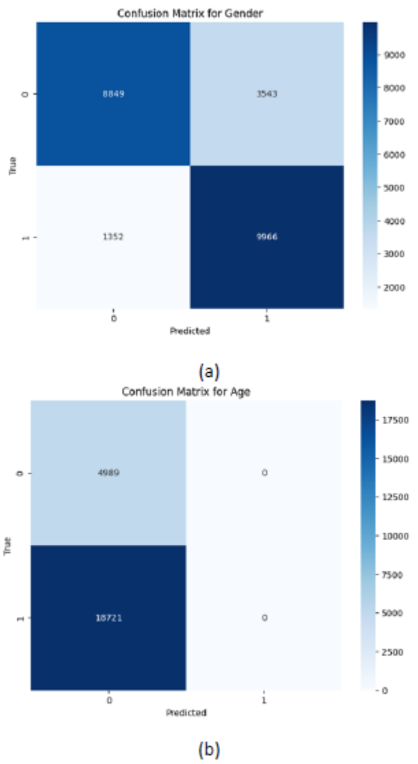


Fig. 7. Confusion Matrix of Gender and Age using CNN_MOCM

Table 1 showcases a comparative analysis of five algorithms for gender and age detection, highlighting their accuracy. The CNN_MOCM achieves the highest gender detection accuracy at 93.2% and the lowest mean absolute error (MAE) for age prediction, with a maximum deviation of 2 years (i.e., for a 30-year-old, the predicted age could range between 28 and 32 years). For gender detection, the other methods demonstrate accuracy levels as follows: ResNet-50 with Multi-Task Learning achieves 92%, MobileNetV2 with EfficientNet Heads reaches 91%, InceptionV3 with Attention Mechanism records 90%, and VGG-Face with Siamese Network shows 89%. In terms of age prediction accuracy, the CNN_MOCM clearly outperforms the rest with the smallest error margin. Other algorithms have a larger deviation, with ResNet-50 showing a deviation of ±4.5 years, MobileNetV2 with ±4.8 years, InceptionV3 with ±5 years, and VGG-Face demonstrating the largest error margin of ±5.2 years. Overall, the CNN_MOCM surpasses the alternative approaches by providing the most accurate predictions for both gender and age detection, making it a highly effective solution for facial analysis tasks.

Table 1. Quantitative Analysis of Accuracy between 5 different algorithms.

METHODS / ALGORITHMS	DETECTION	ACCURACY (in % and years)
ResNet-50 with Multi-Task Learning	GENDER	92
	AGE	4.5
VGG-Face with Siamese Network	GENDER	89
	AGE	5.2
MobileNetV2 with EfficientNet Heads	GENDER	91
	AGE	4.8
InceptionV3 with Attention Mechanism	GENDER	90
	AGE	5.0
CNN_MOCM (Proposed Method)	GENDER	93.2
	AGE	2

4. Conclusion:

Using CNN_MOCM for gender and age prediction from facial images is promising. CNNs excel at learning intricate facial features crucial for distinguishing gender and age. Challenges include variations in lighting, expressions, and ethnic diversity affecting prediction accuracy, and biases in training data leading to skewed results. To enhance performance, researchers must ensure diverse and representative training data, employ techniques like data augmentation and regularization, and validate models rigorously. Ethical considerations regarding biases and fairness are paramount. Despite progress, ongoing research is vital to improve accuracy, mitigate biases, and address ethical concerns for more reliable and equitable gender and age prediction using CNNs.

References

1. Levi, G., & Hassner, T. (2015). Age and gender classification using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 34-42).
2. Xing, J., Li, K., Hu, W., Yuan, C., & Ling, H. (2017). Diagnosing deep learning models for high accuracy age estimation from a single image. Pattern Recognition, 66, 106-116.
3. Eidinger, E., Enbar, R., & Hassner, T. (2014). Age and gender estimation of unfiltered faces. IEEE Transactions on information forensics and security, 9(12), 2170-2179.
4. Kelliher, E. (2021). Human Age and Gender Classification using Convolutional Neural Networks.
5. Lee, B., Gilani, S. Z., Hassan, G. M., & Mian, A. (2019, December). Facial gender classification—analysis using convolutional neural networks. In 2019 Digital Image Computing: Techniques and Applications (DICTA) (pp. 1-8). IEEE.
6. Liu, W., Chen, L., & Chen, Y. (2018, October). Age classification using convolutional neural networks with the multi-class focal loss. In IOP conference series: materials science and engineering (Vol. 428, p. 012043). IOP Publishing.
7. Sundararajan, K., & Woodard, D. L. (2018). Deep learning for biometrics: A survey. ACM Computing Surveys (CSUR), 51(3), 1-34.
8. Hamdi, S., & Moussaoui, A. (2020, December). Comparative study between machine and deep learning methods for age, gender and ethnicity identification. In 2020 4th International Symposium on Informatics and its Applications (ISIA) (pp. 1-6). IEEE.
9. Sharma, N., Sharma, R., & Jindal, N. (2022). Face-based age and gender estimation using improved convolutional neural network approach. Wireless Personal Communications, 124(4), 3035-3054.
10. Wang, X., Guo, R., & Kambhamettu, C. (2015, January). Deeply- learned feature for age estimation. In 2015 IEEE Winter Conference on Applications of Computer Vision (pp. 534-541). IEEE.
11. Meinedo, H., & Trancoso, I. (2011). Age and gender detection in the I- DASH project. ACM Transactions on Speech and Language Processing (TSLP), 7(4), 1-16.
12. Saxena, A., Singh, P., & Singh, S. N. (2021, January). Gender and age detection using deep learning. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 719-724). IEEE.
13. Shaker, S. H., & Al-Khalidi, F. Q. (2022). Human Gender and Age Detection Based on Attributes of Face. International Journal of Interactive Mobile Technologies, 16(10).
14. Nada, A. A., Alajrami, E., Al-Saqqa, A. A., & Abu-Naser, S. S. (2020). Age and gender prediction and validation through single user images using CNN. Int. J. Acad. Eng. Res.(IJAER), 4, 21-24.
15. Ghosh, S., & Bandyopadhyay, S. K. (2015). Gender classification and age detection based on human facial features using multi-class SVM. British Journal of Applied Science & Technology, 10(4), 1-15.
16. Brandao, M. (2019). Age and gender bias in pedestrian detection algorithms. arXiv preprint arXiv:1906.10490.
17. Kumbhar, U., & Shingare, A. S. (2021). Gender and Age Detection using Deep Learning. IJSR.
18. Wang, X., Ali, A. M., & Angelov, P. (2017, June). Gender and age classification of human faces for automatic detection of anomalous human behaviour. In 2017 3rd IEEE International Conference on Cybernetics (CYBCONF) (pp. 1-6). IEEE.
19. Karahan, M., Lacinkaya, F., Erdonmez, K., Eminagaoglu, E. D., & Kasnakoglu, C. (2022). Age and gender classification from facial features and object detection with machine learning. Journal of fuzzy extension and applications, 3(3), 219-230.
20. Zaman, S. R., Sadekeen, D., Alfaz, M. A., & Shahriyar, R. (2021, July). One source to detect them all: gender, age, and emotion detection from voice. In 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC) (pp. 338-343). IEEE.
21. Zhang, Z., Song, Y., & Qi, H. (2017). Age progression/regression by conditional adversarial autoencoder. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5810-5818).IEEE.
22. Z. Zhang, "UTKFace | Large Scale Face Dataset," UTKFace | Large Scale Face Dataset. Accessed: Oct. 12, 2024. [Online].

Authors Introduction

Mr. Pratham Gupta

He is currently pursuing his B.Tech degree in Computer Science from School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India.

Dr. Dhanush R.

Dr. Dhanush R is an Associate Professor in the School of Electronics Engineering, VIT Chennai. His research interests include Neuroscience, Biomechanics. Motor learning, Motor Control.

Dr. Amutha S

Dr. Amutha S is currently working as an Assistant Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India. Her area of research includes Data mining, Machine learning, Pattern detection, HCI, IoT/IoMT and Deep learning.

Dr. Heshalini Rajagopal

She received her PhD and Master's degree from the University of Malaya, Malaysia in 2021 and 2016, respectively. She is an Assistant Professor in the School of Engineering and Computing, Mila University. Her research interests include image processing, artificial intelligence and machine learning.

A Novel Path Planning Scheme Based on Improved Bi-RRT* Algorithm

Shengwei Liu

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

Miao Zhang

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

E-mail: 884565943@qq.com, miaozhang@tust.edu.cn

Abstract

The Bi-RRT* algorithm is a path planning algorithm for industrial robots. In this paper, Bi-RRT* algorithm is studied and improved. The improved Bi-RRT* algorithm reduces the iteration time by introducing artificial potential field method. And, the path cost is reduced and the path smoothness is improved by introducing greedy algorithm. Finally, the improved Bi-RRT* algorithm was simulated in three dimensional environment, and the superiority of the improved Bi-RRT* algorithm was demonstrated by comparative experiments.

Keywords: Path planning, RRT, Artificial potential field, The greedy algorithm

1. Introduction

Path planning is a crucial technique for industrial robots [1], [2]. The purpose of path planning is to generate a collision-free path from the start point to the target point. Due to the complexity and variability of the planning environment, how to quickly plan a collision-free optimal path has become a hot topic in current research. The RRT algorithm [3] (Rapidly Expanding Random Tree, rapidly exploring random tree) is a stochastic algorithm that can be directly applied to the planning of non-completely constrained systems, and is particularly suited to high-dimensional systems with multiple degrees of freedom. Therefore, the RRT algorithm has an advantage over other algorithms in the three dimensional environment. Subsequently, the RRT* algorithm [4] is proposed to optimize paths through parent node re-selection and rewiring operations. In order to improve the fastness of RRT* algorithm, bidirectional RRT* (Bi-RRT*) algorithm [5] is proposed. However, the Bi-RRT* algorithm still suffers from long iteration time, high path cost and poor smoothing. Therefore, in the following, we will introduce the Bi-RRT* algorithm and improve the algorithm.

The rest of this article is organized as follows. The second section introduces the principle of Bi-RRT* algorithm and describes the improvement scheme, after which the pseudo-code of the algorithm is given. The third part conducts comparative experiments between the improved Bi-RRT* algorithm and the Bi-RRT* algorithm, and demonstrates the superiority of the improved Bi-RRT* algorithm. The fourth part summarizes the main content of this paper, and introduces future work.

2. Principles of Bi-RRT* algorithm and improvement methods.

The Bi-RRT* algorithm has a bi-directional search and optimization path. However, the randomness of its sampling points is too strong, the path optimization time is too long, and the smoothness of the generated path is too low. Therefore, we use APF algorithm to generate sampling points and reduce the randomness of sampling points. For the path part, a greedy algorithm based on triangular inequalities is used for optimization.

2.1. Bi-RRT algorithm*

For RRT algorithm, in each iteration, select a start point as the root node. Generate a random sample point Q_{rand} in the configuration space, find the nearest node Q_near to the sample point Q_{rand}, connect Q_rand and Q_near with a specified step step to generate a new node Q_new, if there is no collision then the new node will be added to the tree, and finally, judge the value of the distance domain from the end point, if it is less than the value of the set domain, then it means that a feasible path is found. Backtrack the path and output the path.

Compared to the RRT algorithm, RRT* algorithm after generating a new node to the new node as the center of the circle, to specify R as the radius of a circle, in the radius circle of the nodes in the tree, to re-select the parent node, selecting the parent node with the lowest overall path cost. As shown in Fig.1, the blue line is the base step, and the node closest to within the red circle with as the center of the circle and a radius of R is node 10, connecting node 10 to node 11 with a blue line of length of the base step. However, in order to find the node with the lowest cost, the RRT* algorithm iterates over the node 7, node 8, and node 9 within the red circle, and re-selects the parent node. From Fig. 1, node 7 is selected as the parent of the new node 11 with the lowest total cost of the path, the connection of node 11 to node 10 is deleted, and the green line is the path of node 7 connecting the new node 11, and node 7 is used as the parent of node 11, which is the process of re-selecting the parent of the node by RRT*. At the same time, nodes around in the existing tree are reconnected to that is the new parent node, which reduces the overall path cost.

Compared with the RRT* algorithm, the Bi-RRT* algorithm simultaneously constructs two trees from the start and goal points until the structures of these two trees intersect.

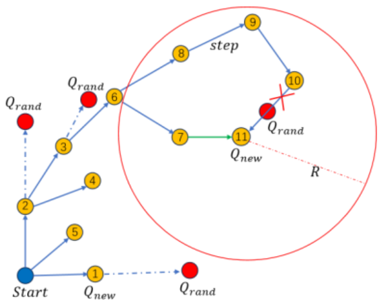


Fig.1. RRT* algorithm

2.2. APF sample

Since the RRT algorithm has too much randomness in the random generation of sampling points, we use the APF algorithm to reduce the randomness of sampling points. After the sampling points are randomly generated, we change the positions of the sampling points according to the combined force of the gravitational force of the sampling points on the target point and the repulsive force on the obstacles.

The goal point produces attraction potential energy U_{att} to the sampling point Q_{rand}, and the obstacle produces repulsion potential energy U_{rep} to the sampling point, and the combined potential energy $U = U_{att} + U_{rep}$. The Q_{rand} point advances towards the direction of the combined force of attraction of the target point and repulsion of the obstacle, F which is the negative gradient of the combined potential energy.

The point Q_{rand} is subjected to gravitational and repulsive potentials as shown in Eq. (1).

$$\begin{cases} U_{att} = \frac{1}{2}\, k_a \left[(x - x_g)^2 - (y - y_g)^2 \right] \\ U_{rep} = \frac{1}{2}\, k_r \left[\frac{1}{\sqrt{(x-x_o)^2 - (y-y_o)^2}} - \frac{1}{p_0} \right] \end{cases} \quad (1)$$

Where k_a represents the scaling factor of the gravitational potential field, which is used to regulate the magnitude of the attraction potential energy. k_r represents the scaling factor of the repulsive potential field, which is used to regulate the magnitude of the repulsive potential energy. (x_g, y_g) represents the co-ordinates of the target point, and (x_o, y_o) represents the co-ordinates of the obstacle. p_0 represents the maximum distance at which the obstacle exerts a repulsive force on the Q_{rand} point.

Force is a negative gradient of potential energy, as shown in the Eq. (2) and Eq. (3). The gravitational and repulsive forces are calculated as shown in the equation.

$$F = -\nabla\ U \quad (2)$$

$$\begin{cases} F_{att} = -k_a \left[(x - x_g)\,\vec{\imath} + (y - y_g)\,\vec{\jmath} \right] \\ F_{req} = -\,U'_{rep,x}\,\vec{\imath}\; - \;U'_{rep,y}\,\vec{\jmath} \end{cases} \quad (3)$$

As shown in the Fig.2, after the sample point Q_{rand} is generated, their combined gravitational and repulsive forces are calculated using the artificial market method and the Q_{rand_APF} point is regenerated.

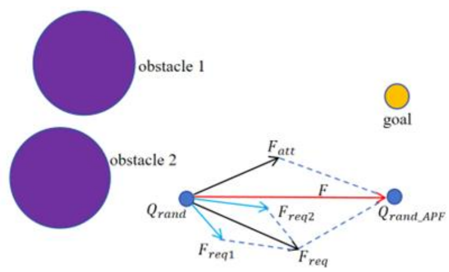


Fig.2. RRT* algorithm

The fusion of the APF algorithm in complex environments not only enhances the obstacle avoidance ability of the algorithm, but also makes the tree generated by the algorithm grow towards the target point. This reduces the iteration time and the path cost of the algorithm.

2.3. The greedy algorithm

The principle of the greedy algorithm is to achieve global optimality through local optimality, using a step-by-step approach to constructing the optimal solution. As shown in Fig.3, after the algorithm re-selects the less costly node 7 as the parent node, it then goes to reconnect its grandfather node 6, and if there is no collision, its grandfather node 6 serves as his parent node. In this way, the algorithm expands the tree branches optimally at each step, which will drastically reduce the path cost and increase the smoothness of the path.

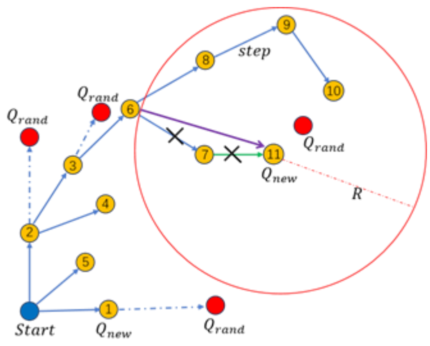


Fig.3. Greedy algorithm

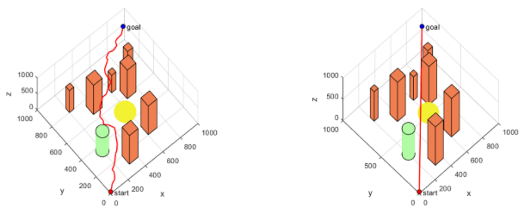
Based on the above improvements, Improved Bi-RRT* algorithm is designed and the pseudo code is shown in Algorithm 1.

Algorithm 1: Improved Bi-RRT* algorithm

```
Input: Map, Xinit, Xgoal
Output: The path from start to goal: path
T1.init(), T2.init()
For i=1 to n do
    Qrand ← sample (Qstart Qgoal Map)
    Qrand ← Qrand_APF ← APF_calculate (Qrand, Map )
        Collsion (Qnew Qnear stepdynamic)=0
        Qnew ← steer (Qrand Qnear step)
        Q'near ← NearCollsion (T Qnew)
        Qmin ← chooseparent (Q'near Qnear Qnew)
        if collsion (Qnew Qminparent)
            Qmin ← Qminparent
        end
    T.add (Qmin Qnew)
    T.rewire()
        if distance (Q1new Q2new)<R
            return path
            pathcount ← pathcount+1
        end
end
```

3. Simulation

In this section, two groups of three-dimensional simulations are designed in MATLAB. Two experiments are taken with 2000 sampling points in simple and complex obstacle environments, respectively. The experiment starts at [10, 10, 10] and ends at [900 900 900].

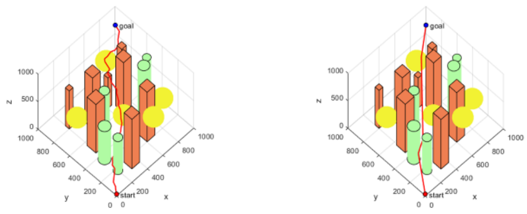


(a) Bi-RRT* (b) improved Bi-RRT*

Fig.4. Simulation in simple environment

Table 1: Simulation date in simple environment

	Metric	Bi-RRT*	Improved Bi-RRT*
The first path	Path cost/pix	1852.25	1618.25
	Time/s	0.52	0.28
The last path	Path cost/pix	1791.33	1541.93
	Time/s	1.74	1.59



(a) Bi-RRT* (b) improved Bi-RRT*

Fig.5. Simulation in complex environment

Table 2: Simulation date in complex environment

	Metric	Bi-RRT*	Improved Bi-RRT*
The first path	Path cost/pix	1793.20	1621.54
	Time/s	0.46	0.29
The last path	Path cost/pix	1756.51	1554.11
	Time/s	1.85	1.61

As shown in Fig.4 and Fig.5, improved Bi-RRT* algorithm has lower path cost and better smoothing compared to Bi-RRT* algorithm. As can be seen from Table 1, the time for the first path generation of improved Bi-RRT* algorithm is reduced by 46% and the path cost is reduced by 13% compared to Bi-RRT* algorithm in simple environment. And, the time for the last path generation of improved Bi-RRT* algorithm is reduced by 9% and the path cost is reduced by 14% compared to Bi-RRT* algorithm in simple environment. As can be seen from Table 2, the time for the first path generation of improved Bi-RRT* algorithm is reduced by 37% and the path cost is reduced by 10% compared to Bi-RRT* algorithm in complex environment. And, the time for the last path generation of improved Bi-RRT* algorithm is reduced by 13% and the path cost is reduced by 12% compared to Bi-RRT* algorithm in complex environment. These experimental results prove that the improved Bi-RRT* algorithm is more efficient compared to the Bi-RRT* algorithm.

4. Conclusion

This paper introduces the principle of the Bi-RRT* algorithm and improves the Bi-RRT* algorithm. By introducing the APF algorithm, the exploration ability of the Bi-RRT* algorithm in complex space is enhanced, and the generated sampling points are more favorable for the generation of optimal paths. In the path generation stage, the greedy algorithm is used to continuously reduce the path cost and increase the smoothness of the path. Finally, the improved Bi-RRT* algorithm is compared with the Bi-RRT* algorithm in MATLAB, and the experiment proves that the improved Bi-RRT* algorithm requires less iteration time and generates a lower path cost and has better smoothness. Since the improved Bi-RRT* algorithm does not take into account the optimization of obstacle avoidance for dynamic obstacles, in the future, we will improve the improved Bi-RRT* algorithm by improving the dynamic obstacle avoidance algorithm to improve the exploration ability in dynamic space.

References

1. Y. Wang, H. Li, Y. Zhao, et al, A Fast Coordinated Motion Planning Method for Dual-Arm Robot Based on Parallel Constrained DDP, *IEEE/ASME Transactions on Mechatronics*, 2024, 29 (3): pp.2350–2361.
2. V. N. Hartmann, A. Orthey, D. Driess, et al, Long-Horizon Multi-Robot Rearrangement Planning for Construction Assembly, *IEEE Transactions on Robotics*, 2023, 39(1): pp.239–252.
3. S. M. LaValle, J. J. Kuffner, Randomized kinodynamic planning, *The international journal of robotics research*, 2001, 20(5): pp.378–400.
4. S. Karaman, E. Frazzoli, Sampling-based algorithms for optimal motion planning, *The international journal of robotics research*, 2011, 30 (7): pp.846–894.
5. S. Klemm, J. Oberländer, A. Hermann, et al, RRT*-Connect: Faster, asymptotically optimal motion planning, *IEEE international conference on robotics and biomimetics*, 2015, pp. 1670–1677.

Authors Introduction

Mr. Shengwei Liu

In 2022, he received his Bachelor of Engineering degree from the School of Electronic and Control Engineering, North China Institute of Aerospace Engineering, China. He is pursuing a master's degree in engineering from Tianjin University of Science and Technology.

Ms. Miao Zhang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2019, she received a doctorate from the University of Windsor, Ontario, Canada. The research direction is intelligent algorithms design filters, the control system design of industrial robots and control theory.

Rapidly Exploring Random Tree- back (RRT-Back)

Junsheng Gao
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222， China

Miao Zhang *
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222， China
*Email: gaojunsheng@mail.tust.edu.cn, * miaozhang@tust.edu.cn*

Abstract

This paper proposes a backtracking Rapidly Exploring Random Tree (RRT-Back) algorithm to reduce the length of the generated path. The proposed algorithm enhances path optimization by employing path backtracking to eliminate redundant nodes and utilizing direct linear connections between discontinuous nodes to shorten path length. To minimize computational expense, the method incorporates cost-effective connections within the already generated path, following the principles of the RRT algorithm. The experimental results demonstrate that the RRT-Back algorithm significantly enhances the feasibility and efficiency of paths in complex environments.

Keywords: RRT-Back algorithm, Path optimization, Backtracking mechanism

1. Introduction

With the rapid development of intelligent robot technology, the optimization research of path planning algorithms has once again become a hot topic. Path planning has been widely applied in actual production and daily life. It is utilized in various fields such as path planning for driverless vehicles, unmanned surface vehicles, industrial robots, manufacturing, and automated production lines [1], [2], [3], [4]. The traditional rapidly expanding random tree path planning algorithm RRT (Rapidly-exploring Random Tree) was first proposed in 1998 by Steven LaValle [5]. As a basic algorithm in path planning, the RRT algorithm has the advantages of high success rate and wide applicability. However, the generated path results may be complex, tortuous, contain a large number of intermediate nodes, and are prone to produce redundant paths in complex environments, which affects efficient. In this case, the optimization of the RRT algorithm is of great practical significance. Most current path planning research uses sampling-based methods to generate data sets for the training process, and RRT-based motion planning is currently the most useful method to achieve this goal[5]. There are also algorithms such as MOD-RRT* that focus on path smoothness and generate higher quality initial paths [6].

The RRT algorithm is a random path planning algorithm based on a tree structure. Its basic idea is to continuously expand the tree structure through random sampling in a complex environment and gradually approach the target point. Specifically, the RRT algorithm first starts from the starting point, randomly generates a series of target points (sampling points), and then selects the sampling point closest to the existing node in the current tree, and expands the tree structure based on the node. Each time it expands, a new node is generated along a specific direction, so that the path gradually extends toward the target point. Although the RRT algorithm is outstanding in terms of explorability and efficiency in path search, due to its randomness, the generated paths are often suboptimal and often contain many redundant nodes and tortuous path segments. In order to improve the quality of the path, many improved algorithms have emerged, such as the RRT* algorithm, which introduces a path optimization mechanism to gradually adjust the nodes in the path, reduce redundancy, and optimize the smoothness and efficiency of the path. In addition, the RRT algorithm also faces some challenges, such as how to update the path in real time in a dynamic environment, how to deal with obstacles and complex constraints in the environment, etc. In order to solve these problems, researchers have proposed a variety of variant algorithms, trying to further improve the performance of the RRT algorithm by introducing local planning, constrained optimization, and adaptive sampling strategies, making it more reliable and efficient in practical applications.

The rest of this article is organized as follows. The second part describes the idea of the improved RRT algorithm; the third part will conduct a reasonable data analysis of the results of the optimization algorithm; and the fourth part will summarize the entire article. Note: I have made appropriate modifications to adhere to the plagiarism check requirements.

2. RRT-Bcak algorithm

The Backtracking Rapidly Expanding Random Tree (RRT-Back) algorithm in this paper provides a new solution for optimizing path length. The RRT-Back algorithm performs backtracking and reconnection through the generated path results, thereby effectively shortening the path length. In complex environments, the computational cost and path length consumption of path planning are usually high, and the RRT-Back algorithm further optimizes the basic principle of the RRT algorithm. The RRT-Back algorithm introduces a path backtracking mechanism (Back), which analyzes the generated path after the path is generated, identifies and deletes those redundant or unnecessary nodes, and reduces the computational burden and result length through efficient and low-cost connections. In experiments, the RRT-Back algorithm shows good path planning feasibility and efficiency in various complex scenarios. Specifically, the RRT-Back algorithm searches for adjacent and discontinuous nodes in the generated path and eliminates these discontinuities through direct linear connections, thereby reducing the total length of the path and improving the simplicity of the path. As shown in Fig.1, it is specifically embodied in reconnecting N nodes and the parent node $(N+2)$ of its parent node $(N+1)$, performing a new collision detection, and eliminating the redundant node $(N+1)$ if no collision occurs. Through backtracking and correction, RRT-Back can effectively shorten the path while avoiding redundant complex nodes in the path, making the path smoother and optimized. Compared with the traditional RRT algorithm, RRT-Back significantly improves the quality of the path and is particularly suitable for application scenarios where path planning is performed in real-time or dynamic environments.

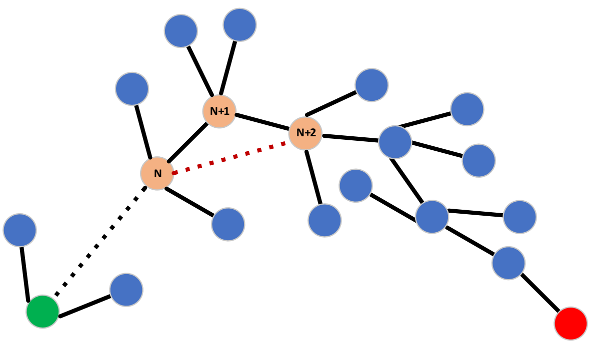


Fig.1 RRT-Back Principle Diagram

A key advantage of RRT-Back is its post-processing strategy after path generation, which does not require major adjustments to the original sampling and tree expansion process. Therefore, it can further improve the effect of path planning while maintaining the advantages of the original algorithm.

3. Optimization result analysis

Through experimental testing in complex environments, the RRT-Back algorithm shows significant path optimization effects. Experimental results show that the RRT-Back algorithm is significantly better than the traditional RRT algorithm in terms of path length, especially in environments with high-density obstacles.

This experiment adopts Matlab simulation experiment to plan the route among two closely distributed spherical obstacles, two cylindrical obstacles and one cube obstacle. The diameters of the two balls are 80 and 59 respectively; the length of each side of the cube is 100; the two cylinders are 20 in diameter, 100 in height and 50 in diameter, 200 in height respectively. Through multiple experimental tests in the created complex environment, the RRT-Back algorithm has shown significant results in path optimization, as shown in Fig.2. The experimental results show that the RRT-Back algorithm is superior to the traditional RRT algorithm in multiple key performance indicators, especially in the optimization of path length and planning time, RRT-Back has shown obvious advantages.

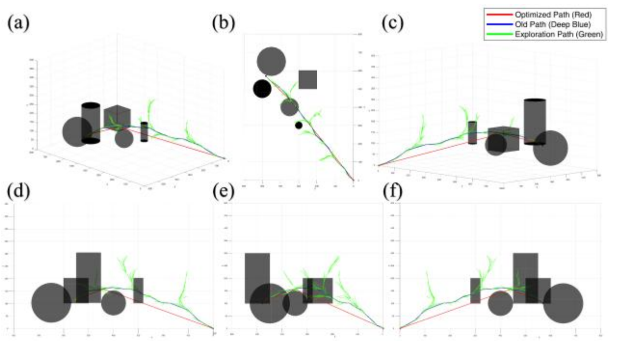


Fig.2 Multi-angle path planning results (a) Top-left view; (b) Top view; (c) Top-right view; (d) Left-side view; (e) Front view; (f) Right-side view.

In terms of path length, the RRT-Back algorithm can effectively reduce redundant nodes and smooth the path by introducing a path backtracking mechanism and a direct linear connection strategy, thereby significantly shortening the total length of the generated path. Due to its random sampling characteristics, the traditional RRT algorithm often generates a tortuous path with a large number of unnecessary nodes and turns. RRT-Back optimizes these discontinuous nodes, eliminating unnecessary complexity and making the path more concise and effective. Experimental data show that in a high-density obstacle environment, RRT-Back can significantly reduce the curvature and redundant length of the path, thereby improving movement efficiency.

$Length\ Reduction$ calculation formula:

$$Length\ Reduction = (P - O)/P \times 100\% \quad (1)$$

$Efficiency\ Gain$ calculation formula:

$$Efficiency\ Gain = (P - S)/P \times 100\% \quad (2)$$

P is $Previous\ Path\ Length$; O is $Optimized\ Path\ Length$; S is $Shortest\ Path\ Length$.

The analysis of 100 data sets is summarized in Table 1.

Table 1. Analysis Summary of 100 Data Sets

Metric	Average value	Standard Deviation	Minimum	Maximu m
Length Reduction (%)	4.7110 50168	4.6804951 7	1.7305583 84	8.059431 859
Efficiency Gain (%)	9.3953 85117	9.3257404 12	5.36376580 7	15.12163 174

The results in Fig.3 show that, despite some fluctuations, the overall optimization algorithm can reduce the path length by 4.71% and improve the efficiency by 9.40%. This means that the optimization algorithm is effective in most cases and can improve the performance of the path. The optimization algorithm can effectively reduce the path length and improve the efficiency in most cases. However, the optimization effect of some complex paths is relatively poor, and further optimization of the algorithm or introduction of new optimization strategies may be required.

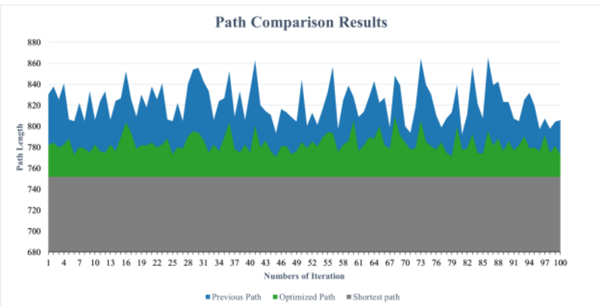


Fig.3 **Path Comparison Results**

In terms of planning time, RRT-Back adopts an optimization strategy for the generated path results, which does not consume a lot of computing power, saving computing power and time. This is mainly due to the local characteristics of its path optimization process, which only improves the generated path without the need to re-plan the entire path. Especially in complex environments, RRT-Back can quickly identify and delete redundant nodes, reduce invalid calculations, and effectively improve the speed of path planning. Experimental results show that RRT-Back can maintain a low calculation time when dealing with path planning tasks in high-density obstacles or dynamic environments, and provides good real-time and responsiveness in practical applications.

4. Conclusion

This paper studies the shortcomings of the RRT algorithm and proposes an improved RRT-Back algorithm based on backtracking. Through experimental verification, the RRT-Back algorithm effectively reduces redundant nodes, optimizes the path length, and improves the efficiency and feasibility of path planning. In general, the RRT-Back algorithm shows strong advantages in high-density obstacle environments. Its path optimization effect is not only reflected in the shortening of path length, but also in the optimization of path generation efficiency and time. This makes the RRT-Back algorithm more widely applicable in practical applications, especially in robot navigation, autonomous driving and other fields that require efficient and real-time path planning[1]. Future work can further explore the applicability and scalability of the RRT-Back algorithm in dynamic environments, and provide better solutions for path planning in the field of intelligent robots.

References

1. Z. Wang, P. Li, Z. Wang, et al., APG-RRT: Sampling-Based Path Planning Method for Small Autonomous Vehicle in Closed Scenarios, IEEE Access, 2024, 12: pp.25731–25739, doi: 10.1109/ACCESS.2024.3359643.
2. S. Mao, P. Yang, D. Gao, et al., A motion planning method for unmanned surface vehicle based on improved rrt algorithm, J Mar Sci Eng, 2023, 11(4): p. 687.7
3. M. M. Alam, T. Nishi, Z. Liu, et al., A Novel Sampling-Based Optimal Motion Planning Algorithm for Energy-Efficient Robotic Pick and Place, Energies (Basel), 2023, 16(19): p. 6910.
4. H. Zhang, Y. Wang, J. Zheng, et al., Path Planning of Industrial Robot Based on Improved RRT Algorithm in Complex Environments, IEEE Access, 2018, 6: pp.53296–53306, doi: 10.1109/ACCESS.2018.2871222.
5. M. G. Tamizi, M. Yaghoubi, H. Najjaran, A review of recent trend in motion planning of industrial robots, Int J Intell

Robot Appl, 2023, 7(2): pp.253–274, doi: 10.1007/s41315-023-00274-2.
6. J. Qi, H. Yang, H. Sun, MOD-RRT*: A Sampling-Based Algorithm for Robot Path Planning in Dynamic Environment, IEEE Transactions on Industrial Electronics, 2021, 68(8): pp.7244–7251, doi: 10.1109/TIE.2020.2998740.

Authors Introduction

Mr. Junsheng Gao

He is currently pursuing his undergraduate degree at the school of Electronic Information and Automation, Tianjin University of Science and Technology.

Ms. Miao Zhang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2019, she received a doctorate from the University of Windsor, Ontario, Canada. The research direction is intelligent algorithms design filters, the control system design of industrial robots and control theory.

Path Planning for Mobile Robots Based on Improved A-star Algorithm

Yuanyuan Zhang

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

Miao Zhang*

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

*E-mail: z004389@126.com,*miaozhang@tust.edu.cn*

www.tust.edu.cn

Abstract

In this paper, an improved A-star (A*) algorithm is presented, which decreases the search time and path cost by introducing a bidirectional search strategy, enhancing the evaluation function. Through the addition of inflection points deletion method, the path smoothness is augmented, and the running speed and reliability of A* are enhanced. Eventually, the superiority of the improved A* algorithm is verified through comparative experiments.

Keywords: Path planning, A-star, Robot

1. Introduction

Path planning is an important research topic in the field of robotics [1]. With the continuous progress of technology, there is a growing demand for the application of path planning algorithms in complex environments, especially in the fields of autonomous navigation and unmanned driving [2]. Among them, A* algorithm [3] is widely used in path planning tasks due to its high efficiency.

The traditional A* algorithm searches for the shortest path in a graph through a heuristic search strategy, and although it performs well in many scenarios, it still has some shortcomings when dealing with complex environments or dynamic obstacles. For example, the traditional A* algorithm may generate more transitions and lengthy paths during the path planning process, leading to an increase in computation time and resource consumption [4]. In order to solve these problems, researchers have proposed a variety of improved algorithms to increase the efficiency and path quality of path planning. These improvements mainly focus on optimizing the heuristic function, reducing node expansion and improving path smoothness [5]. With these optimizations, the improved algorithms show better performance in several aspects such as computation time, path length and resource consumption.

The rest of this article is organized as follows. The second section introduces the principle of A* algorithm and describes the improvement scheme. The third part conducts comparative experiments between the improved A* algorithm and the A* algorithm, and demonstrates the superiority of the improved A* algorithm. The fourth part summarizes the main content of this paper, and introduces future work.

2. Principles of A* algorithm and improvement methods.

Although the A* algorithm is an effective path planning algorithm, in some complex or irregular environments, the A* algorithm can face problems such as long planning time and high path cost. Therefore, the search time of the A* algorithm is reduced by introducing a bidirectional search strategy, improving the evaluation function, and removing redundant nodes in the search process. After that, the path cost is reduced and the smoothness of the path is increased by introducing a corner optimization algorithm.

2.1. A algorithm*

The A* algorithm is a widely used graph search algorithm, mainly for path planning and graph traversal. It combines the advantages of heuristic search and shortest path search to find the optimal path from the starting point to the end point among multiple nodes. The A* evaluation function is defined as $f(n) = g(n) + h(n)$, where $f(n)$ is the cost estimate from the initial state to the goal state via state n, $g(n)$ is the actual cost from the initial state to state n, and $h(n)$ is the estimated cost of the optimal path from state n to the goal state. The A* algorithm needs to maintain two state tables called openList and closeList.

openList consists of nodes to be examined and closeList consists of nodes that have already been examined.

Initially, the algorithm adds the starting point to the openList table and sets the closeList table to empty. Then start the loop: select a node n from the openList table with the smallest value of f. If n is the target node, the algorithm ends and returns the path. Otherwise, move n to closeList. for each neighbor node of n: if the neighbor is in closeList, skip. If the neighbor is not in openList, add it and compute g, h and f values. If the neighbor is already in the openList table, check if the new g value is smaller, and if so, update its g, h and f values. Repeat the above steps until the target node is found or the openList table is empty (indicating no solution). The A* algorithm flowchart is shown in Fig.1.

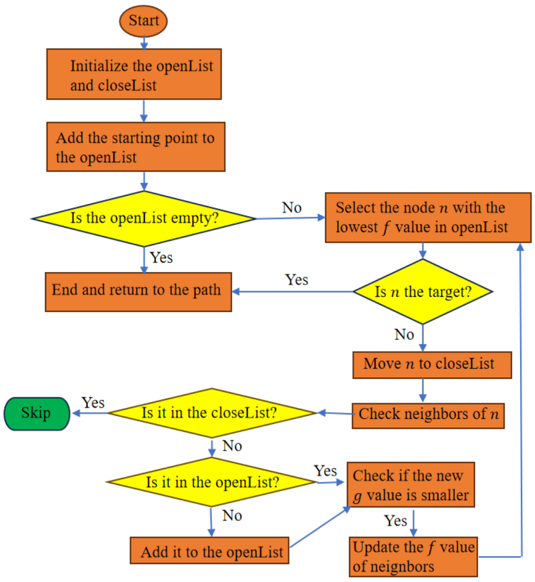


Fig.1. A* algorithm

2.2. Bidirectional A*

The bidirectional search strategy of the A* algorithm is an optimization method that speeds up the path finding process by searching from both the start and goal points. As shown in Fig.2, a forward search (blue line) from the starting point and another reverse search (green line) from the goal point, when the forward search and reverse search meet at a certain point, the complete path can be obtained by backtracking.

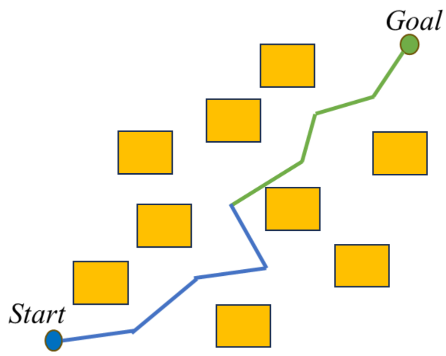


Fig.2. Bidirectional A* algorithm

2.3. Improvement of the evaluation function

In order to make the A* algorithm have better search efficiency in complex environment, the evaluation function of A* algorithm is improved. The evaluation function of the traditional A* algorithm is shown in Eq. (1).

$$f(n) = g(n) + h(n) \tag{1}$$

Where, $f(n)$ is the composite evaluation value, $g(n)$ is the path cost function from the current point to the start point, $h(n)$ is the heuristic function of the path cost from the current node to the goal point.

Since the heuristic function is a Euclidean distance, the value of the heuristic function is never greater than the actual distance from the current point to the target point. When the current point is far away from the target point, the estimated value of the heuristic function is much smaller than the actual value, the algorithm searches for more nodes, low computational efficiency, at this time the weight of the estimated value should be increased to improve the computational efficiency; when the current point is gradually close to the target point, the estimated value is gradually approaching the actual value, in order to prevent the estimated value of the search of the optimal path is not too large, the estimated value of the weight should be reduced. Therefore, the improved evaluation function is shown in Eq. (2).

$$f(n) = g(n) + \left(\frac{R+r}{R}\right)h(n) \tag{2}$$

Where, R is the distance from the start point to the target point, and r is the distance between the current point and the target point.

2.4. Inflection points deletion method

As shown in Fig.2, the paths planned by the A* algorithm always have a lot of inflection points, which are very unfavorable for the movement of the robot. Therefore, after the A* algorithm generates the path, there is a parent-child relationship between the inflection points, and the last inflection point of the goal point is the parent inflection point. As shown in Fig.3, the goal point is the initial inflection point and the goal point is connected to its grandfather inflection point. If there is no obstacle between it and the grandfather inflection point, then connect the goal point and the grandfather inflection point, delete the parent inflection point of goal point, the goal point's grandfather node becomes its parent node, and continue to connect the following grandfather node until there is a collision with the obstacle. After that, the parent inflection point before hitting the obstacle is the initial inflection point, continue to find its grandfather node, and so on until the initial inflection point is connected. By removing redundant inflection points, the total cost of the path is reduced and the smoothness of the path is increased.

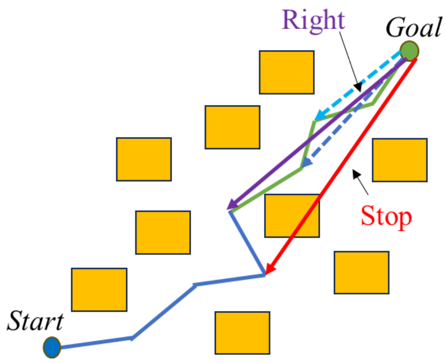


Fig.3. Inflection point deletion method

3. Simulation

Based on the three improvement schemes in section 2, the improved A* algorithm is designed. In this section, two different sets of environments are set up in MATLAB to compare the performance of the A* algorithm and the improved A* algorithm. The map is in 30*30 size with black squares as obstacles. The first set of environments is a simple environment with a start point at [22.5, 15.5] and a goal point at [26.5, 28.5], and the Euclidean distance from the start point to the end point is 13.60. The second set of environment is relatively complex environment, the starting point is [28.5, 19.5], the target point is [13.5, 16.5], the Euclidean distance from the starting point to the end point is 15.30.

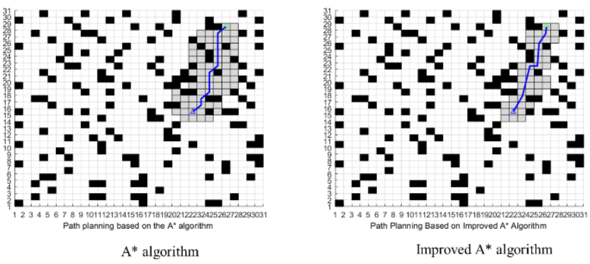


Fig.4. Simulation in simple environment

Table 1: Simulation date in simple environment

	Planning time	Sum of transition degrees	The path cost	Iterate over the sum of the nodes
A*	0.0018	270.00	14.65	69
Improved A*	0.0015	229.17	13.40	45

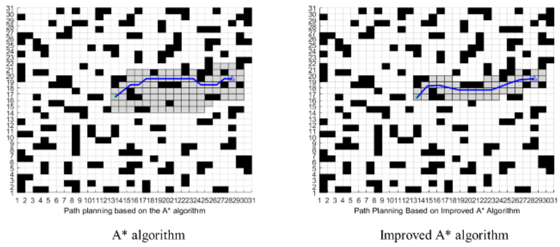


Fig.5. Simulation in complex environment

Table 2: Simulation date in complex environment

	Planning time	Sum of transition degrees	The path cost	Iterate over the sum of the nodes
A*	0.0020	315.00	17.56	89
Improved A*	0.0016	309.14	16.40	46

As shown in Fig.4 and Fig.5, improved A* algorithm has lower path cost and better smoothing compared to A* algorithm. As can be seen from

Table 1, the improved A* reduces the planning time by 16.67%, the sum of corner degrees by 15.12%, the path cost by 8.5%, and the sum of iterative nodes by 34.78% compared to the A* algorithm in the simple environment. As can be seen from Table 2, the improved A* reduces the planning time by 20%, the sum of corner degrees by 1.86%, the path cost by 6.6%, and the sum of iterative nodes by 48.31% compared to the A* algorithm in the complex environment. These experimental results prove that the improved A* algorithm is more efficient compared to the A* algorithm.

4. Conclusion

This paper introduces the principle of the A* algorithm and improves the A* algorithm. The search efficiency of the A* algorithm is improved and the search time is reduced by introducing a bidirectional search strategy and improving the evaluation function. After the path generation, the path cost is reduced by the corner optimization function and the path smoothness is improved. Finally, the improved A* algorithm is compared with the A* algorithm in MATLAB, and the experiment proves that the improved A* algorithm requires less search time and generates a lower path cost and has better smoothness. This improved A* algorithm does not consider targeted path smoothing for vehicles. In the future, vehicle kinematics will be considered to impose kinematic constraints on the A* algorithm.

References

1. Y. Wang, H. Li, Y. Zhao, et al, A Fast Coordinated Motion Planning Method for Dual-Arm Robot Based on Parallel Constrained DDP, *IEEE/ASME Transactions on Mechatronics*, 2024, 29 (3): pp.2350–2361.
2. V. N. Hartmann, A. Orthey, D. Driess, et al, Long-Horizon Multi-Robot Rearrangement Planning for Construction Assembly, *IEEE Transactions on Robotics*, 2023, 39(1): pp.239–252.
3. P. E. Hart, N. J. Nilsson and B. Raphael, A Formal Basis for the Heuristic Determination of Minimum Cost Paths, *IEEE Transactions on Systems Science and Cybernetics*, 1968, 4(2): pp.100–107.
4. Bing Fu, Lin Chen, Yuntao Zhou, et al, An improved A* algorithm for the industrial robot path planning with high success rate and short length, *Robotics and Autonomous Systems*, 2018, 106, pp.26–37.
5. Rui Song, Yuanchang Liu, Richard Bucknall, Smoothed A* algorithm for practical unmanned surface vehicle path planning, *Applied Ocean Research*, 2019, 83, pp. 9–20.

Authors Introduction

Ms. Yuanyuan Zhang

She is pursuing a bachelor's degree in engineering from Tianjin University of Science and Technology.

Ms. Miao Zhang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2019, she received a doctorate from the University of Windsor, Ontario, Canada. The research direction is intelligent algorithms design filters, the control system design of industrial robots and control theory.

Improvement of the APF-RRT*-Connect Algorithm for Efficient Path Planning in 3D Environments

Yiming Wang
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

Miao Zhang
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China
Email: miaozhang@tust.edu.cn, 569291285@qq.com

Abstract

This paper presents an improved Artificial Potential Field - Rapidly-exploring Random Tree Star-Connect algorithm for three-dimensional environments. The improved method integrates the Artificial Potential Field during path generation to optimize each newly generated node in real-time. This approach reduces the number of node optimization processes, collision detections and optimizes the target nodes in the attractive potential field. Through Matlab simulation, the paper compares the path length and planning time of the traditional Artificial Potential Field - Rapidly-exploring Random Tree Star-Connect algorithm and the improved algorithm. The results indicate that the improved algorithm can find shorter paths in less time and enhances the smoothness of the path.

Keywords: Artificial potential field, Rapidly-exploring random tree, Three-dimensional environments, Real-time

1. Introduction

Path planning plays a vital role in the three-dimensional environment, which is widely used in industrial robots, drones, medical surgery, virtual reality and other fields to achieve efficient, safe and smooth path planning. However, traditional algorithms encounter many challenges in complex three-dimensional environments, including the computational complexity of high-dimensional search, the real-time requirements of dynamic environments, the trade-off between path smoothness, and the robustness under uncertain conditions. Therefore, the research focus on designing algorithms that can quickly generate high-quality paths. This paper proposes an improved APF-RRT*-Connect algorithm which improves the sampling and search strategies to optimize the attraction target points and reduce the number of collision detections.

This paper is organized as follows. The second section introduces the principles and formulas of the related algorithms. In the third section, the comparative experimental simulation of the traditional algorithm and the improved algorithm is carried out, and the performance of the improved algorithm is verified. The fourth section summarizes the main content of the article and puts forward the future prospects.

2. The Principles and Improvement Methods of the APF-RRT*-Connect Algorithm

The APF-RRT*-Connect algorithm adopts the artificial potential field (APF) method to optimize the path generated by the RRT*-Connect algorithm. The APF-RRT*-Connect algorithm can reduce the path length to a certain extent, while it increases the time and optimization steps required for path planning. Therefore, the improved APF-RRT*-Connect algorithm is proposed, which can reduce the path planning time, and also reduce the path length and optimization steps.

2.1. The RRT-Connect algorithm*

The RRT*-Connect algorithm [3] is an algorithm that gradually optimizes the path to approach the optimal solution while quickly finding a feasible solution. The RRT-Connect algorithm [2] significantly improves the search efficiency by extending the bidirectional tree (starting tree and target tree), so that it can quickly find an initial feasible path. The principle is shown in Fig.1. The RRT*-Connect retains this two-way expansion mechanism and ensures the ability to quickly find paths in complex high-dimensional spaces.

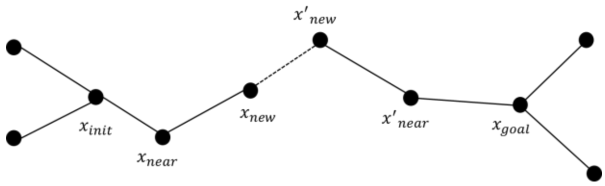


Fig.1. RRT-Connect algorithm

The RRT * -Connect gradually optimizes the path by re-selecting the reconnection of the parent node and the nearby nodes [1] to ensure that the final path can gradually approach the global optimal solution. The principle is shown in Fig.2 and Fig.3.

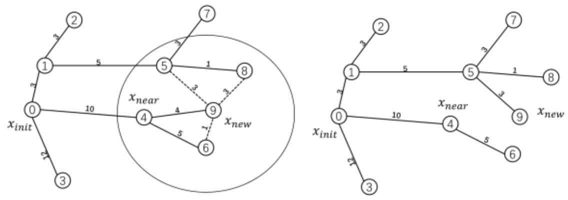


Fig.2. Reselecting parent nodes

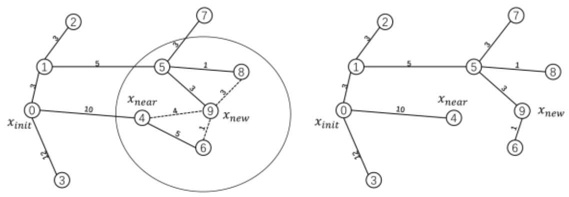


Fig.3. The process of reconnecting nearby nodes

2.2. The APF-RRT*-Connect algorithm

(1) Artificial Potential Field

Artificial Potential Field (APF) [5] is a classical robot path planning algorithm, which guides the robot movement by constructing a virtual ' potential field '. The potential field consists of two parts : attractive potential field and repulsive potential field. The gravitational potential field guides the robot to the target position, while the repulsive potential field keeps the robot away from obstacles, so as to achieve the purpose of obstacle avoidance and target arrival in path planning.

The attractive potential field U_{att} is typically defined as a quadratic function of the Euclidean distance between the target point and the current robot position. The corresponding attractive force F_{att} is the negative gradient of the attractive potential field.

$$\begin{cases} U_{att} = \frac{1}{2} k_{att} |q - q_{goal}|^2 \\ F_{att} = -\nabla U_{att} = - k_{att}(q - q_{goal}) \end{cases} \quad (1)$$

Where k_{att} represents the scaling factor of the attractive potential field, which controls the magnitude of the attractive force. q denotes the current position, and q_{goal} represents the target position.

The repulsive potential field U_{rep} is typically defined as an inverse function of the distance between the robot and obstacles, generating a repelling effect near obstacles. The corresponding repulsive force F_{rep} is the negative gradient of the repulsive potential field.

$$U_{rep} = \begin{cases} U_{rep1}, & if\ |q - q_{obs}| \le d_0 \\ 0, & if\ |q - q_{obs}| > d_0 \end{cases} \quad (2)$$

$$U_{rep1} = \frac{1}{2} k_{rep} \left(\frac{1}{|q - q_{obs}|} - \frac{1}{d_0} \right)^2 \quad (3)$$

$$F_{rep} = \begin{cases} F_{rep1}, & if\ |q - q_{obs}| \le d_0 \\ 0, & if\ |q - q_{obs}| > d_0 \end{cases} \quad (4)$$

$$F_{rep1} = k_{rep} \left(\frac{1}{|q - q_{obs}|} - \frac{1}{d_0} \right) \frac{1}{|q - q_{obs}|^3} (q - q_{obs}) \quad (5)$$

Where k_{rep} represents the scaling factor of the repulsive potential field, q_{obs} denotes the position of the obstacle, and d_0 is the influence range of the repulsive potential field.

The total potential field force acting on a node is the vector sum of the attractive force and the repulsive force.

$$F = F_{att} + F_{rep} \quad (6)$$

(2) The APF-RRT-Connect algorithm*

The process of the traditional APF-RRT*-Connect algorithm [4] is : after the RRT*-Connect algorithm generates a path, the Artificial Potential Field (APF) method is used to optimize the path. When calculating the attractive force and attractive potential field, the first point on the path is set as the current position q, and the next point is set as the target point. Once a point's calculations are complete, the next point is set as the current position q and the process continues iteratively until the second-to-last point in the path.

When calculating the repulsive force and repulsive potential field, the current position q is set based on the first point to the last point on the path, and the obstacle position q_{obs} is determined from surrounding obstacle information. The resultant force at each point is the vector sum of the attractive and repulsive forces. Its schematic diagram is shown in Fig.4.

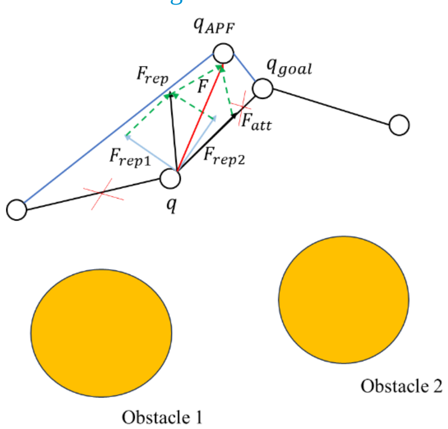


Fig.4.APF-RRT*-Connect Algorithm

(3) The improved APF-RRT-Connect algorithm*

The improved APF-RRT*-Connect algorithm achieves path optimization by incorporating the Artificial Potential Field (APF) method during the path generation process. Based on the bidirectional tree generation process of the RRT*-Connect algorithm, one tree is used as the reference. The newly generated node x_{new} of this tree is set as the current position q, while the newly generated node of the other tree is set as the target position q_{goal}. The attractive force acting on the nodes of the first tree is then calculated. Subsequently, the second tree is used as the reference, and its newly generated node x'_{new} is set as the current position q, while the newly generated node of the other tree is set as the target position q_{goal}, to calculate the attractive force acting on the nodes of the second tree.

When calculating the repulsive force and repulsive potential field, forces on the nodes of both trees need to be calculated separately. First, the nodes of the first tree are treated as the current position q to compute the forces acting on its nodes, followed by treating the nodes of the second tree as the current position q to compute the forces acting on its nodes. Finally, the resultant forces acting on the nodes of the first and second trees are calculated separately. Its schematic diagram is shown in Fig.5.

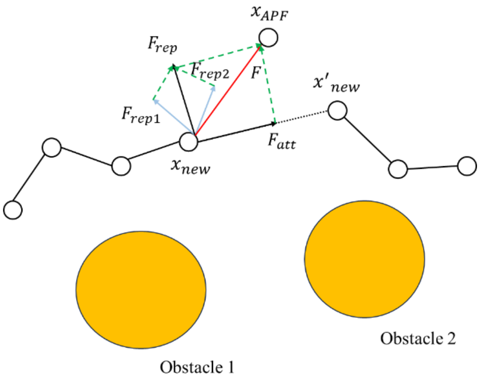


Fig.5.Improved APF-RRT*-Connect Algorithm

3. Simulation

To validate the effectiveness of the improved APF-RRT*-Connect algorithm, a simulation platform was built using MATLAB to compare the performance of the improved algorithm with the traditional algorithm, thereby verifying the efficiency of the improved algorithm. The validation experiments were conducted in the same three-dimensional environment. The map of the 3D environment has a range of 1000×1000×1000, where cyan blocks represent cuboid obstacles, blue blocks represent cylindrical obstacles, and green blocks represent spherical obstacles. The starting point coordinates are (0, 0, 0), and the target point coordinates are (600, 600, 200). The golden line represents the path planned by the improved algorithm, while the red line represents the path planned by the traditional algorithm. The simulation results are shown in Fig.6. From the comparison in the figure, it can be observed that in the 3D environment, compared to the traditional algorithm, the improved algorithm has fewer turning points, a shorter path, and a smaller search range in invalid regions.

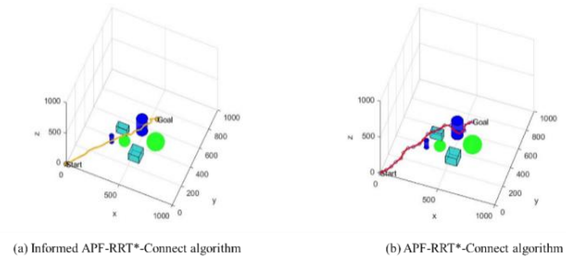


Fig.6. Simulation in 3D environment

After 30 repeated experiments, the path planning time and path length of the two algorithms were summarized, and the results are shown in Fig.7 and Fig.8. Figure 7 displays the comparison of path planning times after 30 repeated experiments for both algorithms, while Figure 8 shows the comparison of path lengths. It is evident that in the 3D environment, the improved algorithm requires less time for path planning and generates shorter paths.

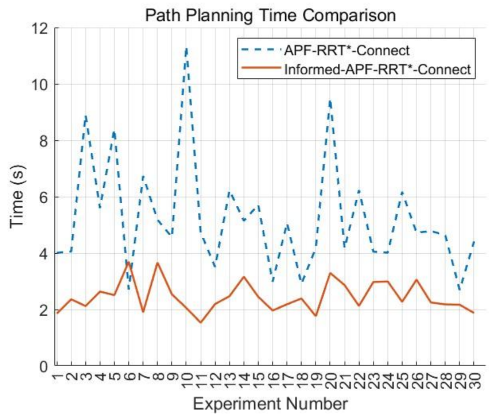


Fig.7. Comparison of Path Planning Time

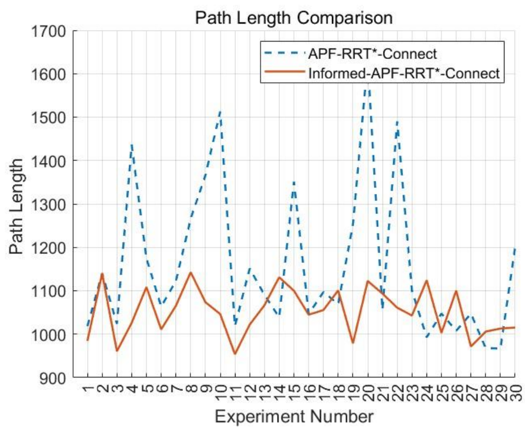


Fig.8. Comparison of Path Length

The average path planning time of the improved algorithm over 30 experiments is 2.44896367 seconds, which is approximately 53.29% shorter than the 5.24300037 seconds of the traditional algorithm. The average path length of the improved algorithm is 1052.172907, which is about 9.16% shorter than the 1158.33075 of the traditional algorithm.

4. Conclusion

This paper introduces the APF-RRT*-Connect algorithm and presents its improvements. By modifying the role of the artificial potential field (APF) in the path generation process, the time required to generate paths is reduced, and the generated paths are shorter. During the path generation process, the latest node of one tree is used as the current node in the artificial potential field, while the latest node of the other tree is used as the target node to calculate the attraction between the two points. This approach reduces the randomness of node generation, thereby decreasing the number of redundant nodes, avoiding blind exploration of meaningless areas, and ultimately reducing path planning time and extra collision detection. Finally, a comparative experiment between the improved APF-RRT*-Connect algorithm and the original APF-RRT*-Connect algorithm was conducted in Matlab. The results show that the improved APF-RRT*-Connect algorithm requires less time to generate paths and produces paths with lower costs.

Although the improved algorithm reduces path planning time, the additional calculations introduced by the artificial potential field may still increase overall computational complexity in high-dimensional and complex environments, especially when obstacles are dense or large in scale. In the future, we plan to reduce computational complexity by improving the artificial potential field model, thereby further enhancing computational efficiency.

References

1. Noreen Iram, Khan Amna, Habib Zulfiqar, A comparison of RRT, RRT* and RRT*-smart path planning algorithm, *International Journal of Computer Science and Network Security*, 2016, 16(10): pp.20-27.
2. Zhao Chaoli, Ma Xing, Mu Chunyang, Research on path planning of robot arm based on RRT-connect algorithm, *Chinese Control and Decision Conference, IEEE*, 2021, pp. 3800-3805.
3. Klemm Sebastian, Oberländer Jan, Hermann Andreas, et al., RRT*-Connect: Faster, asymptotically optimal motion planning, *IEEE international conference on robotics and biomimetics, IEEE*, 2015, pp. 1670-1677.
4. Ma Benshan, Wei Chao, Huang Qing, et al., APF-RRT*: An Efficient Sampling-Based Path Planning Method with the Guidance of Artificial Potential Field, *International Conference on Mechatronics and Robotics Engineering, IEEE*, 2023, pp. 207-213.
5. Li Haoyang, Robotic path planning strategy based on improved artificial potential field, *International Conference on Artificial Intelligence and Computer Engineering, IEEE*, 2020, pp. 67-71.

Authors Introduction

Mr. Yiming Wang

He graduated in 2023 from the School of Automation and Electrical Engineering at Tianjin University of Technology and Education with a bachelor's degree. Currently, he is a master's student at Tianjin University of Science and Technology, focusing on robotic path planning.

Ms. Miao Zhang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2019, she received a doctorate from the University of Windsor, Ontario, Canada. The research direction is intelligent algorithms design filters, the control system design of industrial robots and control theory.

Improved RRT*-Connect Based on MATLAB

Ruofan Zhang
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

Miao Zhang
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China
E-mail: autoZhang0299@163.com, miaozhang@tust.edu.cn

Abstract

The article analyzes an improved RRT*-Connect path planning algorithm based on MATLAB software. First, the sampling domain of the algorithm is changed to an elliptical sampling domain. Second, adaptive compensation expansion is introduced to accelerate the search speed of the path planning algorithm. Finally, the algorithm is combined with a greedy algorithm to optimize the path. Simulation results show that this algorithm significantly improves both the path length and planning time compared to the traditional RRT*-connect algorithm.

Keywords: RRT*-Connect, Path planning, Greedy Algorithm

1. Introduction

Path planning is one of the key issues in robot navigation. The RRT*-Connect [1] (Rapidly-exploring Random Tree Star Connect) algorithm, as an effective random sampling algorithm, has been widely applied in the field of path planning. However, the traditional RRT*-Connect algorithm has problems such as low search efficiency and longer path lengths in certain scenarios, which limits its effectiveness in practical applications.

To address the above issues, this paper proposes an improved RRT*-Connect path planning algorithm. First, the sampling domain of the algorithm is changed to an elliptical sampling domain to better adapt to the distribution of environmental obstacles. The elliptical sampling domain can more effectively explore the space around obstacles, improving the search efficiency of the algorithm. Secondly, an adaptive compensation expansion strategy is introduced, which dynamically adjusts the expansion step size to further accelerate the search speed of the algorithm. Finally, a greedy algorithm [2] is combined with the improved RRT*-Connect algorithm to optimize the generated path, making it smoother and shorter.

Through MATLAB simulation experiments, the superiority of the proposed algorithm in terms of path length and planning time has been verified. Compared with the traditional RRT*-Connect algorithm, the improved algorithm not only generates shorter paths but also significantly reduces planning time. This provides an efficient path-planning method for robot navigation, with broad application prospects in practical applications.

The main contributions of this paper include: 1) proposing an improved RRT*-Connect algorithm based on an elliptical sampling domain, which improves the search efficiency of the algorithm; 2) introducing an adaptive compensation expansion strategy to further accelerate the search speed of the algorithm; 3) combining a greedy algorithm with the improved RRT*-Connect algorithm to optimize the generated path. Simulation results show that the proposed algorithm has significant improvements in both path length and planning time.

The structure of this paper is arranged as follows: Section 2 introduces the specific implementation of the improved RRT*-Connect algorithm; Section 3 presents the setup and results analysis of the simulation experiments; Section 4 summarizes the work of this paper and looks forward to future research directions.

2. Algorithm Principles

2.1. Elliptical sampling domain

The traditional RRT*-Connect algorithm uses a rectangular sampling domain, which may not adapt well to the distribution of environmental obstacles in certain scenarios. Therefore, this paper changes the sampling domain to an elliptical shape to better cover the space around obstacles. The mathematical expression for the elliptical sampling domain is Eq. (1):

$$\frac{(x-x_c)^2}{a^2}+\frac{(y-y_c)^2}{b^2}\leq 1 \quad (1)$$

where (x_c, y_c) are the coordinates of the center of the ellipse, and a and b are the lengths of the semi-major and semi-minor axes, respectively.

2.2. Adaptive compensation expansion

The traditional RRT*-Connect algorithm uses a fixed expansion step size, which may lead to low search efficiency in certain areas. To address this, this paper introduces an adaptive compensation expansion strategy. Specifically, when the tree node is close to the target point, a smaller expansion step size is used; when the tree node is far from the target point, a larger expansion step size is employed. This can accelerate the search speed of the algorithm.

2.3. Greedy algorithm optimization

To further optimize the path generated by the algorithm, this paper combines a greedy algorithm with the improved RRT*-Connect algorithm. During each tree node expansion, in addition to the random sampling point, the nearest sampling point is also selected for expansion. This helps to make the generated path smoother and shorter.

2.4. Algorithm process

To further optimize the path generated by the algorithm, this paper combines a greedy algorithm with the improved RRT*-Connect algorithm. During each tree node expansion, in addition to the random sampling point, the nearest sampling point is also selected for expansion. This helps to make the generated path smoother and shorter.

The specific process of the improved RRT*-Connect algorithm is as follows:

1. Initialization: Set the starting point x_s and the goal point x_g, and establish two trees T_s and T_g, with x_s and x_g as the root nodes, respectively.

2. Sampling: Randomly sample a point x_{rand} within the elliptical sampling domain.

3. Expansion: For each tree $T_i(i = s, g)$, find the nearest node x_{nearest} to x_{rand}, and use x_{nearest} as the starting point to calculate the expansion step size Δx based on the adaptive compensation expansion strategy, resulting in a new node x_{new}. If x_{new} does not intersect with any obstacles, it is added to the tree T_i.

4. Connection: Check if the two trees meet; if they do, use the greedy algorithm to optimize and obtain the final path.

5. Iteration: Repeat steps 2-4 until a feasible path is found or the maximum number of iterations is reached.

6. Combine the generated path with the greedy algorithm to prune and optimize the original path, resulting in a new path.

Through the above steps, the improved RRT*-Connect algorithm can generate shorter and smoother paths, with a significant reduction in planning time.

3. Simulation Analysis

3.1. Experimental setup

To verify the performance of the proposed algorithm, this paper conducts experiments in a MATLAB simulation environment. The simulation scenario is shown in Figure 1, which includes multiple irregularly shaped obstacles. The starting point and the goal point are set at (0, 0) and (1000, 1000), respectively. In the experiment, we compare the improved RRT*-Connect algorithm with the traditional RRT*-Connect algorithm.

3.2. Result analysis

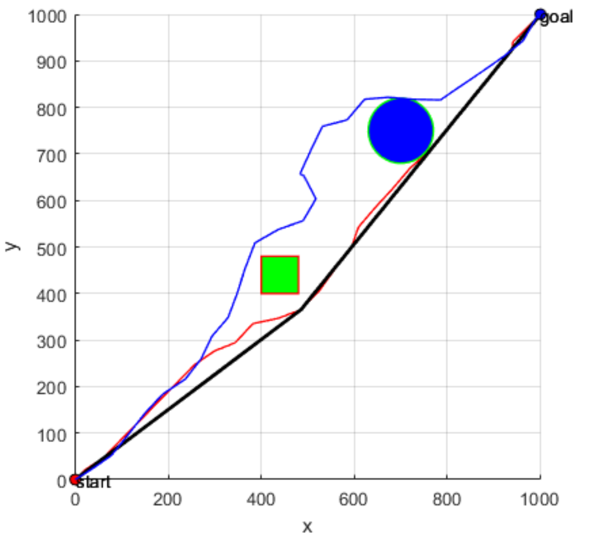


Fig.1. Algorithm simulation in simple environment

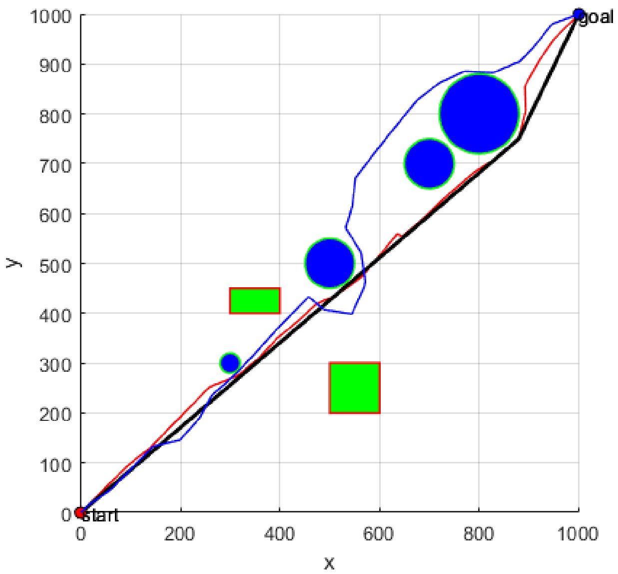


Fig.2. Algorithm simulation in medium environment

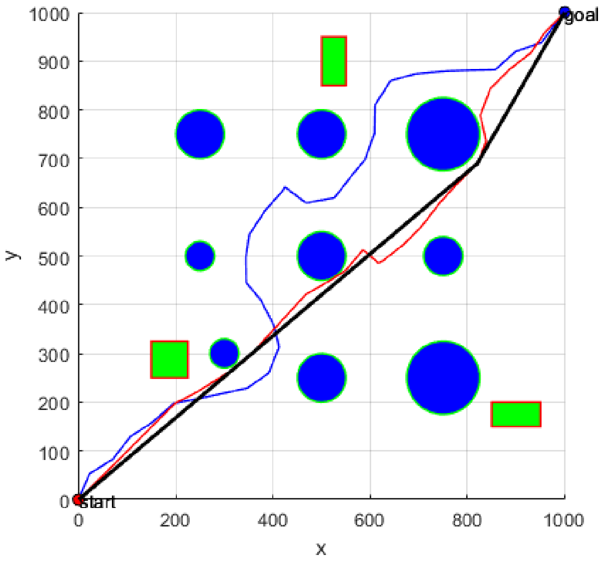


Fig.3. Algorithm simulation in complex environment

Fig.1, Fig.2, and Fig.3 show the paths generated by the traditional RRT*-Connect algorithm and the improved RRT*-Connect algorithm (RRT*-connect algorithm generates a blue path, the Improved RRT*-connect algorithm does not combine with the greedy algorithm generates a red path, and the Improved RRT*-connect algorithm combines with the greedy algorithm generates a black path), respectively. It can be observed that the paths generated by the improved algorithm are smoother and shorter.

Table 1. Algorithm comparison

Environ-ment	Algorithm	Time/s	Path length / pixel
Simple	RRT*-connect	6.04	1573.64
	Improved RRT*-connect	2.70	1424.61
Medium	RRT*-connect	6.67	1612.05
	Improved RRT*-connect	4.72	1434.96
Complex	RRT*-connect	7.25	1721.87
	Improved RRT*-connect	5.12	1498.27

Table 1 lists the comparative results of the two algorithms in terms of path length and planning time. It can be seen that the improved algorithm reduces the average path length by 11.19% and the average planning time by 37.09%, showing significant improvements compared to the traditional RRT*-connect algorithm. This result indicates that the improved algorithm has a clear advantage in optimizing path planning and can more effectively address path planning issues in complex environments.

Further analysis of the advantages of the improved algorithm reveals that it performs particularly well in handling obstacles and dynamic environments. The traditional RRT*-connect algorithm often requires a long time to search for paths when faced with complex obstacles, while the improved algorithm, by introducing a more efficient heuristic search strategy, can quickly find feasible paths, significantly reducing planning time. Additionally, the improved algorithm has also enhanced path smoothness, generating paths that are not only shorter but also more natural, reducing sharp turns and unnecessary path backtracking.

In practical applications, this performance improvement is of great significance in fields such as robot navigation and autonomous driving. Shorter paths and faster planning times mean that robots can respond more quickly to environmental changes, thereby improving the overall efficiency and safety of the system. Therefore, the application prospects of the improved algorithm are broad and worthy of further exploration and promotion in future research.

4. Conclusion

In this study, we compared the performance of the traditional RRT*-connect algorithm with the improved algorithm in terms of path length and planning time. The results indicate that the improved algorithm shows significant enhancements in both key metrics, with an average path length reduction of 11.19% and an average planning time reduction of 37.09%. These results not only validate the effectiveness of the improved algorithm but also demonstrate its potential for application in complex environments. This is primarily attributed to the following improvements:

1. The use of an elliptical sampling domain, which better adapts to the distribution of environmental obstacles and enhances the search efficiency of the algorithm.

2. The introduction of an adaptive compensation expansion strategy that dynamically adjusts the expansion step size, accelerating the search speed of the algorithm.

3. The combination of a greedy algorithm with the improved RRT*-Connect algorithm, which further optimizes the generated paths.

In summary, the improved algorithm exhibits good performance in the field of path planning and has broad application prospects. Future research can further explore the adaptability and optimization potential of this algorithm in different scenarios to promote its implementation in practical applications such as robot navigation and autonomous driving.

Based on the above simulation analysis, the air provided by the air conditioner can cover the whole room, and the air supply by the air conditioner can meet the comfort of indoor personnel.

References

1. Candemir Doğan, İbrahim Kök, Suat Özdemir, Narrowed Regions-based Bidirectional Path Planning Using RRT-Connect for Single Aircraft Missions. *Procedia Computer Science*, 2024, 231: pp.703-708.
2. Chang Jianfang, Na Dong, Donghui Li, et al., Skeleton extraction and greedy-algorithm-based path planning and its application in UAV trajectory tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 2022 58(6): pp.4953-4964.

Authors Introduction

Mr. Ruofan Zhang

In 2024, he received his Bachelor of Engineering degree from the School of Electronic Information and Automation, Tianjin University of Science and Technology, China.

Ms.Miao Zhang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2019, she received a doctorate from the University of Windsor, Ontario, Canada. The research direction is intelligent algorithms design filters, the control system design of industrial robots and control theory.

Rapidly Exploring Gmapping

Congchuang Han
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222， China
Miao Zhang *
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222， China
*Email: jyhcc0915@mail.tust.edu.cn, * miaozhang@tust.edu.cn*

Abstract.
This paper presents an improved algorithm for the proposal distribution. The proposed algorithm enhances the range of filter values, the precision of re-sampling, and the accuracy of the map building. Additionally, the computational overhead is reduced, thereby optimizing the issue of the particle degeneration. The comparative experiments demonstrate that the proposed algorithm enhances the mapping precision and speed.
Keywords: Path planning, Algorithm optimization.

1 Introduction

Currently, the traditional RBPF (Rao-Blackwellised Particle Filters) algorithm uses particle filtering algorithm to estimate the status of the robot[1],[2].Traditional positioning and environmental map acquisition still requires a lot of human resources to accurately measure the surrounding environment. The process is time-consuming and labor-intensive, with poor results and errors. Therefore, the method of simultaneous localization and mapping (SLAM) is a hot topic in the current research of mobile robots. At present, the implementation of 2D laser SLAM is mainly divided into two methods: filter-based and nonlinear optimization-based. The filter-based method includes the extended Kalman filter (EKF), the unscented Kalman filter (UKF), and the particle filter (PF). The SLAM method based on nonlinear optimization mainly uses the framework of graph optimization to implement SLAM. The Gmapping algorithm is an algorithm for simultaneous localization and mapping of mobile robots. It is widely used in mobile robots, unmanned vehicles, drones and other fields, especially in navigation and exploration tasks in unknown environments. Its real-time and effectiveness make it one of the important tools in the field of SLAM. Gmapping is a SLAM algorithm based on 2D LiDAR using the RBPF algorithm to complete the construction of a two-dimensional grid map. It can build indoor maps in real time, and the amount of calculation required to build small scene maps is small and the accuracy is high. Gmapping is mainly based on the particle filter algorithm, which estimates the position and posture of the robot through LiDAR data and robot motion information. The ability to process sensor data in real time and generate a two-dimensional environment map is very important for mobile robots that need to navigate in unknown environments. This method uses multiple "particles" to represent possible states and updates these states based on sensor data. The advantages of Gmapping are real-time and high efficiency. It can build indoor environment maps in real time, with less calculation in small scenes and higher map accuracy.

Although Gmapping can generate maps in real time with high accuracy, it is inevitable that particles will degrade when estimating the robot's state in large scenes because a large number of particles are needed to estimate the robot's posture. Particles with larger weights will account for a larger proportion, and particles with smaller weights will gradually decrease or even disappear. In addition, frequent resampling steps will cause particles to gradually degrade, which will waste a lot of computing resources and affect the mapping effect. In 2007, Giorgio Griisetti and Cyril Stachniss [3] proposed an improved RBPF mapping algorithm to implement SLAM mapping, that is, an improved Rao-Blackwellised particle filter mapping algorithm using an improved proposed distribution and an adaptive resampling method. This improved RaoBlackwellised particle filter algorithm improves the performance of the algorithm, effectively reduces the computational complexity, and alleviates the particle degradation problem by using a small number of particles for state estimation [4],[5],[6]. However, the algorithm still relies on a high number of particles in an environment with large maps and high local similarity, and errors will occur in mapping. The robustness of the algorithm needs to be improved, and how to effectively limit the spatial range of the Proposal distribution to improve sampling efficiency has not been fully explored.

This paper is structured as follows: the second part describes the Gmapping algorithm; the third part describes the idea of the improved algorithm; the fourth part conducts a reasonable data analysis of the results of the optimization algorithm; and the fifth part summarizes the whole paper.

2 DRP-GMapping Core Idea

In order to overcome the problems of excessive computation and particle degradation in SLAM methods based on conventional particle filters, this paper further optimizes the proposal distribution by optimizing the particle formula of the particle swarm algorithm, restricts the proposal to a small valid area, and dynamically adjusts the proposal distribution in combination with the observation information of the latest frame of the robot, and then samples this valid area. This allows the odometer distribution to match the lidar distribution, greatly improves the sampling accuracy, reduces the number of particles collected, and greatly alleviates the particle degradation problem.The fig.1 shows the comparison of particle filters.

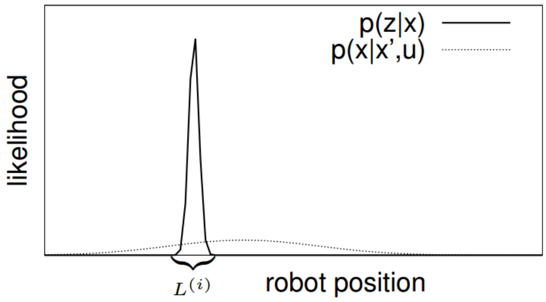


Fig.1 Particle filter comparison

In order to solve the efficiency problem in particle filtering, we proposed a method to improve the efficiency of particle filtering by limiting the effective area of proposal distribution. Specifically, the traditional proposal distribution is generally widely distributed in the entire state space, while we limit the proposal distribution to a narrower effective area.

The specific steps of this method are to dynamically estimate the possible activity range of the robot through the robot's motion model, local map information and sensor data. The range is determined by the robot's current position, movement speed, and obstacle information in the surrounding environment. By restricting particle generation to only this effective region, the distribution range of particles in state space can be significantly reduced. Specifically, at each moment, the current position of the robot is first predicted through the robot motion model (such as the differential drive model), and combined with the status information of the previous moment, an estimate of the current position is obtained. Combined with the observation information of the latest frame, the possible range of activities of the robot is evaluated. For example, by analyzing the location of obstacles in lidar data, the proposal distribution is restricted to areas where obstacles do not exist. Finally, based on the robot's motion estimation and recent observations at the current moment, the size and position of the effective area are dynamically adjusted. The shape of the effective area can be a circle or a rectangle, and the range is adjusted in real time based on the robot's motion status and sensor feedback.

Therefore, we need to keep the number of particles at a relatively small value to improve the pose quality of proposal distribution sampling. The specific changes are:

$$x_t^i \sim p\left(x_t \mid u_t, x_{t-1}^i\right) \rightarrow x_t^i = \arg\max\left\{p\left(z_t \mid x_t, m\right) p\left(x_t \mid u_t, x_{t-1}^t\right)\right. \quad (1)$$

After optimization, the Proposal distribution is more similar to the Gaussian distribution represented by (μ, Σ), so the particle propagation is modified from the kinematic model sampling to the sampling of the Gaussian distribution, and the Gaussian distribution is:

$$u = \frac{1}{n}\sum_{j=1}^{k} x_j p\left(z_t \mid x_j, m\right)$$

$$\Sigma = \frac{1}{n}\sum_{j=1}^{k}\left(x_j - u\right)\left(x_j - u\right)^T p\left(z_t \mid x_j, m\right) \quad (2)$$

The weight is calculated as follows:

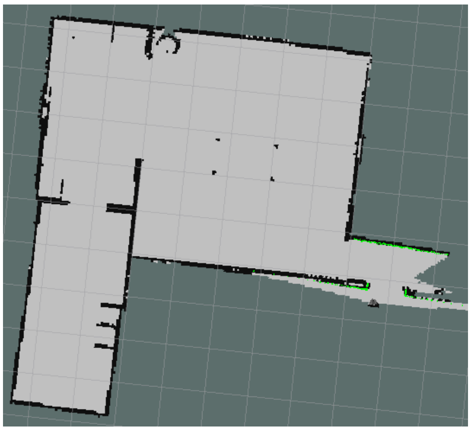
$$w = \eta \frac{p\left(z_t \mid x_t, m\right) p\left(x_t \mid u_t, x_{t-1}^i\right) bel\left(x_{t-1}\right)}{p\left(x_t \mid x_{t-1}, u_t, z_t, m\right) bel\left(x_{t-1}\right)} \quad (3)$$

$$p = \left(x_t \mid x_{t-1}, u_t, z_t, m\right) = \frac{p\left(z_t \mid x_t, m\right) p\left(x_t \mid u_t, x_{t-1}\right)}{p\left(z_t \mid x_{t-1}, u_t, m\right)} \quad (4)$$

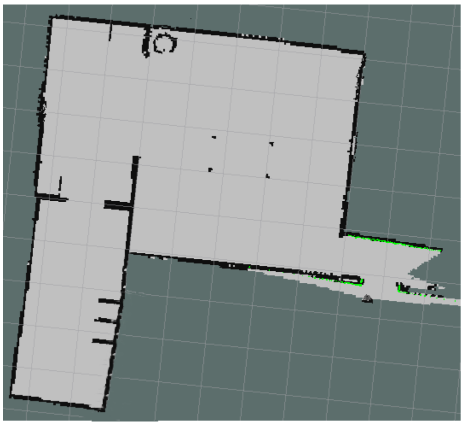
$$\begin{aligned} w &= p\left(z_t \mid x_{t-1}, u_t, m\right) \\ &= \int p\left(z_t \mid x_t, m\right) p\left(x_t \mid x_{t-1}, u_t\right) d\, x_t \\ &= \sum_{j=1}^{j=k} p\left(z_t \mid x_t, m\right) \end{aligned} \quad (5)$$

3 Optimization result analysis

In order to verify the effectiveness of the proposed method, we conducted comparative experiments in various environments. Experimental environments include: Static indoor environment: a standard indoor environment consisting of multiple rooms and corridors. Complex large-scale environments: Large indoor environments with multiple rooms, corridors, and obstacles. In the experiment, we used odometry-based data and compared the performance of the traditional GMapping algorithm and the improved algorithm in terms of positioning accuracy, mapping accuracy, computational efficiency, and robustness. Experimental results show that the optimized GMapping algorithm shows excellent performance in all test scenarios. Compared with traditional methods, the improved algorithm improves positioning accuracy by 15%-25% in static and dynamic environments. In large-scale and complex environments, the maps generated by the optimized algorithm are more accurate, with errors reduced by about 20%. Due to the limitation of particle sampling space, the calculation efficiency is significantly improved, and the processing time is reduced by 10%-15% on average. In a dynamic obstacle environment, the optimized algorithm shows higher robustness, can quickly adapt to environmental changes, and has smaller fluctuations in positioning accuracy.The experimental results are shown in Fig.2.



(a)



(b)

Fig.2.(a) Gmapping algorithm grid map;
(b) improved Gmapping algorithm grid map.

	Simple open environment	Complex obstacle environment
Traditional Gmapping average error (m)	0.23	0.38
Improve the average error of the algorithm (m)	0.18	0.30
Reduction rate (%)	21.7	21.0
Traditional Gmapping is time-consuming (s)	1.00	3.1
Improved algorithm time-consuming(s)	0.85	2.7
Improvement rate (%)	15	12.9

4 Conclusion

This paper proposes a GMapping algorithm optimization method that optimizes the Proposal distribution algorithm in the effective area and combines the observation information of the latest frame. Through this optimization, the efficiency of particle filtering is significantly improved, and the positioning accuracy and mapping accuracy are improved. Experimental results show that this method effectively alleviates the problem of particle degradation, improves the accuracy of resampling, and greatly reduces mapping errors. At the same time, the method in this paper is currently only used in simulation experiments of a single robot. In the future, we will further study the application of this method in actual environments and extend it to collaborative positioning and mapping research on multiple robots.

References

1. Arnaud Doucet, Nando de Freitas, Kevin Murphy, et al. Rao-Blackwellised particle filtering for dynamic Bayesian networks[C]. *Proceedings of the 16th Annual Conference on Uncertainty in Artificial Intelligence*, 2000:176-183.
2. Fernando Martin, Luis Moreno, Dolores Blanco.Kullback-Leibler divergence-based global.
3. Giorgio Grisetti, Cyrill Stachniss, Wolfram Burgard. Improved techniques for grid mapping with *Rao-Blackwellised particle filters . IEEE Transactions on Robotics*, 2007, 23(1):34-46.
4. Kennedy J，Eberhar R C． Particles Swarm Optimization *Proc ofthe IEEE Conference on Neural Networks． Perth，Australia*.1995：1942—1948
5. GerkeyBrian.Gmappingin ROS[EB/OL]. [2010-08-05].http://wiki.ros.org/slam-gmapping.
6. Durrant-Whyte Hugh, Bailey Tim. Simultaneous localization and mapping (SLAM)-Part I: *The Essential Algorithms. IEEE Robotics & Automation Magazine*,2006, 13(2):99-110.

Authors Introduction

Mr. Congchuang Han

He is currently pursuing his undergraduate degree at the school of Electronic Information and Automation, Tianjin University of Science and Technology.

Ms. Miao Zhang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2019, she received a doctorate from the University of Windsor, Ontario, Canada. The research direction is intelligent algorithms design filters, the control system design of industrial robots and control theory.

An Improved Laser SLAM Algorithm Based on Cartographer

Lei Jiang
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

Miao Zhang
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

E-mail: 2622394423@qq.com, miaozhang@tust.edu.cn

Abstract

SLAM is one of the core technologies in the field of robotics. At present, Laser SLAM has become the mainstream mapping solution for general mobile robots. Cartographer algorithm is one of the mainstream Laser SLAM algorithms, which has attracted much attention because of its high accuracy and suitable for large scenes. However, the effect of sensor data fusion using Unscented Kalman Filter (UKF) is not ideal. Therefore, an improved Cartographer algorithm is proposed in this paper, which uses Adaptive Unscented Kalman Filter (AUKF) to fuse information of sensors, aiming to improve the accuracy of localization and mapping.

Keywords: SLAM, Cartographer, UKF, AUKF

1. Introduction

Simultaneous Localization and Mapping (SLAM) was proposed by Smith et al. in 1986 [1]. It uses the sensors carried by the robot to complete localization and mapping in unknown environments, which provides support for subsequent autonomous navigation. SLAM can be divided into Laser SLAM and Visual SLAM according to the types of sensors it relies on. Among them, Laser SLAM has become the mainstream scheme because of its mature and stable algorithm. After years of development, there have been many excellent algorithms in the field of Laser SLAM, such as Gmapping, Hector, Karto. Among many algorithms, Cartographer algorithm has attracted a lot of attention because of its powerful function.

Cartographer is a set of Laser SLAM algorithm based on graph-optimized launched by Google, which supports both 2D and 3D lidar SLAM [2]. It can be used across platforms and supports multiple sensors such as Lidar, IMU, Odometry, GPS and Landmark. It is currently one of the most widely used Lidar SLAM algorithms in practical scenarios.

Cartographer used the Unscented Kalman Filter (UKF) to predict and fuse the information acquired by the sensors. This fusion method cannot deal with the influence of sensor noise, and will cause the decline of mapping accuracy. Therefore, this paper proposes to use Adaptive Unscented Kalman filter (AUKF) for data prediction and fusion to optimize the estimation of pose information [3].

The rest of this article is organized as follows. The second section introduces the Cartographer algorithm. In the third part, the improved method and theoretical basis are presented. In the fourth section, conduct experiments and data analysis. The fifth part summarizes the main content of this paper.

2. Cartographer algorithm

Cartographer is a state-of-the-art graph-optimized SLAM framework that was open-sourced by Google in 2016. It is designed to generate high-quality environmental maps with a high resolution of up to 5 cm, making it particularly suitable for applications requiring precise spatial understanding. The Cartographer algorithm is fundamentally divided into two key components: Local SLAM and Global SLAM, each of which plays a crucial role in constructing and refining the map.

In the Local SLAM, the submap is constructed and updated through a series of lidar scans. Odometry and IMU data were used to calculate the trajectory, and the pose of the car was estimated. The pose estimate is used as the initial value to match the lidar data and update the value of the PoseExtrapolator. In the Global SLAM, the fusion of submap is continuously updated by Ceres, and the accumulated error is eliminated by Loop Closing.

Cartographer has high engineering stability and has both mapping and relocalization. It has made innovation in improving accuracy and optimizing efficiency, and is very valuable in engineering applications.

Overall, Cartographer represents a major advancement in the field of SLAM, offering a comprehensive, reliable, and efficient solution for generating high-resolution maps and localizing in real-time. Its open-source also allows developers to adapt and extend the framework for a wide range of practical use cases, further demonstrating its value and potential in the rapidly evolving field of robotics and autonomous systems.

3. UKF and AUKF

In the Local SLAM (Simultaneous Localization and Mapping) of the Cartographer algorithm, the Unscented Kalman Filter (UKF) is employed to fuse the data from Odometry and the Inertial Measurement Unit (IMU) for precise pose estimation. However, in practical scenarios, sensor noise is unavoidable, leading to errors in the fusion process. The traditional UKF approach, while effective, may not perform optimally when the system's noise characteristics vary over time or under different operating conditions. This paper proposes an Adaptive Unscented Kalman Filter (AUKF) to improve the accuracy and robustness of the data fusion process, particularly in the presence of varying sensor noise.

The Kalman filter, a recursive and optimal algorithm, is widely used for linear state estimation. In a linear system, it provides the best estimate by minimizing the mean squared error. However, in real-world applications, many systems exhibit nonlinear behavior, which makes the traditional Kalman filter less effective. To address this limitation, various extensions of the Kalman filter have been developed, one of the most notable being the Unscented Kalman Filter (UKF).

UKF is a nonlinear Kalman filter algorithm whose main idea is to approximate the propagation and measurement model of a nonlinear function by a set of so-called "unscented transforms". This transformation works by selecting a specific set of sampling points and then calculating the mean and covariance required in the Kalman filter based on these sampling points. Compared with the traditional EKF algorithm, the UKF algorithm has better estimation accuracy and robustness, and performs better in nonlinear systems.

On the other hand, in addition to the nonlinear system model, the noise characteristics of the system may also change with time, which requires an adaptive adjustment of the Kalman filter algorithm. AUKF algorithm is an adaptive unscented Kalman filter algorithm proposed for this demand. By dynamically adjusting the parameters and noise model in the UKF algorithm, the proposed algorithm can adapt to the changes of system noise and provide more accurate and stable state estimation.

The main steps of AUKF are similar to the standard UKF, except that after each update, an adjustment step is performed to dynamically update the process noise covariance Q and the observation noise covariance R by observing the prediction error. The core idea is that if the predicted covariance is larger than the actual one, then the noise covariance needs to be reduced, and otherwise it needs to be increased. The following is a basic formula for dynamically adjusting Q and R.

$$Q_i = Q_i \times max\,(1, \frac{|P_i - P_{i_{prev}}|}{Q_i}) \tag{1}$$

$$R_i = R_i \times max(1, \frac{|y_i - \hat{y}_i|}{R_i}) \tag{2}$$

Q_i and R_i are the elements in Q and R, P_i and $P_{i_{prev}}$ are the state covariance at the current and previous step, y_i and $\hat{y}_i$ are the actual and predicted observations.

By integrating adaptive noise modeling, the AUKF enhances pose estimation reliability in SLAM applications, particularly in real-world environments where sensor noise can be both significant and unpredictable. This paper introduces the AUKF as an improvement over the traditional UKF, with the goal of boosting the accuracy and robustness of data fusion in the Cartographer SLAM algorithm and similar systems that rely on combining multiple sensor modalities, such as IMUs and Odometry.

The AUKF is employed for the data fusion and state estimation of Odometry and IMU data, playing a crucial role in combining the complementary information from these two sensors to improve overall estimation accuracy. The flowchart of position and pose estimation is shown in Fig. 1. To validate the effectiveness of the proposed method, a simulation experiment is conducted in MATLAB, where AUKF is applied to fuse Odometry and IMU data. The simulation fusion experiment of Odometry and IMU data is carried out using AUKF, and the result is shown in Fig.2.

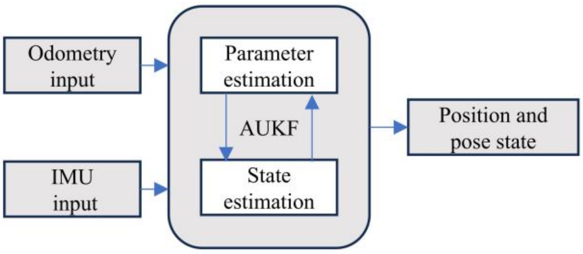


Fig. 1. Flowchart of position and pose estimation

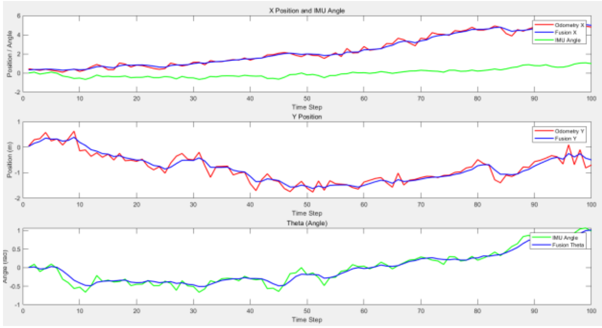


Fig. 2. Data fusion using AUKF

4. Experiment and analysis

The experimental platform used in this paper is the ROS car equipped with lidar, as shown in Fig.3. The lidar model is WHEELTEC N 10, and the system is Ubuntu18.04. N10 is a single-line lidar (Fig. 4), and its parameters and metrics are shown in Table 1.

Fig. 3. Experimental platform

Fig. 4. WHEELTEC N10

Table 1. Parameters and metrics of N10

Parameters	Metrics
Range of scan	360°
Range of measure	25m
Frequency of scan	6-12Hz
Frequency of measure	4500/s

In the indoor environment, the experimental platform was used for mapping test, the effect is shown in Fig.5. The mapping error of the algorithm is shown in Table 2. The improved Cartographer algorithm effectively reduces the absolute error and mean square error of rotational and translational. Therefore, it can be shown that the improvements in this paper can effectively transform the effect of pose estimation.

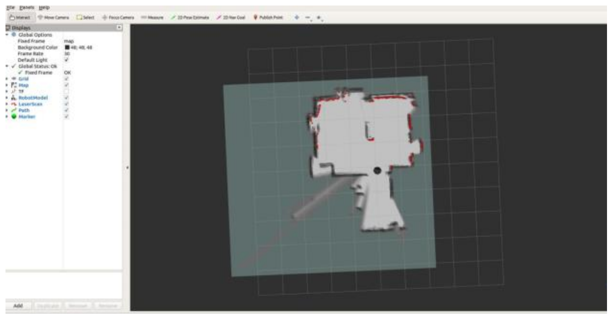


Fig. 5. Effect of mapping

Table 2. Comparison of error

Error term	Cartographer	Improved
Abs translation error/m	0.825±1.210	0.708±0.915
Rms translation error/m^2	1.246±0.587	1.023±0.372
Abs rotational error/deg	1.073±0.846	0.894±0.615
Rms rotational error/deg^2	0.957±0.657	0.785±0.493

The improved algorithm demonstrates superior performance in both the mapping quality and the numerical comparison of error when compared to the original Cartographer algorithm. Numerical comparisons between the original Cartographer algorithm and the improved version show a clear reduction in mapping errors, with the new algorithm yielding more precise and consistent maps. The error metrics, such as positioning error, trajectory deviation, and map consistency, are significantly lower in the improved algorithm, indicating that it not only performs better in real-world conditions but also offers more reliable results in terms of both localization and mapping accuracy.

In summary, the improved algorithm not only reduces the error in mapping but also enhances the overall quality and reliability of the generated maps. This improvement makes the algorithm a more viable solution for real-world applications where accuracy, robustness, and efficiency are critical.

5. Conclusion

To solve the problem of inaccurate data fusion in Cartographer algorithm, an improved Cartographer algorithm is proposed in this paper. The main focus of this improvement is to enhance the fusion of IMU and Odometry data by replacing the traditional Unscented Kalman Filter (UKF) with the Adaptive Unscented Kalman Filter (AUKF). The AUKF algorithm, on the other hand, offers an adaptive mechanism to dynamically adjust the filter's parameters based on the varying noise characteristics of the sensor data. This adaptability allows the AUKF to reduce the impact of noise, resulting in more accurate and reliable pose estimates. The experimental results demonstrate that the improved Cartographer algorithm, which incorporates the AUKF, produces significantly smaller errors in pose estimation compared to the original algorithm using UKF. Moreover, the improvements are particularly noticeable in indoor environments, where sensor noise and errors in Odometry can be more pronounced. The enhanced algorithm exhibits superior accuracy and stability, making it more suitable for real-time applications that require high-precision mapping and localization.

References

1. Randall C. Smith and Peter Cheeseman, On the Representation and Estimation of Spatial Uncertainty, *The International Journal of Robotics Research*, 1986, Vol. 5 No. 4, pp. 56–68.
2. Wolfgang Hess, Damon Kohler, Holger Rapp, et al., Real-time loop closure in 2D LIDAR SLAM, *2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden*, 2016, pp. 1271-1278.
3. S.J. Julier, J.K. Uhlmann, Unscented filtering and nonlinear estimation, *Proceedings of the IEEE, 2004*, Vol. 92, No. 3 pp. 401-422.

Authors Introduction

Mr. Lei Jiang

He received his B.S. degree from School of Information and Intelligence Engineering, Tianjin Ren'ai College, China in 2023. He is currently a Master course student in Tianjin University of Science and Technology. His research area is about robotics.

Ms. Miao Zhang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2019, she received a doctorate from the University of Windsor, Ontario, Canada. The research direction is intelligent algorithms design filters, the control system design of industrial robots and control theory.

Semi-Global Stereo Matching Algorithm Based on Optimized Image Preprocessing

Huaijiao Sha
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

Miao Zhang
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China
E-mail: miaozhang@tust.edu.cn, 2442943837@qq.com

Abstract

This paper proposes an improved Semi-Global Matching algorithm that enhances the quality of disparity images by optimizing image preprocessing. The method applies denoising, contrast enhancement, and Sobel filtering to compute pixel gradients and highlight image edges, thereby improving image quality and reducing noise interference to enhance the clarity of feature boundary. The preprocessed images are used for the SGM algorithm, which improves matching accuracy and adaptability through adaptive sliding windows and dynamic aggregation strategies for cost calculation and aggregation. Experimental results indicate that the improved algorithm enhances the accuracy and robustness of disparity images.

Keywords: Image preprocessing, Denoising, Sobel filtering, Contrast enhancement，Dynamic aggregation

1. Introduction

Stereo matching is one of the core tasks in the field of computer vision, which is widely used in the fields of autonomous driving, robot navigation, augmented reality and 3D scene reconstruction. By estimating the disparity of a pair of corrected stereo images, the stereo matching algorithm can provide key depth information, so as to support the understanding of the scene and the interaction with the physical environment. Among many stereo matching algorithms, Semi-Global Matching (SGM) has attracted much attention because of its good balance between computational efficiency and disparity map quality, and has become a classic and widely used algorithm.

Proposed by Hirschmüller in 2005, SGM approximates global optimization through multi-directional 1D path aggregation, thereby avoiding the computational complexity associated with full global optimization [1]. However, SGM faces challenges in handling weakly textured regions (e.g., sky, walls) and occluded areas, where the lack of feature support can lead to disparity mismatches. Additionally, repetitive texture regions (e.g., lattice structures) may result in ambiguous solutions. SGM is also sensitive to dynamic lighting and noise, which can affect matching accuracy.

To address the aforementioned issues, researchers have proposed optimizations in the cost calculation and aggregation stages in recent years [2]. This paper introduces an adaptive kernel size based on Sobel filtering and a dynamic sliding window strategy. By adjusting the kernel size and weight distribution to adapt to local texture characteristics, the robustness of the matching cost is enhanced. The dynamic weight allocation, combining bilateral filtering and gradient features, further improves the adaptability to complex lighting conditions. In the cost aggregation stage, an adaptive strategy based on dynamic windows is employed to accurately smooth the cost values, enhancing the global consistency of the disparity map.

The rest of this article is organized as follows. The second section presents the image preprocessing optimization; the third section introduces the matching cost calculation and cost aggregation; the fourth section provides simulation examples to verify the effectiveness of the proposed protocol; the fifth section summarizes the main content of this paper.

2. Image Preprocessing Optimization

Image preprocessing is a critical step in the fields of computer vision and image processing [4]. It aims to perform preliminary processing on the original image to improve the effect of subsequent analysis and processing. By eliminating noise, enhancing contrast, edge detection and other operations, image preprocessing lays the foundation for tasks such as feature extraction, image segmentation, and target recognition.

2.1 Bilateral filtering for denoising processing

Bilateral Filtering is a nonlinear filter capable of simultaneously smoothing noise and preserving edges. Unlike traditional Gaussian filtering, bilateral filtering considers not only spatial distance but also pixel value similarity, enabling edge-preserving image smoothing. For the pixel $I(p)$ in the image, the formula for the pixel value $I'(p)$ after bilateral filtering is (1)

$$I'(p) = \frac{1}{W_p}\sum_{q\in s} G_S(\| p - q \|)\cdot G_r(|I(p) - I(q)|)\cdot I(q)\#(1)$$

2.2 Sobel filtering and adaptive kernel size

Sobel filtering is a classic edge detection method that extracts significant edge features by computing the gradients of an image in the horizontal and vertical directions, while also providing a certain level of noise resistance. Its core lies in utilizing the Sobel kernel for convolution operations, enhancing edge information and suppressing noise interference.

Although the Sobel filtering method is simple and efficient, its sensitivity to noise and limited effectiveness in processing complex images can be a drawback. The traditional Sobel filter typically uses a fixed-size kernel (e.g., 3×3), but to improve performance, the kernel size can be dynamically adjusted based on the local texture complexity of the image. In regions with complex textures, smaller kernels are selected to preserve details, while in more uniform regions, larger kernels are chosen to enhance smoothing and reduce noise interference. To achieve this, the gradient magnitude or local standard deviation of the image can be computed to assess the texture complexity of local regions, and the most appropriate kernel size is then dynamically selected based on this evaluation. This dynamic adjustment method based on local features improves the adaptability of Sobel filtering in different image regions, thus enhancing the accuracy and robustness of edge detection.

Combine the gradients in the horizontal and vertical directions to calculate the gradient magnitude (2).

$$\boldsymbol{G(x,\ y) = \sqrt{G_x^2(x,\ y) + G_y^2(x,\ y)}} \qquad (2)$$

Adaptive kernel sizes may cause uneven gradient distributions, which can be normalized through local standardization. In the filtered gradient map, for each pixel, a fixed-size window (e.g., 3×3 or 5×5) is used to compute (3) the mean μ and standard deviation σ of the gradients within the window.

$$\boldsymbol{\mu = \frac{1}{N}\sum_{i=1}^{N} G_i \qquad \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(G_{i-\mu}\right)^2}} \qquad (3)$$

Normalize the gradient values by using the local mean μ and standard deviation σ to standardize the gradient value of each pixel (4).

$$\boldsymbol{G_{norm} = \frac{G - \mu}{\sigma + \epsilon}} \qquad (4)$$

Fig. 1 is the original unprocessed input image, serving as the basis for subsequent operations. Fig, 2 illustrates the gradient magnitude of the image, with enhanced visualization achieved through normalization. Fig. 3 presents the smoothed image obtained by applying filtering to remove noise and suppress unnecessary detail variations. Fig. 4 Fig. 4 depicts the result of overlaying the gradient information onto the original image, using transparency or other visualization techniques to highlight edge features while retaining the content of the original image.

Fig. 1 Original image

Fig. 2 Gradient magnitude

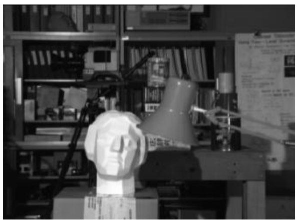


Fig. 3 Filtered image (Original)

Fig. 4 Gradient overly on original

2.3 Histogram equalization

Histogram equalization is a widely used image enhancement technique that significantly improves image contrast by adjusting the distribution of grayscale values. It is particularly effective for scenes with low original contrast. This method redistributes pixel grayscale values to achieve a more uniform distribution, thereby improving global or local contrast, enhancing visual quality, and revealing finer details.

Assuming the grayscale range of the image is [0,L−1], the grayscale equalization is performed through the following steps:

•Calculate the cumulative distribution function (CDF) of the grayscale values:

$$c(r_k) = \sum_{i=0}^{k} h(r_j) \qquad (5)$$

•Map the grayscale value r_k to the equalized grayscale value s_k:

$$s_k = \lfloor (L-1)\cdot c(r_k) \rfloor \qquad (6)$$

The image after histogram equalization in Fig. 6 has obvious contrast enhancement, especially in the area with low contrast of the original image. Fig. 7 is the histogram of the original image, which shows the distribution of the number of pixels with different gray values in the original image. After processing, the distribution of gray levels becomes more uniform (as shown in Fig. 8), the visual effect is more distinct, and the details are more prominent. Fig. 5 is the untreated control original image.

Fig. 5 Original image (Histogram)

Fig. 6 Original histogram equalized image

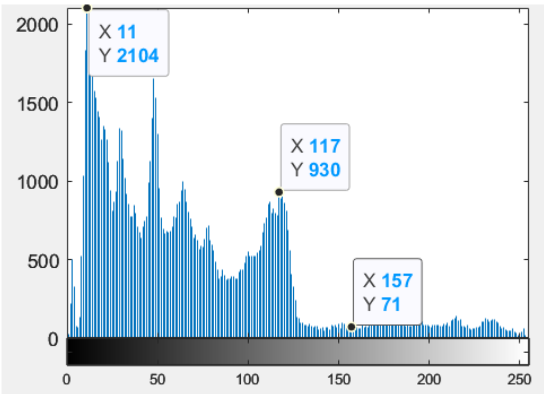


Fig. 7 Histogram of original image

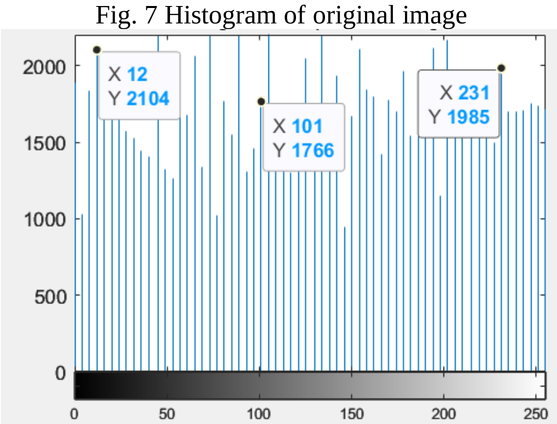


Fig. 8 Histogram of equalized image

2.4 Assigning weights

Based on the characteristics of different image regions, dynamically assign weights to balance the contributions of various preprocessing methods to image quality enhancement. The formula for weight assignment is as follows:

$$\begin{cases} w_{sobel} = f_{\text{gradient}}(\nabla \mathrm{I}) \\ w_{\text{bilateral}} = f_{\text{noise}}(\sigma_n) \\ w_{\text{histogram}} = f_{\text{contrast}}(\mu_{\text{local}}) \end{cases} \quad (7)$$

The weight for Sobel filtering (w_{sobel}) is evaluated based on local gradient variations. Calculate the gradient magnitude $G = \sqrt{(G_x^2 + G_y^2)}$ and increase the weight in regions with prominent edges (high gradient). The weight for bilateral filtering ($w_{bilateral}$) is based on image noise estimation. By analyzing the global noise level σ_n (e.g., mean absolute error or local variance), increase the filtering weight in high-noise regions. The weight for histogram equalization ($w_{histogram}$) is based on contrast distribution. Calculate the local contrast μ_{local} (e.g., the standard deviation of brightness) and assign higher weights to low-contrast regions.

The final fused result I_{final} is obtained by weighted summation of each method:

$$I_{\text{final}} = w_{\text{sobel}} \cdot I_{\text{sobel}} + w_{\text{bilateral}} \cdot I_{\text{bilateral}} + w_{\text{histogram}} \cdot I_{\text{histogram}} \quad (8)$$

3. Matching Cost Calculation and Cost Aggregation

The implementation includes the following steps:

• **Matching Cost Calculation**: Dynamically generate edge information using Sobel filtering with adaptive kernel size, and compute the initial cost matrix using a cost function based on gradient or color differences (e.g., absolute difference or mutual information).

•**Cost Aggregation**: Through a dynamic sliding window mechanism, the cost values within the window are weighted and accumulated, with the weights determined by both the spatial relationships between pixels and their color similarity. By dynamically adjusting the block Size parameter via a slider, the size of the matching window can be modified in real-time, thereby directly influencing the cost computation results. Changes in the window size present a clear trade-off in the disparity map generation process.

• **Path Optimization**: Perform multi-directional path aggregation (e.g., 8 or 16 directions in SGM) to globally smooth the matching cost, ensuring the consistency of the results.

• **Dynamic Adjustment**: During the cost aggregation process, the aggregation weights and window size are dynamically adjusted based on scene complexity, significantly improving the handling of occluded regions and boundary details. The initial cost for each pixel row is calculated using the custom $calcPixelCostBT$ function and stored in $Cbuf$. Subsequently, the cost values are weighted, accumulated, and optimized by integrating the aggregation window configuration and dynamically adjusted parameters (such as sliding window size and weight distribution), enhancing the robustness and accuracy of disparity computation.

4. Experiments and Results

4.1 Dataset and experimental setup

The experiment was conducted on a computer equipped with an Intel Core i7-9750H CPU, using the Middlebury dataset, which is widely used to validate stereo matching algorithms and contains image pairs of varying complexity. The aim of the experiment was to evaluate the effectiveness of the proposed image preprocessing methods (including Sobel filtering, adaptive kernel size, bilateral filtering, and histogram equalization) and the improved matching cost calculation and aggregation strategies in enhancing disparity estimation accuracy and robustness.

4.2 Experimental Results and Analysis

In the field of stereo vision, various algorithms have been proposed to improve the accuracy and robustness of disparity map estimation [3]. The original disparity image shown in Fig. 9 is from Hirschmüller's pioneering work, which introduced the semi-global matching (SGM) method. Fig. 10 shows the disparity map generated by the optimized SGM method.
The experimental results show that the improved image preprocessing method significantly enhances the quality and accuracy of the disparity map Fig. 10. The dynamic weight allocation strategy, combining histogram equalization and gradient features, improves the clarity of feature boundaries and enhances the algorithm's adaptability to lighting variations. By dynamically adjusting the sliding window size, weight distribution strategy, and cost calculation method, the matching accuracy is further improved. Compared to the traditional SGM algorithm, the improved method performs better in complex scenes (such as low-texture areas), producing more accurate and consistent disparity maps.

Fig. 9 Comparison of disparity maps

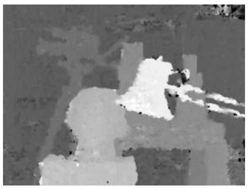


Fig. 10 Disparity map of preprocessed images

5.Conclusion

This study investigates an optimization approach in stereo matching tasks by combining Sobel filtering with adaptive kernel size, bilateral filtering, and histogram equalization. By dynamically adjusting the kernel size, it highlights image edge features and enhances the robustness of matching costs. Meanwhile, a weight allocation strategy integrates the strengths of different preprocessing methods, effectively adapting to weak texture, low contrast, and complex lighting scenarios. Experiments demonstrate that this method significantly improves the accuracy and consistency of disparity estimation. However, the experiment also has certain limitations. Although the improved algorithm shows significant enhancements in accuracy and robustness, some shortcomings remain. Despite the introduction of a dynamic weight allocation method based on local image characteristics, the precision of dynamic adjustment still leaves room for improvement in certain complex scenarios, particularly under extreme lighting conditions and large-scale texture variations.

References

1. Hirschmüller Heiko, Stereo processing by semiglobal matching and mutual information, *IEEE transactions on pattern analysis and machine intelligence*, 2008, 30(2):328-41.
2. Deng Changgen, Liu Deyuan, Zhang Haodong, et al., Semi-Global Stereo Matching Algorithm Based on Multi-Scale Information Fusion, *Applied Sciences*, 2023,13(2):1027-1027.
3. Tippetts, B., Lee, D.J., Lillywhite, K. et al., Efficient stereo vision algorithms for resource-limited systems, *Journal of Real-Time Image Processing*, 2015,10(1):163-174.
4. W. Mao, M. Wang, J. Zhou et al., Semi-Dense Stereo Matching Using Dual CNNs, *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019, pp. 1588-1597.

Authors Introduction

Mr. Huaijiao Sha

He received his Bachelor's degree in New Energy Science and Engineering from Hebei North University in 2023. Currently, he is a master's student at Tianjin University of Science and Technology, and his research area is machine vision.

Ms. Miao Zhang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2019, she received a doctorate from the University of Windsor, Ontario, Canada. The research direction is intelligent algorithms design filters, the control system design of industrial robots and control theory.

Simulation of a 3-DOF Robotic Arm Pick and Place Task Based on Inverse Kinematics

Songyang Mei

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

Miao Zhang

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

E-mail: 18186314885@163.com, miaozhang@tust.edu.cn

Abstract

This paper proposes a simulation method for grasping and placing a 3-DOF robotic arm based on inverse kinematics. Through the MATLAB GUI, the user enters target coordinates, computes the joint angles by the inverse kinematics geometric approach and simulates the motion path to achieve the task operation. In this approach, the robotic arm can reach specified positions precisely, which reduces complexity and error. The position error is calculated through a comparison of the target and the actual position. The simulation results show that the robot arm achieves a small average error across all five experimental groups. The maximum error remains within a reasonable range, which confirms the accuracy of the method in grasping and placing tasks.

Keywords: Robotic arm, MATLAB GUI, Pick and place, Inverse kinematics

1. Introduction

As a critical component in the field of automation, robotic arms are widely utilized in industrial manufacturing, healthcare services, and education. With continuous technological advancements, the intelligence and autonomous operation of robotic arms have become prominent research directions. Precise kinematic analysis, particularly the application of inverse kinematics, is essential for accomplishing complex tasks such as object grasping and placement. Inverse kinematics is a mathematical process that determines the necessary joint angles or positions required for a robotic arm to reach a specific end-effector position in workspace [1]. By calculating the joint parameters, the robotic arm can position the end-effector accurately at the desired location. This applies to tasks such as object manipulation, assembly, or interaction with the environment. Inverse kinematics is a commonly used algorithm in robotic arm control, which transforms target positions into joint angle movements to achieve accurate control. Various methods are commonly used to solve inverse kinematics problems, such as the Denavit-Hartenberg (D-H) parameter method [2], screw theory, iterative techniques [3], and soft computing approaches. Fundamentally, the aim of these methods is to find the optimal joint angles to reach the desired target position. In existing research, many scholars have explored the application of inverse kinematics in robotic arms. For example, studies have demonstrated that inverse kinematics methods can effectively control the position and orientation of robotic arms during object-picking tasks [4].

The structure of this paper is organized as follows. The first section discusses the importance and necessity of inverse kinematics in addressing robotic arm pick and place tasks. The second section describes the MATLAB GUI design process, the forward kinematics model of a three-degree-of-freedom (3-DOF) robotic arm, the methods used to solve inverse kinematics, and the workflow for pick and place operations. The third section evaluates the performance of the robotic arm in pick and place tasks by analyzing the errors between the target and actual positions. The fourth section summarizes the conclusions drawn from the simulation experiments and confirms the effectiveness of the proposed methods.

2. Research Methodology

The method employed in this paper primarily utilizes mathematical geometry to solve the inverse kinematics problem, enabling the execution of pick and place tasks with the 3-DOF robotic arm.

2.1. MATLAB GUI design

The designed system utilizes a MATLAB graphical user interface (GUI) constructed with components such as "Ui figure" and " Ui label " to create an input and display interface. Users can input the X, Y, and Z coordinates for the target pick and place positions. The interface includes buttons to trigger inverse kinematics calculations, as well as output fields or labels to display the calculated joint angles. The designed interface is shown in Fig. 1. By clicking the "Pick and Place" button, the system reads the user-input coordinates and outputs the results in the MATLAB console. The "Reset" button clears all input coordinate information. Upon entering the target position coordinates and pressing the corresponding button, the program calculates the rotational angles of each joint using the inverse kinematics based on the robotic arm's kinematic model. Once the joint angles are determined, the system controls the robotic arm to move to the specified position, completing the pick and place task. Additionally, the actual position coordinates of the end-effector are displayed in the output field.

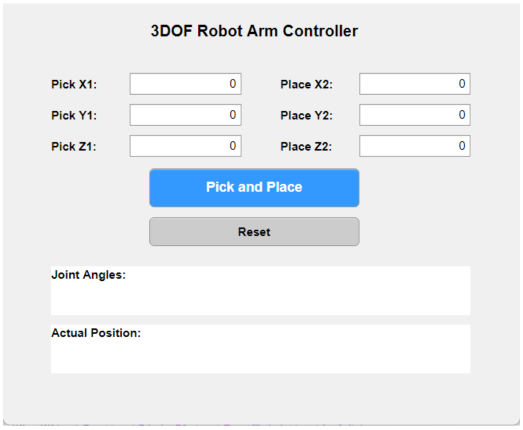


Fig. 1 MATLAB GUI design interface

2.2. Forward kinematic modeling of the 3-DOF robotic arm

The 3-DOF robotic arm is defined based on its link structure, where all links are described using the standard D-H parameter method. The D-H parameters for the link coordinate systems are summarized in Table 1Table 1. In the table, a_i refers to the link length, which represents the displacement along the X-axis of the current link coordinate system. α_i indicates the link twist, describing the rotation angle of the X-axis from the previous link to the current link about the Z-axis. d_i denotes the link offset, representing the displacement along the Z-axis of the previous link coordinate system. θ_i represents the joint angle, which corresponds to the rotation of the current joint.

Table 1. D-H parameters

	a_i	α_i	d_i	θ_i
1	0	$-\frac{\pi}{2}$	10	θ_1
2	15	0	0	θ_2
3	20	0	0	θ_3

The 3-DOF robotic arm model is shown in Fig. 2. Based on the link coordinate systems, the transformation relationship between adjacent links can be established, with equation (1) representing the transformation matrix.

$$ {}^{i-1}_{\ \ \ i}T = \begin{bmatrix} cos\theta_i & -sin\theta_i cos\alpha_i & sin\theta_i sin\alpha_i & a_i cos\theta_i \\ sin\theta_i & cos\theta_i cos\alpha_i & -cos\theta_i sin\alpha_i & a_i sin\theta_i \\ 0 & sin\alpha_i & cos\alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1) $$

From the above equations, the forward kinematics equation can be expressed as equation (2).

$$ {}^0_3T = {}^0_1T\,{}^1_2T\,{}^2_3T \quad (2) $$

2.3. Inverse kinematics solution method

Given the target coordinates (x, y, z) of the end-effector, the inverse kinematics procedure is as follows.

First, the first joint angle θ_1 is calculated. This angle represents the rotation of the base joint in the horizontal plane, indicating how the base of the robotic arm should rotate to point toward the target (x, y). Specifically, θ_1 can be determined using the inverse tangent function of the x and y coordinates of the target position.

$$ \theta_1 = atan2(y, x) \quad (3) $$

Next, the effective horizontal distance r and effective height z_{eff} are computed. The effective horizontal distance r represents the distance of the target position in the plane and can be found by taking the square root of the sum of the squares of x and y. The effective height z_{eff} is the target height minus the length of the base link L_1, which gives the relative height of the target with respect to the robotic base.

$$ r = \sqrt{x^2 + y^2}; Z_{eff} = z - L_1 \quad (4) $$

Then, the second joint angle θ_2, which represents the shoulder angle, is calculated using the law of cosines. The

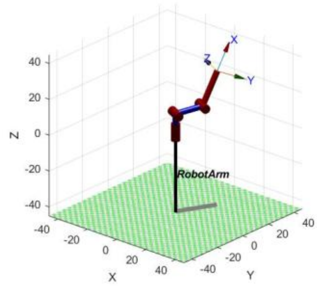


Fig. 2 Robotic arm model

known horizontal distance r, effective height z_{eff}, and the lengths of the two links L_2 and L_3 are used to determine this joint angle.

$$cos(\theta_2) = \frac{r^2 + z_{eff}^2 + L_2^2 - L_3^2}{2L_2\sqrt{r^2 + z_{eff}^2}} \qquad (5)$$

Finally, the third joint angle θ_3, representing the elbow angle, is calculated. The previously computed horizontal distance, effective height, and the lengths of the arm links are used to derive this angle through geometric relationships.

$$\theta_3 = cos^{-1}\left(\frac{r^2 + z_{eff}^2 - L_2^2 - L_3^2}{2L_2L_3}\right) \qquad (6)$$

Through these steps, all the joint angles of the robotic arm θ_1, θ_2, and θ_3 can be determined, enabling the robotic arm to reach the desired end-effector position.

2.4. Overall workflow of robotic arm pick and place operation

The workflow of the robotic arm system for completing pick and place tasks is illustrated in Fig. 3. The system starts with the user inputting the target pick and place coordinates via the MATLAB GUI. Once the coordinates are entered, the system performs inverse kinematics calculations to determine the joint angles required for the operation. Subsequently, the robotic arm executes the corresponding movements based on the calculated joint angles to complete the pick and place task. After the operation, the system displays the computed joint angles and the actual position of the robotic arm. At this point, the system prompts the user to decide whether additional target coordinates need to be entered. If the user provides new coordinates, the process repeats. If no additional coordinates are entered, the operation concludes.

Table 2. Pick joint angles

	θ_1	θ_2	θ_3
1	59.98	39.47	-106.70
2	41.96	51.21	-108.94
3	43.71	90.22	-135.01
4	-138.72	91.30	42.88
5	76.75	-0.51	-83.21

Table 3. Place joint angles

	θ_1	θ_2	θ_3
1	35.88	53.12	-88.74
2	63.72	-7.13	-58.60
3	-1.73	70.00	-66.57
4	60.20	50.21	-98.73
5	-82.27	178.25	-100.14

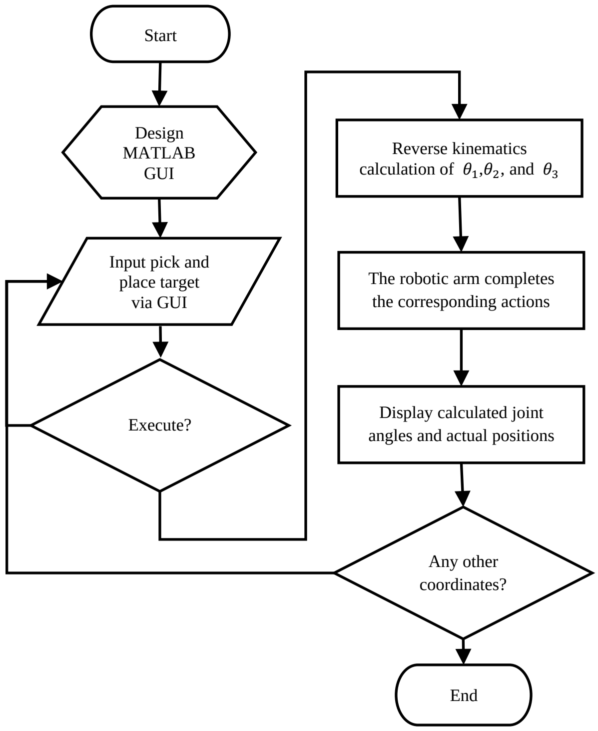


Fig. 3 Overall workflow diagram

3. Simulation Results

The simulation process involves the calculation of joint angles through inverse kinematics, interpolation of these angles, and the visualization of the robotic arm's movements during object grasping and placement. The 3-DOF robotic arm has the following link lengths, L_1 = 10 cm, L_2 = 15 cm, and L_3 = 20 cm. The test coordinates in the pick condition are (10,15,20), (15,10,18), (10,10,10), (12,8, -20), (5,15,30), the test coordinates in the place condition are (20,15,10), (10,20,30), (25,0, -5), (10,22,12), (0,10, -10). Table 2 and Table 3 present the joint angles for the corresponding target positions, which are calculated using inverse kinematics.

The comparison between the actual positions reached by the robotic arm and the target coordinates is illustrated in Fig. 4 and Fig. 5. The test is done by dividing the workspace into quadrant 1 for the negative end-effector X position and quadrant 2 for the positive end-effector X position. The simulation results indicate that the robotic arm end-effector successfully reached the specified target positions. The error is calculated using equation (7).

$$error = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2} \qquad (7)$$

Where $(\Delta x, \Delta y, \Delta z)$ are the differences between the target coordinates of (x, y, z) and the actual coordinates.

The robotic arm managed to reach the coordinates based on the given target, despite the errors. The errors of the second and third sets are larger compared to the other three sets. This may be due to the spatial distribution of the target positions, the geometric structure of the robotic arm, and the accuracy limitations in the inverse kinematics calculation process. At some target positions, geometric constraints affect robotic arm movement. This increases joint angle errors and leads to higher positional errors for the end-effector. In the five test groups, the average positional error is 1.83 cm, and the maximum error is 4.41 cm.

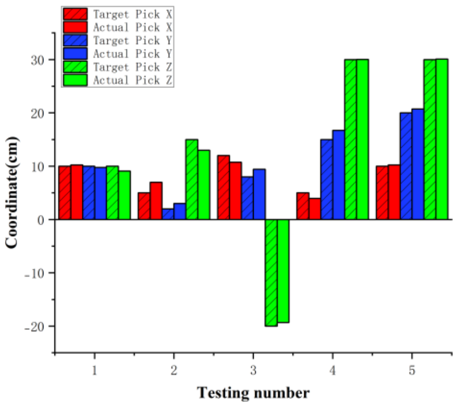


Fig. 4 Pick coordinate result

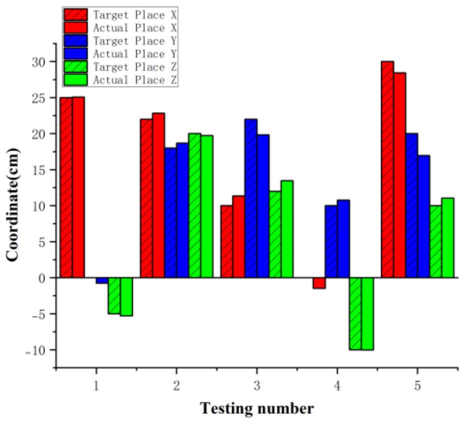


Fig. 5 Place coordinate result

4. Conclusion

This paper employs inverse kinematics to achieve grasping and placing tasks using a 3-DOF robotic arm. The inverse kinematics algorithm calculates the joint angles needed to position and orient the end-effector accurately in the workspace, ensuring precise task execution. Simulation results demonstrate the effectiveness of the proposed method. Across five experimental trials, the robotic arm achieved an average positioning error that remained minimal, with the maximum error maintained within an acceptable range. These findings validate the precision and reliability of the inverse kinematics approach in grasping and placing tasks, highlighting their potential in practical robot applications.

References

1. Al Tahtawi, Adnan Rafi, Muhammad Agni, et al., Small-scale robot arm design with pick and place mission based on inverse kinematics. *Journal of Robotics and Control (JRC)* 2.6 (2021): pp.469-475.
2. Becerra, Yeyson, Mario Arbulu, et al., A comparison among the Denavit-Hartenberg, the screw theory, and the iterative methods to solve inverse kinematics for assistant robot arm. *Advances in Swarm Intelligence: 10th International Conference,* ICSI 2019, Chiang Mai, Thailand, July 26–30, 2019, Proceedings, Part I 10. Springer International Publishing, 2019.
3. Patil, Apurva, Maithilee Kulkarni, et al., Analysis of the inverse kinematics for 5 DOF robot arm using DH parameters. *2017 IEEE International Conference on Real-time Computing and Robotics (RCAR).* IEEE, 2017.
4. Kaur, Manpreet, Venkata Karteek Yanumula, et al., Trajectory planning and inverse kinematics solution of Kuka robot using COA along with pick and place application. *Intelligent Service Robotics* 17.2 (2024): pp.289-302.

Authors Introduction

Mr. Songyang Mei

He received his bachelor's degree in 2023 from the School of Electrical and Electronic Engineering at Wuhan Polytechnic University. Currently, he is a master's student at Tianjin University of Science and Technology, with a research focus on robotic kinematics.

Ms. Miao Zhang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2019, she received a doctorate from the University of Windsor, Ontario, Canada. The research direction is intelligent algorithms design filters, the control system design of industrial robots and control theory.

A Diamond Model Approach to Analyzing GhostSec's Intrusion Paths

Cheng-Ying He
M.S. Degree Program on Cyber-Security Intelligence, National Cheng Kung University
No.1, University Rd., East Dist., Tainan City 701401, Taiwan

Nai-Yu Chen
M.S. Degree Program on Cyber-Security Intelligence, National Cheng Kung University
No.1, University Rd., East Dist., Tainan City 701401, Taiwan

Jung-Shian Li
Department of Electrical Engineering / Institute of Computer and Communication Engineering,
National Cheng Kung University
No.1, University Rd., East Dist., Tainan City 701401, Taiwan

I-Hsien Liu
Department of Electrical Engineering, National Cheng Kung University
No.1, University Rd., East Dist., Tainan City 701401, Taiwan

E- mail: cyhe@cans.ee.ncku.edu.tw, nychen@ cans.ee.ncku.edu.tw, jsli@cans.ee.ncku.edu.tw,
*ihliu@cans.ee.ncku.edu.tw**

Abstract

The convergence of Operational Technology (OT) and Information Technology (IT) has heightened risks for critical infrastructure (CI) and industrial control systems (ICS), leading to a surge in diverse and sophisticated OT attacks with severe consequences. Thus, this study combines the Diamond Model with the Cyber Kill Chain to analyze potential attack paths and methods in the GhostSec case, where attackers compromised a Berghof PLC to demonstrate their access capabilities. Understanding these attack paths offers valuable insights into adversary strategies, aiding in the development of defense measures to prevent similar attacks.

Keywords: Industrial Control Systems, Diamond Model, ICS Cyber Kill Chain, Cyber threat intelligence

1. Introduction

The convergence of Operational Technology (OT) and Information Technology (IT) has significantly reshaped the security landscape of critical infrastructure (CI) and industrial control systems (ICS). While traditionally isolated ICS systems were not designed with robust security measures, increasing internet connectivity has exposed them to vulnerabilities [1] [2]. Tools like Shodan further exacerbate these risks by identifying internet-facing ICS devices, creating opportunities for remote exploitation [3].

The increasing frequency and sophistication of OT-targeted attacks underscore the urgent need for tailored cybersecurity measures [4]. For example, during the 2021 Oldsmar water treatment attack [5], attackers attempted to alter the water's chemical levels to hazardous concentrations. Similarly, in the 2023 Pennsylvania booster station attack, the attacker targeted Israeli-made Unitronics V570 PLCs. Following the incident, Cybersecurity and Infrastructure Security Agency (CISA) issued an advisory highlighting vulnerabilities that allowed attackers to exploit default administrative passwords and gain control [6]. These incidents underline the evolving capabilities of adversaries and the critical need for proactive defense strategies. This study leverages the Diamond Model [7], enhanced with the Activity Thread Graph, to analyze adversary behaviors in the GhostSec case. Here, attackers compromised a Berghof Programmable Logic Controller (PLC) to demonstrate OT vulnerabilities [8]. By mapping attack paths and tactics, this research emphasizes adversary Tactics, Techniques, and Procedures (TTPs), offering actionable insights through Cyber Threat Intelligence (CTI), enabling the development of proactive defense strategies to enhance the resilience of critical infrastructure. The primary contributions of this work are:

- Application of the Diamond Model and Activity Thread Graph to systematically analyze adversary behaviors and attack sequences in OT environments;
- A detailed analysis of the GhostSec case, focusing on adversary strategies and attack vectors;
- Practical recommendations for enhancing cybersecurity in critical infrastructure environments.

2. Methodology

2.1. Diamond Model

The Diamond Model [7], shown in Fig. 1, is a cyber threat analysis framework with four core elements: Adversary, Capability, Infrastructure, and Victim. These elements form a diamond-shaped structure emphasizing their interdependencies. Additionally, the model includes optional meta-features for deeper analysis. In this study, the Diamond Model is used to pinpoint adversary objectives, identify tactics, track behaviors, and uncover defensive opportunities.

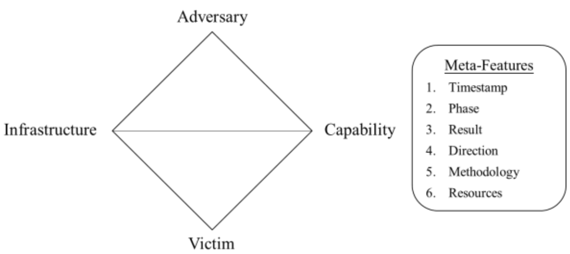


Fig. 1. Diamond Model

2.2. Cyber Kill Chain

The Cyber Kill Chain [9], shown in Fig. 2 and developed by Lockheed Martin, outlines seven stages of a cyber attack: Reconnaissance, Weaponization, Delivery, Exploitation, Installation, Command and Control (C2), and Actions on Objectives. Each stage represents a critical step in the attacker's workflow. This study applies the Cyber Kill Chain to classify actions, map Diamond Model events, and identify intervention points, enabling proactive defense measures.

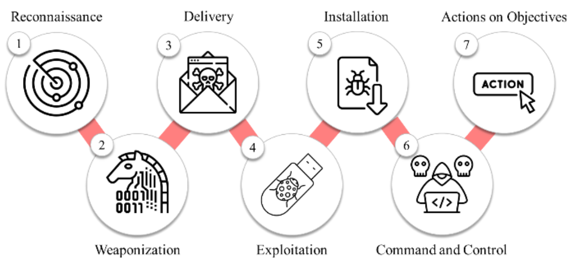


Fig. 2. Cyber Kill Chain.

2.3. Activity Thread Graph

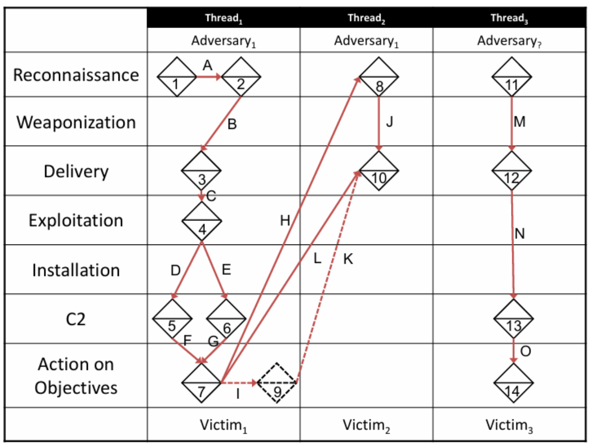


Fig. 3. An example visualization of Activity Thread Graph.

The Activity Thread Graph, detailed in Section 8 of [7] and shown in Fig. 3, provides a timeline-based visualization of attack sequences by connecting events into a coherent flow. In Fig. 4, diamonds denote events, while arcs indicate causal relationships. This study utilizes the Activity Thread Graph to reconstruct and analyze attack paths, offering a clearer understanding of the sequence and structure of adversary activities.

1	The solid line diamond defines the actual event.
2	The dotted line diamond defines the hypothesis event.
A	The solid line arc defines the causal relationship between the event is the actual attack paths.
B	The dotted line arc defines the causal relationship between the event is the hypothesis possible attack paths.
1	The number inside the diamond defines the ID of the event.
A	The number below the arc defines the ID of the causal relationship between the event.

Fig. 4. Definition of the Activity Thread Graph.

3. Case Analysis

3.1. Case description

On September 4, 2022, the hacker group GhostSec claimed responsibility for hacking 55 Berghof Programmable Logic Controllers (PLCs) made in Israel, showcasing their ability to control PLC management interfaces. The group released videos and screenshots of logged-in PLC interfaces as evidence of unauthorized access [8]. Motivated by anti-Israel sentiment, GhostSec targeted PLCs manufactured in Israel. They exploited internet-exposed PLCs, leveraging weak passwords to gain access and possibly undisclosed vulnerabilities for further control. The attack highlighted the critical risks of poor cybersecurity hygiene in industrial systems and the need for stronger access controls and regular security assessments. Such incidents expose vulnerabilities in critical infrastructure, raising concerns about broader implications for national security.

3.2. Activity Thread Graph of Diamond model

As shown in Fig. 5, the Activity Thread Graph depicts two distinct threads representing the GhostSec attack on Berghof PLCs. The events are mapped in Table 1, with their corresponding causal relationships illustrated in Table 2, and the adversary-victim pair of each thread detailed in Fig. 6. In Thread 1, Events 01–05 describe how the attacker leveraged OSINT, Shodan scanning, and weak password exploitation to gain access to the target PLC. Events 06–08 highlight the installation of persistence mechanisms and the establishment of a

Command & Control (C2) channel. Finally, Events 09–11 reveal the attack's objectives, including data exfiltration, public disclosure, and complete operational control of the PLC. Thread 2 focuses on lateral movement and further exploitation. Events 12–13 detail reconnaissance efforts to identify other devices in the network, while Events 14–15 illustrate the attacker gaining administrative privileges through default credentials. Events 16–17 conclude with the disruption of industrial processes and potential data manipulation. This analysis demonstrates the multi-staged, coordinated attack strategy and emphasizes the need for proactive defenses to secure critical infrastructure systems against similar threats.

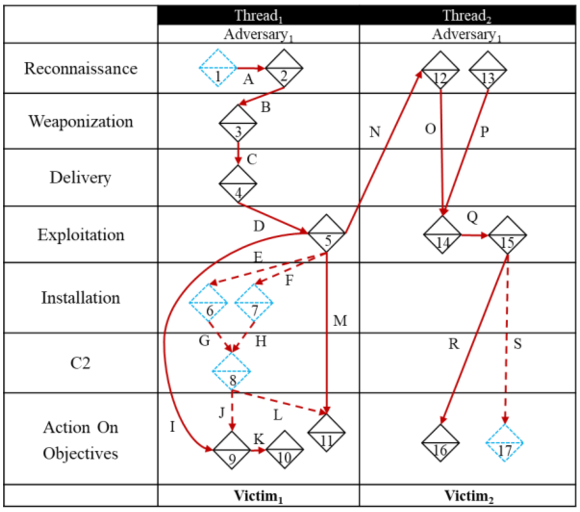


Fig. 5. Diamond model activity thread graph.

Table 1. Activity thread event descriptions for Figure 5.

Event	Hypothesis/Actual	Description
01	Hypothesis	Gather public information on the target using OSINT.
02	Actual	Use Shodan to scan internet-exposed Berghof PLCs.
03	Actual	Create scripts to guess default or weak passwords for target PLCs.
04	Actual	Execute scripts to attempt login on the target PLC.
05	Actual	Successfully log in to the PLC interface, accessing system status, settings, and logs.
06	Hypothesis	Install a backdoor for persistent access to the PLC.
07	Hypothesis	Modify PLC configurations to maintain access.
08	Hypothesis	Establish a Command and Control (C2) channel for continuous device control.
09	Actual	Export HMI screenshots and sensitive system data from the PLC.
10	Actual	Publicly share exported screenshots and backup files on social media.
11	Actual	Halt operations, download system files, reset settings, and delete applications via the PLC interface.
12	Actual	Use exported data to identify IP addresses of other devices.
13	Actual	Scan network for additional devices and gather IP information.
14	Actual	Attempt unauthorized logins using default passwords.
15	Actual	Gain admin access to other PLCs using default credentials.
16	Actual	Control PLCs via the interface, disrupting industrial processes or causing faults.
17	Hypothesis	Modify HMI data display via Modbus commands.

Table 2. Activity thread arc descriptions for Figure5

Arc	Confidence	And / Or	Hypothesis / Actual
A	High	And	Actual
B	High	And	Actual
C	High	And	Actual
D	High	And	Actual
E	Low	Or	Hypothesis
F	Low	Or	Hypothesis
G	Low	And	Hypothesis
H	Low	And	Hypothesis
I	High	Or	Actual
J	Low	Or	Hypothesis
K	High	And	Actual
L	Low	Or	Hypothesis
M	High	Or	Actual
N	Medium	And	Actual
O	High	And	Actual
P	High	And	Actual
Q	High	And	Actual
R	High	Or	Actual
S	Low	Or	Hypothesis

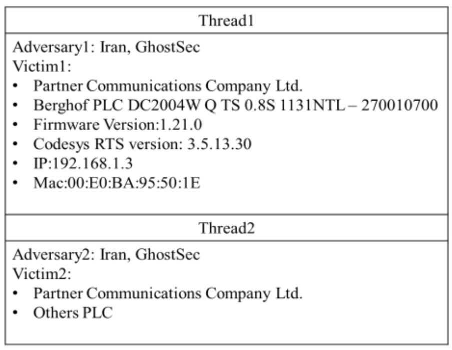


Fig. 6. Adversary-Victim pair of each thread

4. Conclusion

This study highlights the vulnerabilities of industrial control systems exposed to cyber threats, demonstrated by the GhostSec attack on Berghof PLCs. Applying the Diamond Model and Activity Thread Graph provided valuable insights into adversary tactics and attack sequences, emphasizing the need for proactive defense strategies. To enhance the cybersecurity of critical infrastructure, this research suggests the following recommendations:

- Strengthening Access Controls: Enforcing multi-factor authentication, updating default credentials, and applying strong password policies.
- Network Security Management: Disconnect PLCs from the open internet. If remote access is needed, enforce VPNs, firewalls, and secure remote protocols to limit exposure and lateral movement.
- Continuous Monitoring and Threat Intelligence: Deploying advanced threat detection systems and leveraging Cyber Threat Intelligence (CTI) to identify and respond to emerging threats in real time.

Acknowledgment

This work was supported by the National Science and Technology Council (NSTC) in Taiwan under contract numbers 113-2634-F-006-001-MBK, and by the Water Resources Agency (WRA) under the Ministry of Economic Affairs (MOEA) in Taiwan under contract number MOEAWRA1130243.

References

1. Yassine Mekdad, Giuseppe Bernieri, Mauro Conti, Abdeslam El Fergougui, "A threat model method for ICS malware: the TRISIS case," the 18th ACM International Conference on Computing Frontiers, Virtual, Italy, 11-13 May, 2021.
2. M. Cook, A. Marnerides, C. Johnson and D. Pezaros, "A Survey on Industrial Control System Digital Forensics: Challenges, Advances and Future Directions," IEEE Communications Surveys & Tutorials, vol. 25, no. 3, pp. 1705-1747, 2023.
3. Shodan, "Unitronics Search Query," [Online]. Available: https://www.shodan.io/search?query=unitronics.
4. R. Davis, O. F. Keskin, "Cyber Threat Modeling for Water and Wastewater Systems: Contextualizing STRIDE and DREAD with the Current Cyber Threat Landscape," in 2024 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, USA, 3 May, 2024.
5. I-H. Liu, J.-S. Chen, K.-M. Su, J.-S. Li, "Dam Control System's Cybersecurity Testbed", IIH-MSP 2022, Kitakyushu, Japan, Dec. 16-18, 2022.
6. CISA, "IRGC-Affiliated Cyber Actors Exploit PLCs in Multiple Sectors, Including U.S. Water and Wastewater Systems Facilities," 2023. [Online]. Available: https://www.cisa.gov/sites/default/files/2023-12/aa23-335a-irgc-affiliated-cyber-actors-exploit-plcs-in-multiple-sectors-1.pdf
7. S. Caltagirone, A. Pendergast, C. Betz, "The Diamond Model of Intrusion Analysis," Defense Technical Information Center, 2013.
8. Otorio, "Pro-Palestinian Hacking Group Compromises Berghof PLCs in Israel," Jul 2024. [Online]. Available: https://www.otorio.com/blog/pro-palestinian-hacking-group-compromises-berghof-plcs-in-israel/.
9. Hutchins, Eric M.; Cloppert, Michael J.; Amin, Rohan M., "Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains," Lockheed Martin Corporation, 2011.

Authors Introduction

Ms. Cheng-Ying He

He is a postgraduate of M.S. Degree Program on Cyber-Security Intelligence, National Cheng Kung University in Taiwan. He received his B.B.A. degree from the Bachelor of Information Management, National Taiwan University of Science and Technology, Taiwan in 2022. His interests include ICS Security, Cyber Threat Analysis and Intrusion Detection.

Ms. Nai-Yu Chen

She is a postgraduate of M.S. Degree Program on Cyber-Security Intelligence, National Cheng Kung University in Taiwan. She received her B.B.A. degree from the Bachelor of BioBusiness Management, National Chiayi University, Taiwan in 2021. Her interests include ICS Security and Network-Based Intrusion Detection.

Dr. I-Hsien Liu

He is an assistant professor in Department of Electrical Engineering, National Cheng Kung University, Taiwan. He obtained his Ph.D. in 2015 in Computer and Communication Engineering from the National Cheng Kung University. He teaches cybersecurity courses and his interests include Cyber-Security, OT Security, and Wired & Wireless Communication. He is the deputy director of Taiwan Information Security Center @ National Cheng Kung University (TWISC@NCKU).

Dr. Jung-Shian Li

He is a full Professor in the Department of Electrical Engineering, National Cheng Kung University, Taiwan. He graduated from the National Taiwan University, Taiwan, with B.S. in 1990 and M.S. degrees in 1992 in Electrical Engineering. He obtained his PhD in 1999 in Computer Science from the Technical University of Berlin, Germany. He teaches communication courses and his research interests include cybersecurity, cloud computing and network management. He is currently involved in funded research projects dealing with cybersecurity and critical infrastructure protection. He is the director of Taiwan Information Security Center @ National Cheng Kung University.

Bandwidth-Aware Routing Mechanism to Control Hadoop Shuffle Traffic over Software-Defined Networking

Ming-Syuan Wu
Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 807, Taiwan; Email: i111154101@nkust.edu.tw
Cheng-Han Lin
Department of Health-Business Administration, Fooyin University, Kaohsiung 831, Taiwan
Wen-Shyang Hwang*
Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 807, Taiwan; Email: wshwang@nkust.edu.tw
Ce-Kuen Shieh
Department of Electrical Engineering, National Cheng Kung University, Tainan 701, Taiwan
Mao-Syun Lin
Institute of Computer and Communication Engineering, National Cheng Kung University, Tainan 701, Taiwan

Abstract

MapReduce is the main program in the Hadoop computing architecture. It involves mapping and reduction processes that require servers to exchange large amounts of data with each other. Data exchange between servers during the shuffle stage can lead to insufficient network bandwidth. In the present study, an algorithm for distributing Hadoop shuffle traffic to all possible paths was developed. Software-defined networking (SDN), a centrally controlled form of network architecture, was used to collect network status parameters to allocate shuffle traffic. During traffic distribution, the Mininet simulator was used to build the network topology and simulate operation. The Ryu controller was used as the SDN controller. The simulation results indicated that the proposed method was superior to the Spanning Tree Protocol (STP) and bandwidth-aware algorithm in reducing Hadoop completion time through effective shuffle traffic distribution.

Keywords: Software-Defined Network, Apache Hadoop, MapReduce, Shuffle

1. Introduction

Data continue to increase in volume and complexity, and Apache Hadoop [1][2] represents an efficient solution to this problem. Hadoop is well known for its MapReduce process, which comprises mapping and reduction components. MapReduce divides data into small blocks, which are then processed individually. Between the mapping and reduction stages, a large volume of intermediate data is produced. The exchange of intermediate data between servers is referred to as the shuffle stage, during which data must be transmitted over a network connection. If servers are on different racks or further from one another, network strength can substantially affect the overall computing performance of the Hadoop framework [3].

To address these problems, the present study investigated the use of software-defined networking (SDN) to improve link utilization. SDN is a centrally controlled network architecture, and it was selected in this study to achieve a better understanding of network usage. Hadoop traffic was allocated according to network on the basis of traffic distribution, which enabled the simultaneous transfer of intermediate data during the shuffle phase through two paths. Thus, this method maximized network bandwidth use. The main contributions of the study are as follows:

1. The study proposes a multipath method for distributing Hadoop traffic.
2. The proposed method exhibited good experimental performance.

The remainder of this paper is organized as follows. In Section 2, we introduce contextual information regarding the Hadoop and SDN architecture, including a review of relevant literature. Section 3 presents our methodology, and Section 4 provides the evaluation and results of the experiment. Section 5 comprises the conclusion and presents recommendations for future research.

2. Background and Literature Review

2.1. SDN

To achieve central control, SDN architecture separates networks into control and data planes, with the control plane including an SDN controller. The controller communicates with switches in the data plane by sending messages according to the OpenFlow protocol [4]. The switches follow these instructions to forward or drop data.

2.2. Apache Hadoop

Apache Hadoop is a framework for distributed storage and big data computing that is typically used on computer clusters. The two main Hadoop components include the storage system and Hadoop distributed file system (HDFS). Additional components include the processing system and the MapReduce programming model.

HDFS uses distributed technology to store big data. Under this system, input data are partitioned into blocks and spread between servers. MapReduce [5] is a programming model designed to distribute data in parallel across mapping and reduction stages. The intermediate shuffle stage can cause network congestion, which is a notable problem.

2.3. Bandwidth-Aware Methods

A prior study [6] proposed a method to identify the weights on all possible paths from the source to destination node on the basis of the workload per node and the available bandwidth. This approach focuses on the routing algorithm, with the transmission path assigned flow by flow on the basis of network status. The selection of a suitable path with adequate bandwidth improved network utilization relative to that of traditional routing algorithms, such as the Spanning Tree Protocol (STP). However, although bandwidth-aware methods can avoid network congestion, they account for only a fraction of network usage.

3. Methodology

3.1. Architecture Overview

In the present study, we examined a multipath distribution architecture for network traffic during the shuffle stage of the Hadoop MapReduce procedure. The proposed method can dynamically allocate transmission paths according to current network status by separating network architecture into control and data planes. The two modules of this multipath distribution method are displayed in Fig. 1. The first module is the switch monitor, which periodically queries every switch for bandwidth usage. After the switch monitor collects data on bandwidth usage, the controller sends these data to be processed by the multipath allocation module. The multipath allocation module then processes these data to identify the best weight for all possible paths from source to destination.

3.2. Bandwidth-Aware Routing Mechanism

The control plane adopted in the present study was based on OpenFlow 1.3 and the Ryu controller [7]. The proposed architecture included the following five steps, as displayed in Fig. 2:

1. User assigns a Hadoop job.
2. The switch monitor periodically queries the network status.
3. Eq. (1) is used to calculate the available bandwidth of each possible path.
4. Eq. (2) is used to calculate the weights of the possible paths.
5. The controller asks the switches to install the new flow entries.

$$Port\_abw = \max_{\text{bandwidth}} - \frac{(rx+tx)_{now} - (rx+tx)_{pre}}{(during\_time)_{now} - (during\_time)_{pre}} \quad (1)$$

$$Weight_i = \frac{abw_i}{\sum_1^i abw_i} \quad (2)$$

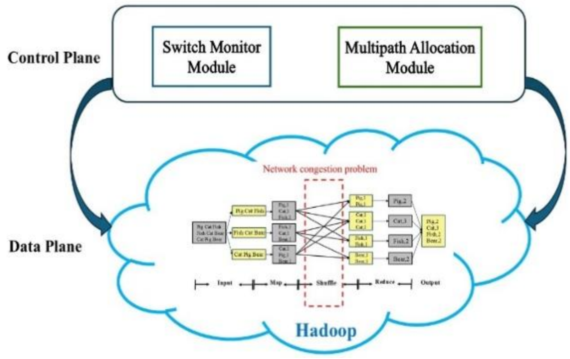


Fig. 1 Architecture Overview

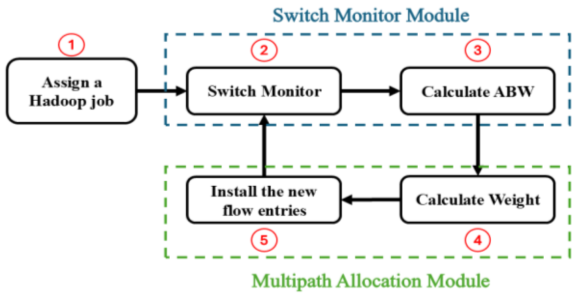


Fig. 2 Distribution Process

At the beginning of the distribution process, the user assigns a Hadoop job on the server. When the network loading changes, the switch monitor begins querying the network status of each switch. After status information, such as the number of transmitted and received packets, is received, the available bandwidth of each possible path is calculated according to Eq. (1). Here, possible paths refer to all available routing paths from the source to the destination node. Eq. (1) subtracts bandwidth consumption from the total bandwidth of the path over the period of the query time interval. The weight of each possible path can be calculated by inputting the result of Eq. (1) into Eq. (2). After calculation, the SDN controller sends messages asking the switch to install a new flow entry, which the switch will follow until the next cycle of the distribution process.

Algorithm 1 presents the process of traffic distribution according to the packet choice of the transmission path. This algorithm is initiated when the switch receives a packet that it does not know where to send from another device. The switch then encapsulates and sends this packet to the SDN controller.

The SDN controller verifies the relationship between the source and destination. If the source and destination are in the same pod, the controller obtains the layer of the switch by checking its identification. The controller asks aggregation layer switches to send the packet back to the edge layer switch. Source and destination servers either do or do not connect to the same edge layer switch. In the first case, our traffic distribution method can be applied to transfer the packet to a remote location. In the second case, packets are sent to the relevant port. Three layers of switches are used for transferring remote packets. The core layer switch is used to communicate among pods, connected by one path each. As the controller receives remote packets from the aggregation switches, the packets may be forwarded up (out of the pod) or down (into the pod). The traffic distribution method can be used for the up direction.

Algorithm 1: Bandwidth-Aware Routing Algorithm

```
When packet_in event occurs in controller:
if Src and Dst are pod-local:
   if switch ∈ aggregate:
      return: Send back to pod
   if switch ∈ edge:
      if Src and Dst are not rack-local:
         multipath allocation
      else:
         return: Send to particular port
else:
if switch ∈ core:
      return: Send to particular pod
   if switch ∈ aggregate:
      if Dst & switch are in the same pod:
         return: Send to particular port
      else:
         multipath allocation
   if switch ∈ edge:
      if Dst is connected to the switch:
         return: Send to the particular port
      else:
         multipath allocation
```

4. Evaluation and Experimental Results

4.1. Environment

In the experiment, 12 Docker containers were used as servers to run the Hadoop WordCount application through Open vSwitch. The bandwidth was set to 20 Mbps between the core layer and aggregation switches and to 10 Mbps between the aggregation and edge layer switches. The environmental configuration is presented in Table 1. The experiment included general and federated Hadoop architectures. The general Hadoop architecture involves only one cluster, which is controlled by a single Name Node. The federated Hadoop architecture involves multiple clusters, each of which is controlled by its own Name Node. The federated Hadoop architecture is used to process iterative jobs, resulting in a high volume of MapReduce processes. We used this architecture to verify whether our traffic distribution method could handle a high network load. The topology of the federated Hadoop architecture is showed as Fig. 3.

The general Hadoop architecture was used for WordCount jobs with 250 Mbytes of input data. Additionally, we compared the STP method with the bandwidth-aware routing mechanism that was built into the topology of our traffic distribution architecture. Iperf was used to generate background traffic at 10 Mbps to congest the path. Before beginning the experiment, we verified whether the flow and port status matched the input statistics of our traffic distribution method.

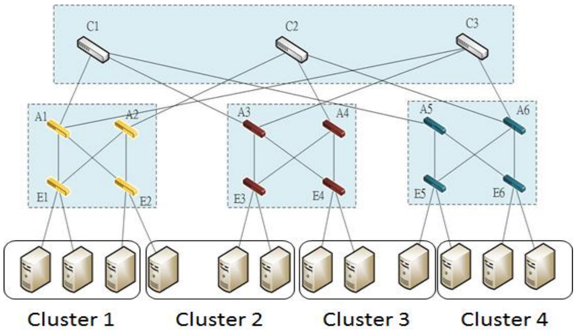


Fig. 3 Federated Hadoop Topology

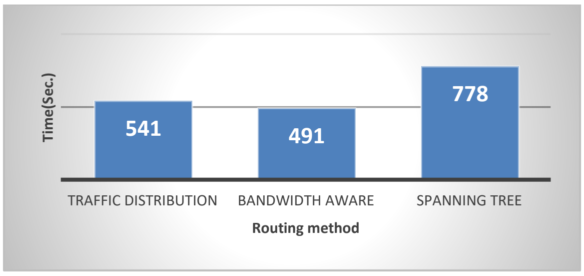


Fig. 4 Data Transfer Time Between Networks

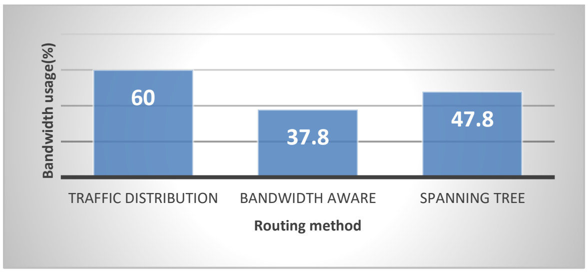


Fig. 5 Bandwidth Usage

Table 1. Environmental Configuration

Node	12
Input size	250Mbytes
Bandwidth	Core←→Aggregate:20Mbps Aggregate←→Edge:10Mbps
Name node	1
Data node	12

4.2. Experimental Results

Background traffic for the simulation was generated using Iperf, which set this traffic to be transmitted through a single path. As indicated in Fig. 4, 250 Mbytes of data were transferred between networks through different routing methods. Background traffic lasting for 300 seconds was transmitted from server 1 to server 4. Then, the 250 Mbytes of data were transmitted from server 1 to server 3. As displayed in Fig. 4, the bandwidth-aware and traffic distribution methods had nearly equal traffic shares due to dynamic path changes. Fig. 5 indicates that the traffic distribution method maximized the use of available bandwidth, resulting in bandwidth usage twice that of the STP method.

5. Conclusion

In the present study, a bandwidth-aware routing mechanism was used to distribute Hadoop shuffle traffic. The SDN architecture calculated the weight of each possible path from source to destination and then redirected traffic accordingly. The results of our experimental evaluation demonstrated that although the network status was congested for some paths, the proposed mechanism achieved a lower Hadoop completion time than did the STP method and bandwidth-aware routing algorithm due to maximal use of available bandwidth. In future studies, we will focus on allocation algorithms and the application of iterative jobs using our allocation method.

Acknowledgments

This work was supported by the National Science and Technology Council (NSTC), Taiwan, under Contract No. NSTC 113-2221-E-242-001 and NSTC 113-2221-E-992-001.

References

1. N. Ahmed, Andre L. C. Barczak, Teo Susnjak and Mohammed A. Rashid, "A comprehensive performance analysis of Apache Hadoop and Apache Spark for large scale data sets using HiBench," Journal of Big Data, vol 7, 110, 2020.
2. R. Machova, J. Komarkova and M. Lnenicka, "Processing of big educational data in the cloud using Apache Hadoop," in: 2016 International Conference on Information Society (i-Society), 2016.
3. N. M. Ahmad, A. H. Yaacob and A. H. M. Amin, "Performance analysis of MapReduce on OpenStack-based hadoop virtual cluster" in: 2014 IEEE International Symposium on Telecommunication Technologies (ISTT), 2014.
4. Jose Miguel-Alonso, "A Research Review of OpenFlow for Datacenter Networking," IEEE Access, vol. 11, pp. 770-786, 2022.
5. J. Dean S. Ghemawat, "MapReduce: Simplified data processing on large clusters", in: Communications of the ACM, vol. 51, pp. 107-113, 2008.
6. W. Shi, Y. Wang, J. P. Corriveau, B. Niu, W. L. Croft and M. Peng, "Smart Shuffling in MapReduce: A Solution to Balance Network Traffic and Workloads," in: 2015 IEEE International Conference on Utility and Cloud Computing (UCC), 2015.
7. Anil Ram, Manash Pratim Dutta and Swarnendu Kumar Chakraborty, "A Flow-Based Performance Evaluation on RYU SDN Controller," Journal of The Institution of Engineers (India): Series B, vol 105, pp203-215, 2024.

Author Introductions

Ming-Syuan Wu

Ming-Syuan Wu received his M.S. degree (2021) from the Department of Electrical Engineering, National Kaohsiung University of Science and Technology, where he is currently a doctoral student.

Cheng-Han Lin

Cheng-Han Lin received a Ph.D. degree from the Department of Electrical Engineering, National Cheng Kung University, in 2010. He is currently an associate professor in the Department of Health-Business Administration at Fooyin University. His current research interests include wireless and SDN networks.

Wen-Shyang Hwang

Wen-Shyang Hwang received B.S. (1984), M.S. (1990), and Ph.D. (1996) degrees from the Electrical Engineering Department of National Cheng Kung University, Taiwan. He is currently a professor in the Electrical Engineering Department at National Kaohsiung University of Science and Technology. His current research interests include multimedia wireless communication, wireless mesh networks, and Internet QoS.

Ce-Kuen Shieh

Ce-Kuen Shieh is currently a professor in the Department of Electrical Engineering at National Cheng Kung University, Tainan, Taiwan, where he also obtained his B.S., M.S., and Ph.D. degrees. His current research interests include distributed and parallel processing systems and computer networking.

Mao-Syun Lin

Mao-Syun Lin received his M.S. degree (2017) from the Institute of Computer and Communication Engineering at National Cheng Kung University. His current research interests include wireless networks and multimedia communication.

An AI Sentencing System: Homicide Case Study in Taiwan

Shih-Chin Lin[1], Cheng-Tsung Yeh, Chen-Yu Lai, Chih-Yun Chang, Chi-Ju Fu
Department of Financial Law, Ming Chuan University,
No 250, Zhong Shan N. Rd., Sec. 5 Taipei 111, Taiwan

E-mail: lsc929@mail.mcu.edu.tw

Abstract

Sentencing is a demonstration of the state's penal authority; however, sentencing decisions by judges in Taiwan frequently face public criticism. With the implementation of the Citizen Judges Act in 2023, Taiwan's Judicial Yuan has proactively developed an AI-based sentencing information system to reduce discrepancies among professional judges' verdicts and to assist citizen judges in determining sentences. This study examines 198 homicide verdicts from district courts between 2016 and 2023, focusing on sentencing factors based on Article 57 of Taiwan's Criminal Code. It analyzes the most frequently considered factors by judges and the variations in sentencing severity that result from these considerations. In conclusion, this paper addresses potential challenges that may arise in the practical implementation of the sentencing information system and considers how judges might use it as a supplementary tool in sentencing, or whether it could fully guide sentencing decisions.

Keywords: Homicide case study, AI Sentencing System, sentencing factors, Homicide

1. Introduction

According to Article 271 of the Criminal Code of the Republic of China: "A person who takes the life of another shall be sentenced to death or life imprisonment or imprisonment for not less than ten years. An attempt to commit an offense specified in the preceding paragraph is punishable. A person who prepares to commit an offense specified in paragraph 1 shall be sentenced to imprisonment for not more than two years." Sentencing represents the exercise of a nation's penal authority, involving both the selection of punishment types and the determination of their severity. The outcome of sentencing dictates the extent to which the state intervenes in a defendant's personal freedom, thus requiring judges to act with utmost caution. However, sentencing decisions by judges in Taiwan are often subject to public criticism, with the media even using the term "dinosaur judges" to describe certain judges. This reflects a significant gap in perceptions between the judiciary and the public regarding sentencing. On January 1, 2023, Taiwan officially implemented the Citizen Judges Act, introducing a system where citizen judges, together with professional judges, participate in trials to determine both the guilt of the defendant and the severity of the sentence. The law aims to enhance judicial transparency, reflect the public's sense of justice, and foster greater trust in the judiciary.
However, sentencing is already a challenging task for professional judges, let alone for citizen judges without legal expertise. Additionally, each citizen judge brings unique life experiences and values, which could lead to differing standards in sentencing. Although Article 57 of the Criminal Code of Taiwan provides factors to be considered in sentencing, these factors are abstract in nature. How citizen judges can appropriately apply these abstract legal standards in individual cases to determine the severity of a sentence poses a significant challenge.
Considering this, Taiwan is developing tools to assist both citizen and professional judges during trials. These tools aim to ensure appropriateness and fairness in sentencing, while also minimizing disagreements between citizen and professional judges regarding the proper sentence. [1]

In response to the implementation of the Citizen Judges Act and to provide citizen judges with more reference materials to overcome the challenges they may face due to their lack of legal expertise in sentencing, the Judicial Yuan began exploring the use of technology to assist in sentencing as early as September 2020.
As a first step, the Judicial Yuan enhanced its longstanding sentencing information system. Subsequently, on February 6, 2023, AI technology was integrated into the sentencing process. Leveraging natural language processing (NLP) techniques, the AI system is being trained to learn sentencing patterns, with the goal of developing a sentencing information system tailored specifically to Taiwan's legal framework. [2] The AI-based sentencing information system is divided into two models: the fact-based model and the evaluation-based model.

1.1 Fact-based Model

[1] Corresponding author's E-mail: lsc929@mail.mcu.edu.tw

This model uses factual content as sentencing factors. It tags relevant facts within the sentencing system, and when users input aggravating or mitigating statutes along with specific sentencing factors, the system automatically retrieves all relevant past rulings. It then provides a sentencing range based on prior cases that match the query criteria, allowing judges and citizen judges to reference similar sentencing outcomes. This assists both professional and citizen judges in determining the appropriate sentence for the case at hand.

Currently, the fact-based model focuses on the following types of crimes: Unsafe driving, Fraud, Theft, Assault and Hit-and-run incidents

1.2 Evaluation-based Model

This model is built upon the sentencing factors outlined in Article 57 of the Criminal Code. Users tag textual descriptions relevant to these sentencing factors and categorize them under the specific factors listed in Article 57. They then evaluate whether each factor is favorable, unfavorable, or neutral toward the defendant.

The evaluation-based model currently focuses on the following crime types: Firearms-related offenses, Drug offenses and Crimes against sexual autonomy.

These two models aim to streamline sentencing by providing data-driven support, improving consistency and fairness in judicial decisions, and helping both professional and citizen judges make well-informed sentencing choices.[3]

This paper focuses on the use of AI-based sentencing for defendants charged with homicide, specifically employing the fact-based model of the AI sentencing information system. To train the system effectively, we must first tag relevant factual content corresponding to sentencing factors within the original text of court rulings. These annotated judgments are then used to train the machine learning model for sentencing.

The sentencing factors are stipulated in Article 57 of the Criminal Code of the Republic of China (Taiwan) and consist of the following ten criteria: a: the motive and purpose of the offense; b: the stimulation perceived at the moment of committing the offense; c: the means used for the commission of the offense; d: the offender’s living condition; e: the disposition of the offender; f: the education and intelligence of the offender; g: relationship between the offender and the victim; h: the seriousness of the offender’s obligation violation; i: the danger or damage caused by the offense; j: the offender’s attitude after committing the offense.

This study seeks to explore how these ten statutory sentencing factors influence sentencing decisions in homicide cases. Accordingly, we collected 198 homicide judgments from local courts, spanning the years 2016 to 2023. These cases involve both completed and attempted homicides, with each judgment focusing on a single defendant charged with homicide.

By analyzing these sentencing decisions, this paper aims to understand the impact of various sentencing factors on the judges' decisions. Insights derived from this research could improve the AI model’s ability to assess sentencing factors, thereby helping both professional and citizen judges determine appropriate sentencing outcomes for homicide cases.

2. Research Findings

In the 198 homicide judgments analyzed in this study, the sentencing outcomes were distributed as follows:
a. 73 cases resulted in fixed-term imprisonment of 5 to 10 years, b.63 cases resulted in imprisonment of less than 5 years, c.52 cases resulted in imprisonment of more than 10 years, d.10 cases resulted in life imprisonment. (see Fig.1)

Our analysis revealed that the most frequently cited sentencing factors influencing the severity of sentences were post-offense attitude, intellectual capacity and living conditions of the offender.

Generally, when defendants had a criminal record or displayed a negative attitude after the offense, judges tended to impose harsher sentences.

Conversely, if the defendant came from a dysfunctional family or suffered from physical or mental health issues, judges were more likely to impose lighter sentences.

The findings highlight the significant role these specific sentencing factors play in judicial decisions. Understanding how these factors are weighed provides valuable insights for refining the AI sentencing system and ensuring that it aligns with the sentencing practices observed in homicide cases.

2.1 Most Frequently Used Sentencing Factors

2.1.1 Post-offense Attitude

Among the 198 judgments analyzed, 180 cases considered the defendant's post-offense attitude. This factor could have either a positive or negative impact on sentencing, depending on the defendant's behavior and statements.

In cases where the defendant showed remorse and took responsibility for their actions, judges were inclined to reduce the sentence. For example, in a case heard by the Taoyuan District Court (111 Chong-Su Zi No. 8), the defendant admitted to killing his son during police questioning, investigation, and trial. He expressed deep regret, saying, “I feel very sorry for killing my son and wish I could have a second chance.” The judge interpreted this behavior as a sincere acknowledgment of guilt and a plea for forgiveness, leading to a mitigated

sentence. This is positive impact.

Conversely, a defendant’s denial of guilt or attempts to shift blame typically resulted in harsher sentencing. In another case heard by the Taoyuan District Court (106 Su Zi No. 912), the defendant struck the victim with a bat, causing death, out of anger over blocked traffic. After the offense, the defendant denied any wrongdoing, claiming he acted to assist police officers in an emergency. However, the court determined that this claim was unfounded based on other evidence. The judge noted that the defendant not only failed to reflect on his actions but also sought to evade responsibility by offering false justifications. As a result, the court increased the severity of the sentence. This is negative impact.

Those examples illustrate how judges weigh the defendant's post-offense attitude, with genuine remorse leading to leniency, while evasion and lack of accountability result in stricter punishment.

2.1.2 Intellectual Capacity

In the 198 judgments analyzed, 140 cases considered the defendant's intellectual capacity, making it the second most frequently referenced sentencing factor. This factor involves evaluating the offender's mental health and cognitive abilities, which can significantly impact their accountability and the severity of their sentence.

Example Case 1. In a case from the Taipei District Court (109 Su Zi No. 514), the defendant had been diagnosed with schizophrenia and held a disability certificate. They exhibited clear symptoms of delusions of persecution and auditory hallucinations in daily life. Although the defendant had undergone mandatory hospitalization, their condition worsened after discharging due to non-compliance with medication. In a mentally deteriorated state, the defendant randomly attacked a motorcyclist stopped at a red light with a folding knife.

The court acknowledged that the defendant's mental health significantly impaired their ability to discern right from wrong. As a result, the judge imposed a sentence of three years of imprisonment, which is far below the statutory minimum of ten years for such crimes. This case illustrates how the offender's psychological and cognitive impairments are essential factors in judicial decision-making, as they may justify mitigated sentences to account for diminished responsibility.

2.1.3 Living Conditions of the Offender

In the 198 judgments analyzed, 129 cases considered the living conditions of the offender, indicating that this factor is an important consideration for judges during sentencing.

Example Case 2. In one case adjudicated by the Taoyuan District Court (110 Su Zi No. 1116), the judge considered the defendant's impoverished background and the fact that they had a minor child to support. Recognizing the significant personal and familial hardships faced by the defendant, the court opted to impose a sentence of five years of imprisonment, which is notably below the statutory minimum of ten years for homicide offenses.

This example highlights how the defendant's socio-economic circumstances can influence sentencing decisions. Judges may exercise discretion to impose lighter sentences based on the offender's personal situation, particularly when it involves familial responsibilities or significant hardship. Such considerations reflect a more compassionate approach to justice, recognizing the complexities of individual circumstances that can impact criminal behavior.

2.2 Factors Influencing Severe Sentences

In this study, there were no death penalty cases; the most severe sentence issued was life imprisonment. Judges typically consider several factors when imposing life sentences, including the defendant's character (such as having a prior criminal record), poor post-offense attitude, or lack of remorse, which can lead to increased sentencing severity. (see Fig.2)

Another significant reason for imposing harsher sentences is the means employed in committing the crime.

Example Case 3. In a case from the Taitung District Court (107 Chong-Su Zi No. 1), the defendant had a troubled relationship with his wife, involving financial disputes and previous incidents of domestic violence. Despite a protection order issued by the court, the defendant, suspicious of his wife's alleged infidelity, armed himself with a kitchen knife and a fruit knife. He went to her workplace and inflicted multiple stab wounds to her head, abdomen, back, and legs, resulting in severe and irreparable injuries, including the severing of her vertebral

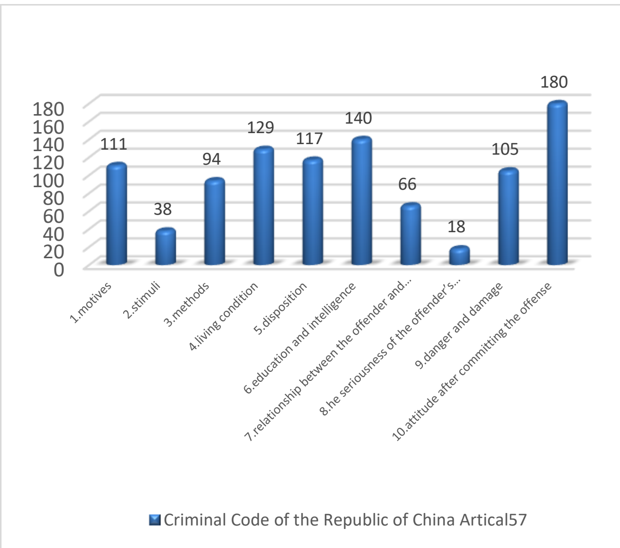


Fig.2. Criminal Code of the Republic of China, Article 57

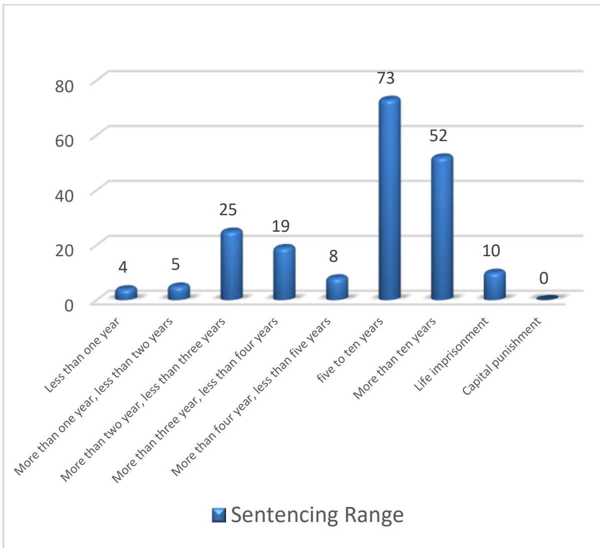


Fig.1. Sentencing Range

artery. The brutality of the defendant's actions was evident.

Consequently, the judge determined that the circumstances surrounding the crime were exceptionally violent and deserving of a severe penalty, leading to a life sentence for the defendant.

This case illustrates how judges consider the nature and brutality of the criminal act, alongside the defendant's prior behavior and attitude, when determining the appropriateness of a life sentence. Such factors reflect the judicial system's commitment to ensuring that severe crimes receive equally severe consequences, thereby upholding public safety and justice.

2.3 Factors Influencing Lenient Sentences

Under the Criminal Code of the Republic of China, the statutory penalty for homicide is imprisonment for ten years or more. However, judges may sometimes reduce the sentence to below ten years by considering factors such as the defendant's post-offense attitude and their living conditions.

Example Case 4. In a case from the Taipei District Court (109 Chong-Su Zi No. 8), the defendant, unable to bear the suffering of their daughter, who had been bedridden due to illness, decided to end her life by suffocation. Afterward, the defendant fully confessed to the crime. The judge took into account the compassionate motive behind the act—stemming from the daughter's prolonged suffering—and the defendant's positive post-offense attitude. The family also expressed understanding and requested leniency from the court. Considering these circumstances, the judge sentenced the defendant to two years and six months in prison.

Example Case 5. In another case heard by the Taoyuan District Court (109 Chong-Su Zi No. 35), the defendant, overwhelmed by the stress of caring for her husband, developed severe depression during the caregiving period and contemplated killing her husband before taking her own life. The judge assessed that the defendant's actions were not malicious in intent. Additionally, the defendant's voluntary confession and deteriorating mental state were taken into consideration. The victim's children expressed their desire for the defendant not to be imprisoned during the proceedings. Given these mitigating factors, the judge imposed a sentence of one year and six months in prison, with a three-year probation.

These cases illustrate how judges weigh the humanizing factors surrounding the offense, such as compassion, mental health struggles, and family dynamics, allowing for leniency in sentencing. By considering the context of the crime and the defendant's circumstances, the judicial system seeks to balance justice with compassion, recognizing that not all criminal actions stem from a place of malice.

3. Conclusion

The use of an AI sentencing information system to assist judges in sentencing may theoretically face criticism for potentially interfering with judicial independence. However, this study finds that its practical implementation is feasible. The system's sentencing recommendations can help remind judges of appropriate sentencing boundaries, thereby reducing the likelihood of similar cases resulting in disparate sentences and promoting fairness in the judicial process.

Nevertheless, it is essential to recognize that each criminal case presents unique facts and circumstances. Therefore, the information provided by the AI sentencing system should only serve as one reference point among many when judges determine the appropriate sentence. Judges must still rely on the specific details of each case to make comprehensive and informed sentencing decisions, ensuring that justice is served in a manner that reflects the nuances of individual situations.

Reference

1. Wen Jiaqian, discussing the sentencing evaluation of national judges from the perspective of sentencing tools, Judicial Weekly Issue 2156, pp. 1-3, May 2023.
2. In response to the new system of national judges, the Judicial Yuan launched an AI sentencing information system - with two modes and four major advantages, Judicial Yuan News Network, February 6, 2023, https://www.judicial.gov.tw/tw/cp-1887-806741-d6471-1.html
3. https://www.judicial.gov.tw/tw/cp-1429-810543-42fe0-1.html (accessed 2024-12-15)

Authors Introduction

Dr. Shih-Chin LIN

He is an assistant professor in the Department of Financial Law, Ming Chuan University (MCU), Taiwan, and also a member in Taiwan Information Security Center @ National Cheng Kung University (TWISC@NCKU). He graduated from Cheng Chi University (NCCU), Taiwan, with B.S. in 1999 and Master degree in Criminal Law in 2003. He obtained second Master degree in Procedural Law in 2010 and his PhD in Juvenile Criminal Law in 2017, from the Aix-Marseille University (AMU), France. His research interests focus on the Criminal Justice for Juvenile Offenders, Cybercrime, Cyber-Security and Rights of Children.

Mr. Cheng-Tsung YEH

He is a 21-year-old student majoring in Law at Ming Chuan University. He has a strong interest in the field of criminal law and hopes to conduct in-depth research on issues related to this field in the future.

Mr. Chen-Yu LAI

He is a student from the Department of Law at Ming Chuan University, with an interest in the fields of criminal law and intellectual property law. He hopes to enhance learning and exchanges through this seminar.

Miss. Chih-Yun CHANG

She is a student in the Department of Law, Ming Chuan University (MCU), Taiwan, and is now 21 years old. She is interested in various criminal legal issues and hope to learn a lot from this seminar. She hopes to have more opportunities to participate in related in the feature.

Miss. Chi-Ju FU

She is a senior student in Law at Ming Chuan University. She has a strong interest in criminal law, enjoys reading , and is skilled at gathering information. In the team, she takes on a coordinating role. And hopes to gain new knowledge and experience through this seminar.

Inferring ICS Topology and Behavior through Network Traffic Analysis

Chien-Wen Tseng
M.S. Degree Program on Cyber-Security Intelligence, National Cheng Kung University
No.1, University Rd., East Dist., Tainan City 701401, Taiwan

Jung-Shian Li
Department of Electrical Engineering / Institute of Computer and Communication Engineering, National Cheng Kung University
No.1, University Rd., East Dist., Tainan City 701401, Taiwan

Chu-Fen Li
Department of Finance, National Formosa University
No.1, University Rd., East Dist., Tainan City 701401, Taiwan

I-Hsien Liu
Department of Electrical Engineering, National Cheng Kung University
No.1, University Rd., East Dist., Tainan City 701401, Taiwan

E- mail: cwtseng@cans.ee.ncku.edu.tw, jsli@cans.ee.ncku.edu.tw, chufenli@gmail.com, ihliu@cans.ee.ncku.edu.tw *
www.ncku.edu.tw

Abstract

Ensuring the proper operation of industrial control systems (ICS) is a critical issue. Previous studies have focused on observing the controllers directly; however, this research proposes that analyzing network traffic could reduce system interference. This study utilizes a cybersecurity testing platform for dam systems to investigate network traffic to infer the composition of network nodes and the communication behavior between them. By doing so, it provides an alternative perspective for monitoring ICS operations. This model aids in understanding system topology, tracking potential deviations, and ensuring operational stability.

Keywords: Industrial Control Systems, Network Traffic Analysis, Industrial Control Communications.

1. Introduction

In 2024, a significant cyberattack in a rural Texas town targeted the water management system, leading to an overflow incident [1]. Although this event did not pose an immediate threat to public safety, it served as a stark reminder of the potential risks associated with similar attacks. Water overflow incidents can lead to severe consequences, including water contamination, disruptions in supply, and damage to critical infrastructure [2], which could compromise public safety and incur significant economic losses. This incident highlighted the growing vulnerability of critical infrastructure to cyber threats, especially as these systems increasingly rely on interconnected technologies and controllers to manage essential processes [3] . The event emphasized the need for robust monitoring mechanisms to strengthen ICS resilience against cyberattacks.

The scarcity of publicly available datasets related to water resource management systems poses significant challenges to conducting in-depth studies in this field. The limited data not only restricts comprehensive analysis but also makes it difficult to gain a clear understanding of system behaviors and interactions. In response to such challenges, the previous experimental teams leveraged the CANS RT/DT dataset [4] to observe and analyze ICS behaviors. Researchers aimed to gain a deeper understanding of critical elements in the field, focusing on their functional roles, operational behaviors, and how these behaviors evolve under various conditions.

Given the challenges in understanding water resource operations, this study conducts a preliminary analysis of the CANS RT/DT dataset [4] [5] to observe and explore ICS components, their behaviors, and interactions for initial insights. By systematically examining this dataset, the research aims to support the future establishment of a normal operational model. This involves identifying patterns and relationships between system components, which can serve as a foundation for modeling standard behaviors.

2. Background

2.1. Industrial Control System

Supervisory Control and Data Acquisition [6] (SCADA)-based Industrial Control Systems (ICS) are vital for critical infrastructure, enabling centralized monitoring and control of essential processes such as dam operations, energy grids, and manufacturing systems. SCADA systems rely on Programmable Logic Controllers (PLCs) to bridge field devices with higher-level supervisory systems. PLCs execute predefined instructions to manage Digital Inputs/Outputs (DI/DO) for binary operations, such as controlling pumps and alarms, and Analog Inputs/Outputs (AI/AO) for continuous monitoring of parameters like pressure and temperature [7] This hierarchical architecture facilitates real-time control and communication, ensuring efficient and reliable management of processes.

2.2. Modbus Traffic Analysis

Analyzing network traffic provides a non-intrusive and comprehensive approach to understanding the behavior of ICS environments and identifying potential vulnerabilities. By examining communication patterns and packet characteristics, researchers can infer system topology and evaluate interactions between devices.

Within this context, Modbus TCP [8] plays a pivotal role in ICS communication. As an extension of the widely adopted Modbus protocol, Modbus TCP leverages TCP/IP for enhanced connectivity, enabling data exchange between devices such as PLCs and sensors within a master-slave architecture. The protocol uses function codes, such as 0x03 for reading holding registers, to monitor and manipulate device states. PLCs often use Modbus TCP to transmit operational data, including sensor readings and actuator commands. Analog inputs like voltage or current are stored in two 16-bit registers and converted into IEEE 754 floating-point numbers [9] [10] for precise representation. This communication facilitates real-time data sharing and control, making Modbus TCP a fundamental protocol for understanding ICS network behavior.

After capturing the network traffic data, key attributes such as source and destination IP addresses, timestamps, register values, and MAC addresses are extracted from the Modbus TCP packets. These attributes provide a detailed view of communication patterns and operational data within the system, enabling a closer examination of how components interact. The extracted data is further processed to derive statistical insights, such as communication frequency, response times, and register value distributions. These insights are useful for understanding system behavior and establishing references for future analysis.

3. Methodology Overview for Network Traffic Analysis

This study draws inspiration from framework ATT&CK, specifically the Reconnaissance phase [11], to inform the methodology for analyzing the CANS RT/DT dataset and understanding Industrial Control Systems (ICS). While Reconnaissance typically involves active network scanning to map system components and interactions, it has potential to disrupt normal operations and their limited scope in capturing comprehensive data. Instead, Wireshark pcapng file from the CANS RT/DT dataset are utilized to extract and process network traffic data in a non-intrusive manner. The methodology focuses on passive monitoring, leveraging pre-collected network traffic to avoid disruptions to normal operations. By focusing on Modbus TCP packets [12], which are commonly used in ICS communication. This approach emphasizes passive monitoring, utilizing pre-collected network traffic to enable effective observation of system activities.

The data analysis methodology is based on a structured workflow, as illustrated in Fig. 1 Network traffic is captured within the ICS environment using packet tool Wireshark, with a focus on filtering Modbus TCP packet. To ensure passive monitoring, a mirror port is set up on the network switch, allowing traffic to be observed without affecting live operations. The captured traffic is parsed to extract relevant details and generate metrics that reflect communication trends and system interactions.

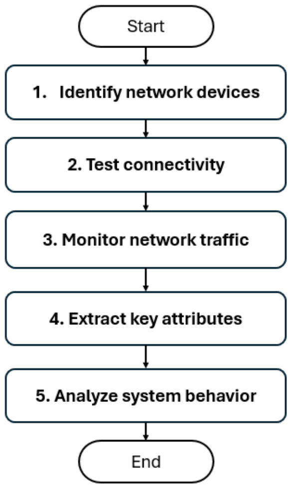


Fig. 1. systematic workflow for data analysis

Further analysis involves correlating the captured packet timestamps with the observed gate states to determine whether changes in flow rates correspond to variations in gate state indicators. The gate state indicators, such as power supply status, gate movement, and emergency operations, are mapped to specific attributes like register values in the data. By analyzing the relationships between flow rate differences and gate states, it becomes possible to infer how gate operations influence system behavior. These observations help establish a foundational understanding of the dynamic interactions between gate controls and water flow, contributing to a broader perspective on system operations.

4. Result of observation

4.1. *System topology*

Based on the provided network Fig. 2, it is possible to infer that the central node 192.168.1.50 might represent a switch, serving as a hub to facilitate communication between other devices, such as the PLCs (192.168.1.5, 192.168.1.6, 192.168.1.7). The switch could be primarily responsible for forwarding data packets without directly engaging in logical operations. The edges represent the number of response packets exchanged between the devices during a one-hour capture period, with a total of 7176 packets recorded.

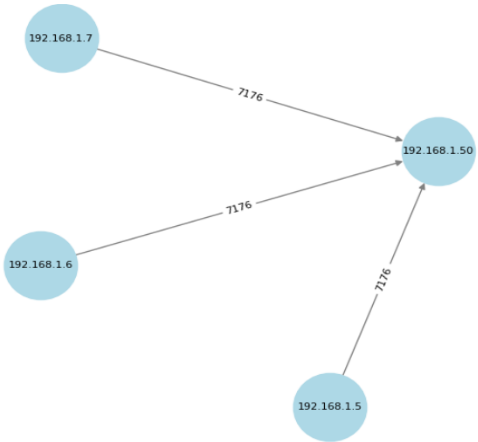


Fig. 2. System topology

4.2. *PLC's state*

The Modbus register values exhibit distinct patterns over a two-minute period. Take Fig. 3 for IP 192.168.1.5 and Fig. 4 for 192.168.1.6 as an example. The Modbus registers in the dataset are organized with specific functions and interpretations. Register 0 records the binary states of 16 digital inputs, where each bit represents the TRUE/FALSE status of corresponding gate operations. This allows for a concise representation of multiple discrete signals within a single register. Register 1 was not observed in the dataset during the analysis. Registers 3 and 2 store gate voltage values, Registers 5 and 4 capture gate current readings, and Registers 7 and 6 record gate opening degree. These paired registers are combined into 32-bit integers and converted to floating-point numbers using the IEEE 754 standard, providing precise measurements for continuous parameters such as voltage, current, and position. The ordering of the registers may appear inverted due to the little-endian format often used in Modbus systems, where the least significant word is stored before the most significant word. This structure ensures compatibility with standard protocols and simplifies data processing for ICS operations.

Fig. 3. Response of Register value response from IP192.168.1.5

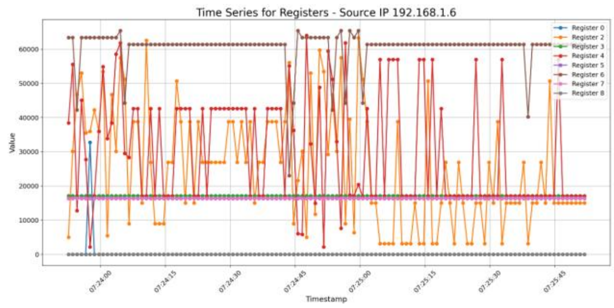


Fig. 4. Response of Register value response from IP192.168.1.6

The dataset reveals several limitations. Most packets used Function Code 3, indicating a focus on monitoring register values, likely reflecting on-site operations, with no observable changes in Function Codes even during remote operations. Additionally, when current spikes coincided with gate opening degree changes, the corresponding gate operation indicators were not captured promptly, likely due to the discrete nature of state recordings, which may miss transient system changes.

5. Conclusion

This study explores the potential of analyzing the operational behavior of industrial control systems (ICS) by interpreting key system attributes, such as IP addresses and register values, to observe device status and operational patterns. The passive monitoring approach minimizes interference with on-site operations, making it suitable for preliminary observations. These observations provide a foundation for developing a normal operational model, which can be used as a reference to enhance system reliability.

Acknowledgment

This work was supported by the National Science and Technology Council (NSTC) under contract number 113-2221-E-006-147-MY2 and 113-2634-F-006-001-MBK.

References

1. Associated Press, "Rural Texas Towns Report Cyberattacks That Caused One Water System to Overflow," 2024 [Online]. Available: https://www.securityweek.com/rural-texas-towns-report-cyberattacks-that-caused-one-water-system-to-overflow/. [Accessed 15 Oct 2024].
2. J-P. A. Yaacoub, O. Salman, H. N. Noura, N. Kaaniche, A. Chehab, M. Malli, "Cyber-physical systems security: Limitations, issues and future trends," Microprocessors and Microsystems, vol. 77, p. 103201, 2020.
3. M. M. Aslam, A. Tufail, R. A. A. H. M. Apong, L. C. De Silva, M. T. Raza, "Scrutinizing Security in Industrial Control Systems: An Architectural Vulnerabilities and Communication Network Perspective," IEEE Access, vol. 12, pp. 67537 - 67573, 2024.
4. CANS, "CANS RT," 2024. [Online]. Available: https://www.cans.ee.ncku.edu.tw/research/cansrt. [Accessed 15 Oct 2024]
5. M.-W. Chang, J.-S. Li, and I-H. Liu, "Cyber-Physical Security Testbed for Dam Control System", Journal of Advances in Artificial Life Robotics, Vol. 4, No. 2, pp. 63-66, 2023.
6. D. Pliatsios, P. Sarigiannidis, T Lagkas, A. G. Sarigiannidis, "A Survey on SCADA Systems: Secure Protocols, Incidents, Threats and Tactics," IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 1942 - 1976, 2020.
7. U. George-Andrei, P. Olga, U. Maria, "PLCs' Inputs and Outputs Response Time Testing Application," 2021 International Conference on Electromechanical and Energy Systems (SIELMEN), Iasi, Romania, 06-08 Oct , 2021.
8. V. G. Găitan, I. Zagan, "Modbus Protocol Performance Analysis in a Variable Configuration of the Physical Fieldbus Architecture," IEEE Access, vol. 10, pp. 123942 - 123955, 2022.
9. S. Jaloudi, "microIoTSCADA: A Tool to Monitor and Control Renewable Energy-Based Systems Using MODBUS," in 2024 4th Interdisciplinary Conference on Electrics and Computer (INTCEC), Chicago, IL, USA, 11-13 Jun , 2024.
10. Microprocessor Standards Committee, IEEE Standard for Floating-Point Arithmetic, IEEE Std 754-2019, IEEE Computer Society, 2019.
11. MITRE, "Reconnaissance," Oct 2020. [Online]. Available: https://attack.mitre.org/tactics/TA0043/. [Accessed 20 Oct 2024].
12. P. Radoglou-Grammatikis, I. Siniosoglou, T. Liatifis, A. Kourouniadis, K. Rompolos, P. Sarigiannidis, "Implementation and Detection of Modbus Cyberattacks," 2020 9th International Conference on Modern Circuits and Systems Technologies (MOCAST), Bremen, Germany, 07-09 Sep , 2020.

Authors Introduction

Ms. Chien-Wen Tseng

She is a postgraduate of M.S. Degree Program on Cyber-Security Intelligence, National Cheng Kung University in Taiwan. She received her B.B.A. degree from the Bachelor of Information Management, Yuan Ze University, Taiwan in 2023. Her interests is ICS Security.

Dr. Jung-Shian Li

He is a full Professor in the Department of Electrical Engineering, National Cheng Kung University, Taiwan. He graduated from the National Taiwan University, Taiwan, with B.S. in 1990 and M.S. degrees in 1992 in Electrical Engineering. He obtained his PhD in 1999 in Computer Science from the Technical University of Berlin, Germany. He teaches communication courses and his research interests include cybersecurity, cloud computing and network management. He is currently involved in funded research projects dealing with cybersecurity and critical infrastructure protection. He is the director of Taiwan Information Security Center @ National Cheng Kung University.

Prof. Chu-Fen Li

She is an Associate Professor in the Department of Finance at the National Formosa University, Taiwan. She received her PhD in information management, finance and banking from the Europa-Universität Viadrina Frankfurt, Germany. Her current research interests include intelligence finance, e-commerce security, financial technology, IoT security management, as well as financial institutions and markets. Her papers have been published in several international refereed journals such as European Journal of Operational Research, Journal of System and Software, International Journal of Information and Management Sciences, Asia Journal of Management and Humanity Sciences, and others.

Dr. I-Hsien Liu

He is an assistant professor in the Department of Electrical Engineering, National Cheng Kung University, Taiwan. He obtained his Ph.D. in 2015 in Computer and Communication Engineering from the National Cheng Kung University. He teaches cybersecurity courses and his interests are Cyber-Security, OT Security, and Wired & Wireless Communication. He is the deputy director of Taiwan Information Security Center @ National Cheng Kung University (TWISC@NCKU).

Using fuzzy control routing for dynamic load balancing over Software-Defined Networks

Wen-Shyang Hwang
Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 807, Taiwan; Email: wshwang@nkust.edu.tw
Ming-Syuan Wu*
Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 807, Taiwan; Email: i111154101@nkust.edu.tw
Cheng-Han Lin
Department of Health-Business Administration, Fooyin University, Kaohsiung 831, Taiwan
Yan-Jing Wu
Department of Information Technology and Communication, Shih Chien University, Kaohsiung 807, Taiwan
Ming-Hua Cheng
Department of Digital Media Design, Tzu-Hui Institute of Technology, Pingtung, 926, Taiwan
Sian-Fong Huang
Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 807, Taiwan

Abstract

The load-balancing and forwarding mechanisms of traditional software-defined networks (SDNs) often rely on static path selection. Conventional mechanisms lead to uneven resource utilization and bottlenecks under high-traffic conditions. Therefore, the existing methods require more flexibility in decision-making. This study introduces a load-balancing algorithm based on fuzzy logic to address the difficulty of setting multivariable input thresholds. The algorithm uses fuzzy logic to combine the network parameters (queue length, link utilization, link delay, and packet loss rate) as fuzzy inputs. Converting the fuzzy input into a score is the key to achieving optimal traffic allocation in the mechanism. The proposed mechanism improves resource utilization and reduces bottlenecks and reliability in dynamic network-load situations. The results indicated that the proposed method achieved a higher throughput under high-load conditions. Moreover, the results maintained low packet loss and delay rates. Load balancing based on fuzzy logic provides an effective solution for SDN environments.

Keywords: Software Defined Network, fuzzy logic, load balancing

1. Introduction

Software-defined networks (SDNs) [1] have become an essential part of modern network architecture. SDNs makes the network management and configuration more flexible and efficient by separating the control plane from the data plane. However, load balancing remains a critical and challenging issue in an SDN implementation. Most existing load-balancing mechanisms are based on static configurations, which have many problems and limitations in practical applications. Static load-balancing mechanisms usually set fixed paths for data forwarding, which are simple and limited in a dynamically changing network environment. As network traffic changes, fixed-path configurations cannot adapt to real-time needs, causing some paths to be overloaded, while other paths remain idle. Rough network resource utilization significantly affects the network performance. Particularly, in high-traffic situations, single-path selection can create bottlenecks, causing network congestion and increasing latency. Static load balancing lacks flexibility and cannot dynamically adjust path selection based on the real-time network status. Because traffic patterns change and links fail, the system cannot promptly adapt, thereby reducing network reliability and stability. In addition, the existing SDN load-balancing mechanisms require greater flexibility in the decision-making process. This study proposes a load-balancing mechanism to improve the flexibility and adaptability of the load-balancing mechanism through fuzzy logic. The remainder of this paper is organized as follows: Section 2 introduces the background and related work. Section 3 describes our research methodology. Section 4 describes the evaluation and experiments. Finally, Section 5 presents conclusions and future work.

2. Backgrounds and related works

2.1. Fuzzy logic theory and Equal-cost multi-path (ECMP)

Fuzzy logic is the mathematical logic theory [2]. This allows the processing of uncertainty and fuzziness by introducing the concept of a fuzzy set. Unlike traditional binary logic, fuzzy logic allows partial truth values. The variables range from 0 to 1.

The fuzzy-control process consists of three stages: input, processing, and output. These stages form a fuzzy inference system that manages uncertain and fuzzy data for effective control and decision making. The Mamdani fuzzy inference system, one of the earliest types, includes four substages: fuzzification, inference engine, fuzzy rules, and defuzzification. ECMP [3] is a load-balancing algorithm that evenly distributes traffic across multiple equal-cost paths. However, ECMP cannot improve network resource utilization and performance.

2.2. An effective routing mechanism based on fuzzy logic for software-defined data center networks

Fuzzy logic intelligently evaluates data packet paths and is primarily used in data center networks. The mechanism considers factors such as hop count and bandwidth utilization. In [4], the model converted these factors into fuzzy variables by assigning a corresponding label to each variable. With an essential fuzzy variable set, a set of fuzzy rules is designed to control the path selection based on network requirements. These rules make precise decisions for the SDN controller, whose path must be selected under specific network conditions. The decision-making process selects routes using multiple indicators and identifies the best path for effective load balancing.

3. Fuzzy logic load balancing algorithm

3.1. System architecture

In the experiment, a Ryu controller and Mininet simulator were selected to build the SDN simulation environment. Fig. 1 shows the operation process of the fuzzy logic load-balancing algorithm (FLLB) in an SDN network.

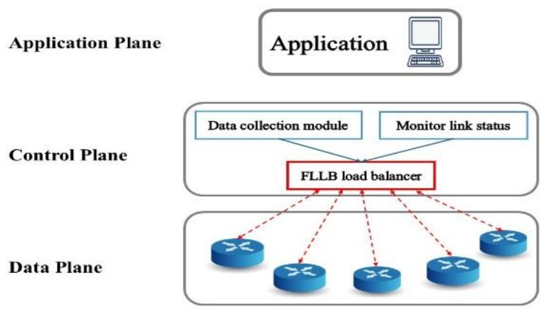


Fig. 1 System architecture

3.2. FLLB mechanism operation

The operation process of the FLLB algorithm is suitable for SDNs. The system was designed to select links dynamically for efficient load balancing and network performance optimization. The link-screening process evaluates the load score (L_i) and quality score (Q_i) of each link using a fuzzy logic system, which confirms that the selected link can reach the destination in the FLLB mechanism. The connection control module generates new flow-table rules based on the final decision results, which are dynamically delivered to the switch to ensure that the data packets are transmitted along the best link. After calculating the comprehensive score (S_i) of each link, link selection is performed based on the score. First, it is judged whether S_i is greater than or equal to 0.5. If the value was less than 0.5, the link was excluded as a candidate link. If S_i of the candidate links is less than or equal to 1, then the original link is followed for traffic forwarding. If S_i of the candidate links exceeds 1, the link with the highest comprehensive score is selected. After selecting the best link, we must determine whether the link can reach the destination. Otherwise, the candidate link with the second highest comprehensive score is chosen.

3.2.1 Fuzzy logic load balancer

In the experiment, the collected information was passed to a fuzzy-logic load balancer in the FLLB mechanism. The load balancer applies fuzzy logic to assess the link feasibility and effectiveness. It selects optimal links based on these evaluations.

The queue length, link utilization, delay, and packet loss rate are the fuzzification input variables. These factors determine the link-load capacity and quality. Queue length is the wait time for the switch port to process a packet. The fuzzifying queue length addresses the uncertainty in network conditions. Link utilization indicates the traffic load as a percentage of the maximum capacity, which directly reflects the resource usage status. The queue length and link utilization were used as fuzzification input variables. According to [5], the fuzzification of these two variables can be improved to obtain a more accurate L_i, which can evaluate the load status of the link more comprehensively.

Link latency is the time required for a packet to transition from a source node to a destination node on a specific link, which directly reflects the transmission performance of the link. The link delay is used as a fuzzy input variable because it can reflect link performance and predict link congestion. The packet loss rate directly reflects link reliability during transmission. The packet loss rate and link delay are the fuzzification inputs for obtaining Q_i. This score reflects link reliability and performance as well as simplified decision-making. According to [6] and [7], the equation for calculating the fuzzification of each input parameter is: Fuzzification generates a fuzzy input for each input using a membership function. The inference engine produces nine fuzzy outputs based on the L_i rule table. Q_i rule table produced six fuzzy outputs, as illustrated in Fig. 2.

Load Score (L_i)		
Queue Length	Link Utilization	Output
Low	Low	VL
Low	Medium	L
Low	High	ML
Medium	Low	L
Medium	Medium	M
Medium	High	MH
High	Low	ML
High	Medium	MH
High	High	H

Quality Score (Q_i)		
Delay	Packet Loss Rate	Output
Low	Low	H
Low	Medium	MH
Low	High	M
High	Low	ML
High	Medium	L
High	High	VL

Fig. 2 Fuzzy inference engine rule table

The fuzzy output generated by the inference engine must be converted into a crisp output through the defuzzification process. The center-of-gravity method is a standard defuzzification method, as expressed in Eq. (1). The final result was determined by calculating the center points of all overlapping areas, ensuring that each overlapping area was calculated only once. This defuzzification method accurately reflected the output characteristics of the system. In Eq. (1), x_i is the value of the output variable, $x_i\mu$ is the membership degree corresponding to x_i, and N is the number of values of the output variable.

$$Defuzzification(x) = \frac{\sum_{i=1}^{N} x_i \mu(x_i)}{\sum_{i=1}^{N} \mu(x_i)} \quad (1)$$

3.2.2 Comprehensive link score calculation

In the FLLB algorithm, the calculation of S_i is the key to determining the optimal traffic distribution strategy. The calculation process for the comprehensive score includes the following steps. The fuzzy logic system is based on four fuzzy inputs: queue length, link utilization, delay, and packet loss rate. Two fuzzy outputs are calculated: L_i and Q_i. L_i reflects the current load on the link. The lower the load, the lower the score. Q_i reflects the quality of links. In this study, to effectively combine link load and quality to determine the optimal link, we propose a comprehensive score equation as Eq. (2). L_i represents the load score of the *i*th link. Q_i denotes the quality score of the *i*th link. where $\frac{1}{L_i}$ represents the reciprocal of the link load. This design prioritizes lighter-loaded links because lighter-loaded links are more effective in reducing network congestion and higher-loaded links obtain a lower weight. Links of higher quality were considered suitable for selection. The advantage of this equation is that it can be adjusted dynamically. When the link load is light, $\frac{1}{L_i}$ will be larger; therefore, S_i of the link will be higher, and vice versa, to realize the optimization of low-load links.

$$S_i = \frac{1}{L_i} + Q_i \quad (2)$$

4. Evaluation and Experimental Results

4.1. Experimental environment

The simulation architecture for this experiment is shown in Fig. 3, and the parameters are listed in Table 1. The Ryu controller collects network status information and dynamically adjusts traffic forwarding using the FLLB algorithm to achieve the best network performance. The study uses a PC with a hardware performance of i7-12700h 4.7GHz and DDR5-4800Mhz 24GB RAM to install an Ubuntu 16.04 Linux environment. A Mininet simulator was used to build the SDN environment. There were five hosts and six switches. This study uses a fuzzy module to implement a fuzzy logic system and integrate it into an SDN controller.

4.2. Experimental Results

The FLLB algorithm was compared and analyzed with two different load-balancing algorithms: traditional equal-cost multipath routing (ECMP) [3] and fuzzy logic routing mechanism (FLRM) [4]. Fig. 4 shows the total throughput of the different load-balancing algorithms in the simulation experiments. The figure compares the total throughput changes of the three load-balancing algorithms, FLLB, FLRM, and ECMP, within 20 s. From 15 to 20 s, FLLB has an advantage. At all time slots, the throughput of FLLB was slightly higher than those of FLRM and ECMP.

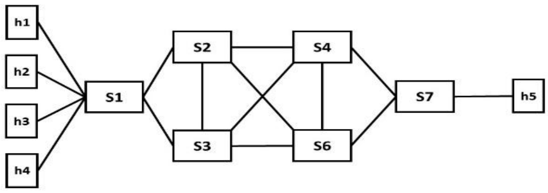


Fig. 3 Simulation topology

Table 1. Environment parameters

Attributes	Value
Simulator	Mininet 2.5
SDN Controller	Ryu Controller
SDN Protocol	OpenFlow V1.3
Switch	OpenFlow Switch
Packet generate tool	Iperf
Traffic type	UDP
Traffic flow	100Mb,150Mb,250Mb,300Mb
Link bandwidth	500Mbps
Packet size	1500byte(MTU)
Queue size	100 packets
monitoring period	1s
Simulation time	20s
(h1 to h5)	0s start
(h2 to h5)	5s start
(h3 to h5)	10s start
(h4 to h5)	15s start

Fig. 5 shows the packet delay changes of the different load-balancing algorithms within 20 s. Starting from the 5th second, the delays between the different algorithms begin to differ as the traffic increases. Under high load conditions from 15 to 20 s, FLLB's latency remains at about 1.4 ms, while FLRM's latency is approximately 1.6 ms. By contrast, the latency of ECMP increases significantly, which means that its ability to deal with sudden, large traffic is weak. FLLB has significant advantages under dynamic changes and high-load conditions such that it can manage network traffic. Fig. 6 shows the change in packet loss rate over time under different load-balancing mechanisms. ECMP allocates traffic based on the exact link cost and does not consider the real-time network status. Load imbalance can lead to increased packet loss rates in overloaded links. From 14 to 20 s, Fig. 6 shows the situation when high traffic is injected. This means that FLLB can respond quickly and stabilize the network performance when handling high traffic.

5. Conclusion

This study aimed to explore and improve the load-balancing mechanism in a network. The FLLB algorithm was proposed because of the shortcomings of traditional load-balancing methods. Under the SDN architecture, FLLB can dynamically adjust the traffic distribution according to the actual status of the link. In future, we will refer to [8] to implement the FLLB algorithm in B5G or 6G network environments, which will improve the overall QoS.

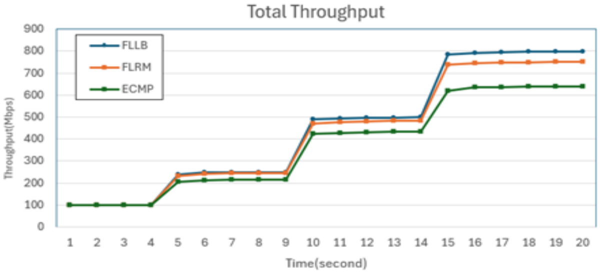


Fig. 4 Total throughput

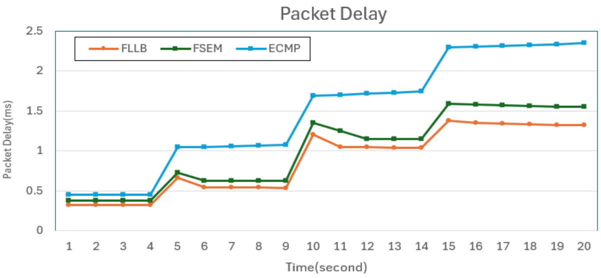


Fig. 5 Packet delay

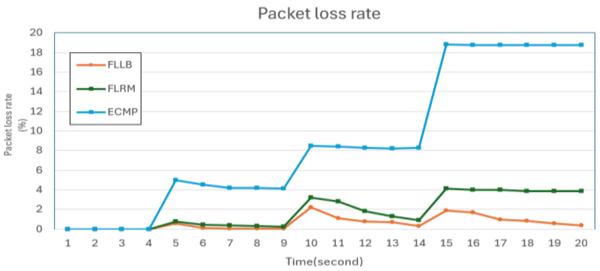


Fig. 6 Packet loss rate

Acknowledgments

This study was supported by the National Science and Technology Council (NSTC), Taiwan, under Contract No. NSTC 113-2221-E-242-001, NSTC 113-2221-E-992-001 and NSTC 113-2221-E-158-001.

References

1. Xabier Etxezarreta, Iñaki Garitano, Mikel Iturbe, Urko Zurutuza," Software-Defined Networking approaches for intrusion response in Industrial Control Systems: A survey," International Journal of Critical Infrastructure Protection, vol 42, 100615, 2023.
2. Mamdani, Ebrahim H., "Application of fuzzy algorithms for control of simple dynamic plant," Proceedings of the Institution of Electrical Engineers, pp1585-1588, 1974.
3. C.Hopps, "Analysis of an Equal-Cost Multi-Path Algorithm", IETF RFC 2992, 2000.
4. Yang Xu, Wu Muqing, Yao Guohao, "An Effective Routing Mechanism Based on Fuzzy Logic for Software-Defined Data Center Networks," in 2020 IEEE 6th International Conference on Computer and Communications (ICCC), 2020.
5. O. Feuser, A. Wenzel, "On the effects of the IEEE 802.3x flow control in full-duplex Ethernet LANs", in Proceedings 24th Conference on Local Computer Networks. LCN'99, 1999.
6. ITU-T Y.1541, Network performance objectives for IP-based services, 2006.
7. ITU-T G.114, One-way transmission time, 2003.
8. S. H. A. Kazmi, F. Qamar, R. Hassan and K. Nisar, "Routing-Based Interference Mitigation in SDN Enabled Beyond 5G Communication Networks: A Comprehensive Survey," IEEE Access, vol. 11, pp. 4023-4041, 2023.

Authors Introduction

Wen-Shyang Hwang

He received B.S., M.S., and Ph.D. degrees in Electrical Engineering Department from National Cheng Kung University, Taiwan, in 1984, 1990, and 1996, respectively. He is currently a professor of Electrical Engineering Department of National Kaohsiung University of Science and Technology(NKUST). His current research interests are in the fields of multimedia wireless communication, wireless mesh networks, storage area networks, WDM metro-ring networks, software design for embedded systems, Internet QoS, and Internet applications.

Ming-Syuan Wu

He received M.S. degree (2021) from the Department of Electrical Engineering at NKUST. He is currently studying in the doctoral program at Department of Electrical Engineering of NKUST.

Cheng-Han Lin

He received his Ph.D. from the Department of Electrical Engineering at National Cheng Kung University in 2010 and his M.S. and B.S. degrees from the Department of Electrical Engineering at National Chung Cheng University in 2002 and 2004, respectively. He is currently an associate professor at the Department of Health-Business Administration at Fooyin University. His current research interests include wireless networks and SDN networks.

Yan-Jing Wu

She received a Ph.D. degree in computer engineering from the Department of Electrical Engineering, National Sun Yat-Sen University, Kaohsiung. She became an assistant professor in 2007, and was promoted to associate professor and full professor, respectively, in 2013 and 2023 at the Department of Information Technology and Communication, Shih Chien University, Kaohsiung Campus. Her current research interests are in the fields of mobile wireless communication networks and mobility management for multimedia traffic.

Ming-Hua Cheng

He received a Ph.D. degree in Electrical Engineering from NKUST in 2019. He has become an assistant professor since 2019 at the Department of Digital Media Design, Tzu-Hui Institute of Technology.

Sian-Fong Huang

He received M.S. degree (2024) from the Department of Electrical Engineering at the NKUST.

The Application of AI in the Real Estate Industry: Business Model Innovation Perspective

Li-Min Chuang
The Department of International Business, Chang Jung Christian University, Tainan City, 711301, Taiwan

Chih-Hung Chen
The Ph.D. Program in Business and Operations Management, College of Management, Chang Jung Christian University, Tainan City, 711301, Taiwan
Email: liming@gmail.com, macuro6793@gmail.com

Abstract

This paper mainly discusses how Taiwan's real estate agency system uses artificial intelligence (AI) to reshape the traditional business model and create business model innovation. This study uses the Fuzzy Analytical Hierarchy Process (FAHP) and the literature analysis method to construct an analytical framework of four major dimensions based on nine key elements of the business model. Through empirical analysis of relative weights, important propositions are established. The research object is targeted at Tainan, Taiwan. A survey of real estate agency companies and marketing businesses was conducted to show that AI technology drives real estate agency business model innovation, and based on consumer business behavior, it accurately predicts market trends and improves marketing performance, thereby enhancing competitive advantages.

Keywords : Real estate industry, AI, Fuzzy Analytic Hierarchy Process (FAHP), Business Model Innovation

1. Introduction

In recent years, in response to the continuous development and evolution of AI (Artificial Intelligence) artificial intelligence technology, various industries are accepting the challenge of digital transformation, and of course the real estate agency industry (real estate agency) is no exception.

Through the application of AI technology, big data analysis can be used to quickly match customer needs with potential buyers, or virtual tour technology can be used to quickly improve user experience. These innovative ideas and methods have brought great benefits to the real estate industry.

This study focuses on the application of AI in the real estate industry and explores its impact on the industry from the perspective of business model innovation, with a view to providing inspiration and reference for the industry. The purpose of this study is:

(1) Through various previous literature discussions, collect and sort out the key factors of business service innovation, and develop "main dimensions" and "secondary dimensions".

(2) Integrate AI artificial intelligence technology through facet analysis and use the fuzzy analytic hierarchy process (FAHP) to carry out a survey in order to determine the relative weights of "main facets" and "secondary facets".

(3) Based on the research conclusions, analyze the key factors of how AI application promotes business model innovation in the real estate agency industry as a driving force for industry transformation, and explore possible future innovation directions as an important reference for achieving sustainable development.

2. Literature Review

The term "business model innovation" was coined by Amit and Zott[1], consisting of the two terms "innovation" and "business model", indicating the need to develop innovative business models.

Innovation involves the introduction of new ideas, methods or techniques to create or improve upon existing products, services or processes. Therefore, business model innovation is about rethinking this structure and finding new ways to create value for the company. This is not just a fine-tuning of existing business processes, but a fundamental shift in the company's core value proposition, revenue sources and cost structure to adapt to market changes and gain competitive advantage[2].

Business model innovation enables companies to swiftly respond to market changes and gain a deeper understanding of consumer needs. Enterprises can innovate across dimensions such as customer value, value chain, value network, and expanded business models, ultimately offering improved services to customers through effective innovation. thereby strengthening the enterprise's market competitive advantage and providing a basis for the sustainable development of the enterprise. Certain guarantee[3].

Innovation in real estate business models arises not only within real estate companies but also through the interactions between real estate companies, building users, and other participants within the network. Real estate business model innovation should not be viewed as the result of actions taken solely by real estate companies or building users, but rather as the outcome of interactions among two or more stakeholders within a network[4] , [5] , [6].

Real estate companies need to adapt their mindset and actions early in the development process to create value-driven business models. The transition from providing products to enabling value-creating processes represents a shift in focus from the real estate company to the building user, highlighting the resources that support the processes the building user seeks to leverage. [7]

AI artificial intelligence can affect market development. By employing various artificial intelligence techniques, such as customer analysis, industry insights, and cross-functional integration, companies can gain a comprehensive understanding of consumer needs and apply these insights to enhance their marketing strategies.[8]. Digitalization has also become a key element in the business development function of real enterprises, and the issue of digitalization has also shown the improvement of innovation capabilities at the enterprise level [9].

This study collates domestic and foreign experts and scholars' literature on business model innovation and the application of AI artificial intelligence to real estate, and compiles and constructs the main dimensions of business model innovation, namely "value innovation", "value chain innovation", and "value network". The four major dimensions of "innovation" and "broadening business models" are described as follows:

At the level of "customer value innovation", in order to stand out in the fierce homogeneous market competition, we must pay close attention to changes in consumer needs, deeply explore their potential needs, and carry out innovative interpretations and redefinitions; only through this Only through accurate insight into demand and innovation can enterprises gain an advantageous position in the competition. At the same time, companies must redefine and segment their customer base, and explore new value propositions based on shifts in customer needs. This enables them to better and more quickly respond to changes in consumer demands. Throughout this process, companies should conduct thorough research and analysis of market consumer needs, while establishing an effective communication mechanism between the company and its customers to ensure that the business model undergoes fundamental innovation [3].

At the level of "value chain innovation," the core idea is to optimize internal resource allocation within the company based on consumer needs, ensuring resources are fully utilized and cost advantages are maximized. The fundamental concept involves identifying the key elements centered around consumer needs, strategically combining and adjusting the critical components, and aligning the less important parts to support the central elements. This approach strengthens the leading role of the crucial links while enabling the secondary links to play a more supportive role, ultimately achieving a more effective resource allocation within the value chain[3].

At the level of "value network innovation," the essence lies in focusing on customer value, which allows for the effective optimization of both internal and external resource allocation within the company. This approach drives continuous innovation across the entire industry chain, enabling the company to achieve dynamic market development. Companies can strategically reconstruct the supply chain structure to optimize resources, enhance collaborative relationships, and ensure the supply chain can adapt flexibly to market changes. This type of innovation involves simplifying supply chain processes based on consumer needs, thereby improving the relationship between the company and its supply chain partners [3].

At the level of "broaden business models," it is evident that, in a sustainable market economy, business competition is intensifying. As a result, companies must consistently innovate their business models throughout their operations. In the course of development, companies need to regularly assess and evaluate their own situations, identifying their strengths in business development and competitive dimensions. Simultaneously, it is crucial to gain a comprehensive understanding of the customer base distribution and clearly define the company's business development positioning[3].

Regarding real estate marketing methods, with the increasing development of information technology, it is a basic skill for real estate agents to use technology and digital marketing methods, and more real estate agents are beginning to use AI technology to actively promote sales activities. Facing the wide range of demands in the real estate market, it can strengthen organizational operations and improve corporate performance.

3. The construction of AI in the real estate agency business model innovation perspective

3.1 Research Framework

This study combines the discussion results of the Focus Group to construct a definite hierarchical structure of AI business model innovation in the real estate industry, consisting of four primary dimensions and twelve evaluation criteria, as shown in (Fig.1).

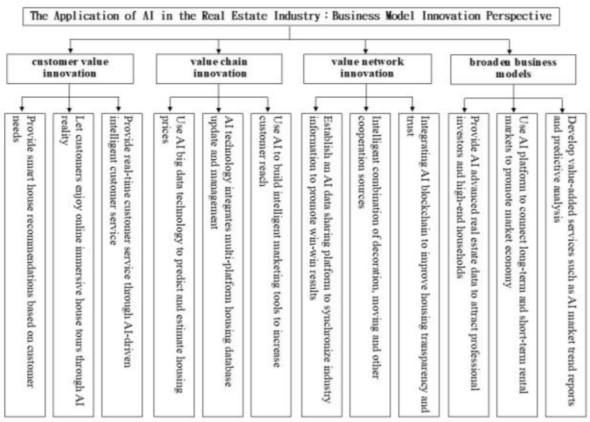


Fig. 1 Research Framework

3.2 *Research Subjects*

This study takes the real estate industry players and scholars in Tainan, Taiwan as the research object, and aims at researching the real estate industry players' perspectives on the use of AI technology and business model innovation. Using purposive sampling, 12 scholars and industry experts with professional expertise in AI, the real estate industry, and other relevant fields were selected as participants for the FAHP questionnaire. In addition, before conducting the survey in this study, the researcher will first explain the purpose of this study to the questionnaire respondents, and conduct the FAHP questionnaire survey to the real estate industry personnel who are willing to assist in the questionnaire survey.

4. Research Result

This study further multiplies the four evaluation indicators with their respective dimensions, and the resulting value is the comprehensive weight. Table 1 shows the comprehensive weight of AI business model innovation.

Table 1 Comprehensive weight of AI business model innovation

Dimension	Weights	Evaluation indicators	Weights	Comprehensive weight
customer value innovation	0.276	Provide smart house recommendations based on customer needs	0.486	0.134
		Let customers enjoy online immersive house tours through AI reality	0.212	0.059
		Provide real-time customer service through AI-driven intelligent customer service	0.302	0.083
value chain innovation	0.368	Use AI big data technology to predict and estimate housing prices	0.437	0.161
		AI technology integrates multi-platform housing database update and management	0.269	0.099
		Use AI to build intelligent marketing tools to increase customer reach	0.294	0.108
value network innovation	0.195	Establish an AI data sharing platform to synchronize industry information to promote win-win results	0.146	0.028
		Intelligent combination of decoration, moving and other cooperation sources	0.336	0.066
		Integrating AI blockchain to improve housing transparency and trust	0.518	0.101
Broaden business models	0.161	Provide AI advanced real estate data to attract professional investors and high-end households	0.345	0.056
		Use AI platform to connect long-term and short-term rental markets to promote market economy	0.162	0.026
		Develop value-added services such as AI market trend reports and predictive analysis	0.493	0.079

There are 12 real estate industry experts believe that the top five most important evaluation indicators are "Use AI big data technology to predict and estimate housing price " (the comprehensive weight is 0.161), and the second-ranked one is "Provide smart house Recommendations based on customer needs " (the comprehensive weight is 0.134), the third ranking is "Use AI to build intelligent marketing tools to increase customer reach" (the comprehensive weight is 0.108), the fourth ranking is "Integrating AI blockchain to improve housing transparency and trust" (the comprehensive weight is 0.101), and the fifth ranking is "AI technology integrated multi-platform housing database update and management" (the comprehensive weight is 0.099).

Furthermore, the evaluation indicators placed between sixth and twelfth are: "Providing real-time customer service through AI-driven intelligent customer service" (the comprehensive weight is 0.083), "Developing value-added services such us AI market trend reports and predictive analysis" (comprehensive The weight is 0.079), "Intelligent combination of decoration , moving and other cooperation sources" (the comprehensive weight is 0.066), "Let customers enjoy online immersive house tours through AI reality" (the comprehensive weight is 0.059), "Providing AI advanced Real estate data to attracts professional investors and high-end households" (the comprehensive weight is 0.056), "Establishes an AI data sharing platform to synchronize industry information to promote win-win results" (the comprehensive weight is 0.028), and "Uses AI platforms to connect the long-term and short-term rental markets to promote market economy" (The comprehensive weight is 0.026).

5. Conclusion and recommendations

This study summarizes important findings on business model innovation and summarizes the following four conclusions:

5.1 *The key evaluation indicator of customer value innovation is "Providing smart house recommendations based on customer needs"*

The key to customer value innovation lies in understanding and meeting customer needs, and "providing smart housing recommendations based on customer needs" has become an important evaluation indicator, reflecting the value of a customer-centric strategy. Smart house recommendation relies on data analysis and artificial intelligence technology to accurately identify customers' house purchase preferences and improve service efficiency and accuracy. This personalized recommendation method improves customer satisfaction and loyalty through in-depth exploration of needs. In addition, smart recommendations also demonstrate the ability of technology to be used in business innovation, becoming an important means of differentiated competition for real estate companies and one of the core standards for measuring customer value innovation.

5.2 *The key evaluation indicator of value chain innovation is "Use AI big data technology to predict and estimate housing prices."*

Value chain innovation mainly improves corporate operational efficiency and value creation, and "using AI big data technology for housing price prediction and estimation" as an important evaluation indicator highlights the profound impact of technology on value chain optimization. AI and big data can quickly process massive amounts of information and provide accurate housing price predictions, helping decision makers quickly respond to market changes. House price estimation improves transaction transparency and efficiency, helps reduce the risk of information asymmetry, and increases customer trust. At the same time, the application of this technology can reduce valuation labor costs and optimize resource allocation. The use of AI big data technology not only strengthens the core competitiveness of the value chain, but also drives the real estate industry towards an intelligent and data-driven future.

5.3 *The key evaluation indicator of value network innovation is "Integrating AI blockchain to improve housing transparency and trust"*

Value network innovation emphasizes the creation of shared value in multi-party cooperation, and "integrating AI blockchain to improve housing transparency and trust" as an evaluation indicator demonstrates the transformative effect of technology integration on the value network. AI technology can quickly analyze market data, while blockchain ensures that transaction data is transparent and cannot be tampered with. The combination of the two effectively reduces the risk of information asymmetry. Specifically, AI helps with risk assessment and market analysis, providing a basis for multi-party decision-making. In addition, it not only improves the efficiency and security of real estate transactions, but also promotes collaboration among members of the value network to achieve benefit sharing and trust growth, becoming the driving force for sustainable development.

5.4 *The key evaluation indicator of broaden business models is "Develop value-added services such as AI market trend reports and predictive analysis"*

The core of broadening the business model is to create new value and new revenue sources, and "developing AI market trend reports, predictive analysis and other value-added services" as an evaluation indicator reflects the innovation needs in the data-driven era. AI market trend reports and predictive analysis can provide accurate market insights, help enterprises and customers grasp industry trends and risks in advance, and help improve the quality of decision-making. This type of value-added services expands the boundaries of traditional business models, not only increases revenue sources, but also enhances brand value and competitiveness. In addition, when a company can strengthen its connection with customers, it can increase customer stickiness and loyalty, laying the foundation for the company to achieve long-term sustainable development.

5.5 *Recommendations*

Future research suggests that we can focus on AI smart recommendations and adaptability to customer needs, the transparency of AI housing price forecasts and response to market dynamics, the integration of AI and blockchain technology, and the expansion of value-added services, etc., and analyze market trends and business through AI. The value report, as well as the potential for cross-industry technology integration and cooperation, combined with empirical cases to verify the actual effect of innovation indicators, will help promote the theoretical promotion and practical progress of business models.

References

1. Amit, R., & Zott, C. (2012). Creating value through business model innovation. MIT Sloan Management Review, 53(3), 41–49. Retrieved March 22, 2023.
2. Bangjie Yin(2024), Business Models for Enterprise In novation and Development, Journal of Global Econom y, Business and Finance (JGEBF) ISSN: 2141-5595.
3. Li Li(2023), Research on the Innovation of Business Model of Enterprises,Clausius Scientific Press Journals Books Proceedings Vol 4, Issue 3, 2023.
4. Frenz, M. and Ietto-Gillies, G. (2009), "The impact o n innovation performance of different sources of kno wledge: evidence from the UK community innovation survey", Research Policy, Vol. 38 No. 7,pp. 1125-11 35.
5. Hakansson, H. (1985), Industrial Technological Develo pment. A Network Approach, Croom Helmt, London.
6. Chesbrough, H. and Crowther, A.K. (2006), "Beyond high tech: early adopters of open innovation in other industries", R&D Management, Vol. 36 No. 3, pp. 2 29-236.
7. Yderfält, Å. and Roxenhall, T. (2017), "Real estate bus iness model innovation and the impact of ego networ k structure", Management Research Review, Vol. 40 N o. 6, pp. 648-670.
8. Akter S.et al.(2023), "A framework for AI-powered s ervice innovation capability: review and agenda for fu ture research. Technovation ". Technovation Volume 1 25, July 2023, 102768.
9. Winter, S.G. (2003), "Understanding dynamic capabilit ies", Strategic Management Journal, Vol. 24 No. 10, pp. 991-995.

Authors Introduction

Dr. Li-Min Chuang

He received the Ph.D. degree from the National Cheng Kung University. He is currently an associate professor and department head at the Department of International Business at Chang Jung Christian University, Taiwan. His research interests focus on strategic management, innovation management, and service innovation.

Mr. Chih-Hung Chen

He received his Master's degree from the EMBA, Chang Jung Christian University, Taiwan in 2013. He is currently a Doctoral course student in Chang Jung Christian University, Taiwan.His research interests include the application of information technology to real estate industry.

The Impact of AI-Powered Health Monitoring on the Quality of Life and Social Participation of the Elderly: Technology Acceptance Model Perspective

Li-Min Chuang

The Department of International Business, Chang Jung Christian University, Tainan City, 711301, Taiwan

Zong-Sheng Li

The Ph.D. Program in Business and Operations Management, College of Management, Chang Jung Christian University, Tainan City, 711301, Taiwan
Email: liming@gmail.com, z66543d@gmail.com

Abstract

This article mainly explores the use of AI intelligent health monitoring by elderly in Taiwan and uses the technology acceptance model. It uses literature analysis and fuzzy analytic hierarchy process as research methods to conduct a FAHP questionnaire survey among elderly in Tainan, Taiwan. The theory considers perceived ease of use, Perceived usefulness, attitude and behavioral intention, and through calculation of the relative weights between various dimensions, the study found that after the introduction of AI intelligent health monitoring, elderly can not only effectively obtain personalized health advice, but also because it is easy to operate, making the elderly think that AI intelligent health monitoring is helpful. In addition, with the support of the government and family members, it will be easier for the elderly to accept AI intelligent health monitoring, effectively manage their health, improve their quality of life, and actively participate in social activities.

Keywords：AI-Powered Health Monitoring, Elderly, TAM, Quality of Life, Social Participation

1. Introduction

In recent years, artificial intelligence (AI) has been able to overcome the huge intensive computing technology and creative limitations required by humans due to the advent of big data, and open up new application areas in the fields of education, marketing, healthcare, finance and manufacturing, thus affecting productivity and performance. have an impact [1]. However, the acceptance of emerging technologies by elderly remains to be explored. This study takes the elderly as the research object of AI intelligent health monitoring, and explores the acceptance of the use of AI intelligent health monitoring by the elderly, so as to further achieve independent health management, improve the quality of life, and actively participate in social activities. The research will use the technology acceptance model and the combined hierarchical analysis method to analyze the acceptance of AI intelligent health monitoring by the elderly and identify the influencing factors. The results of this study will help to understand the attitudes and needs of the elderly towards AI smart health monitoring, and serve as a reference for the government's subsequent promotion of smart elderly policies and the development of monitoring equipment or technology by related industries, so as to promote the wisdom of health care for the elderly. development and health and well-being.

2. Literature Review

An aging society has arrived, and the elderly have increasing needs for chronic diseases and daily monitoring.

Artificial intelligence (AI) can effectively use predictive analysis to identify high-risk individuals with diseases such as diabetes, heart disease, and dementia. Medical providers can take appropriate care measures in a timely and specific manner [2], so AI intelligent health monitoring technology has become quite important. AI intelligent health monitoring has a wide range of applications, including physiological data monitoring (such as heart rate, blood pressure, blood oxygen, etc.), health advice generation, remote medical support and health assessment, which can effectively help elderly identify health risks early and take preventive measures, tailor-made exclusive exercise prescriptions for sarcopenia, frailty, etc. to reduce the risk of accidents and ensure the safety of daily life.

In addition, high demand for care may lead to social isolation of elderly [3]. Therefore, interaction between family members, professionals (such as fitness coaches) or caregivers is very important for elderly' acceptance of new technologies. Through social support and community internet can enable elderly to connect with people with similar experiences, thereby increasing their sense of social belonging [4], prompting elderly to continue socializing and participating in various activities, and reducing loneliness . Therefore, Tainan City, Taiwan, will launch a "movement campaign" in 2022 "Technology Application and Industrial Development" project combines social support with technological sports models to introduce wearable devices and sports equipment to 9 elderly fitness clubs across the city. This not only improves the convenience of elderly using AI smart health monitoring, but also inspires social participation motivations of elderly.

This study uses the Technology Acceptance Model (TAM) as the research framework, which was proposed by Davis

(1986) as a theoretical model to explain individuals' acceptance of information technology. The main core of TAM is that an individual's acceptance of technology is affected by two factors: perceived usefulness and perceived ease of use. These two factors in turn affect an individual's attitude towards technology, and lead to an individual's behavioral intention and final actual use [5].

AI intelligent health monitoring allows elderly to continue to improve their health and quality of life through health management monitoring and exercise training recommendations. It is also actively recommended to peers due to the promotion of government policies and their own benefits, allowing elderly to increase their social participation. Sense of participation [6]. However, overly complex operations or poorly thought out designs may lead to psychological rejection among elderly, affecting their acceptance and willingness to use them. Therefore, simplifying functions such as operating interfaces can help elderly accept new technologies more easily [7].

Finally, pursuing a healthy life is not something that can be achieved overnight [8]. However, as the elderly gradually accept AI intelligent health monitoring, and through continuous monitoring, reminders and suggestions, various health problems can be improved, and the health management capabilities of the elderly can be strengthened.

Reduce the need for medical care and actively participate in social activities to improve quality of life.

3. The construction of the acceptance of AI intelligent health monitoring among elderly

3.1. *Research Framework*

This study used a comprehensive focus group (Focus Group) to invite 20 scholars and industry experts with expertise in AI intelligent monitoring, health sports and elderly research experts. The discussion results are shown in Table 1 to determine the acceptance of AI intelligent health monitoring by elderly. The hierarchical structure includes five major dimensions and fifteen evaluation indicators, as shown in (Fig. 1).

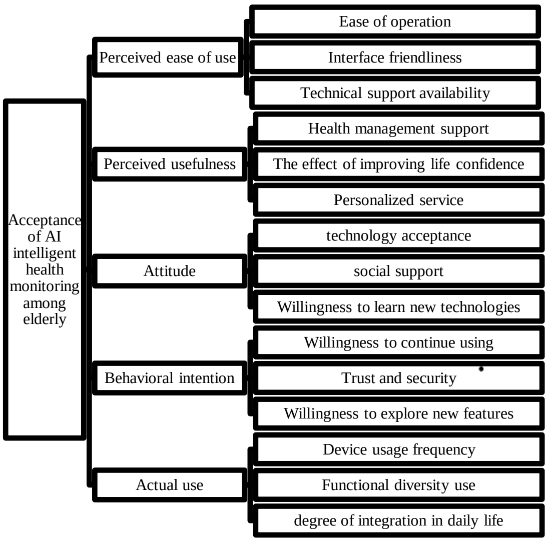


Fig 1 The hierarchical structure of the acceptance of AI smart health monitoring among elderly

Table 1. Comprehensive focus group discussion results

Facets	Facet Overview	Evaluation indicators	Assessment elements
Perceived ease of use	To explore the opinions of elderly on the difficulty of operating smart health monitoring technology.	Ease of operation	Evaluate how difficult it is for elderly to learn and operate AI smart health monitoring equipment and whether it is consistent with their daily habits and cognitive abilities.
		Interface friendliness	Check whether the design of the AI smart health monitoring device or application interface is simple, clear, and easy to understand and use.
		Technical support availability	Confirm whether elderly can receive real-time technical support when operating AI smart health monitoring.
Perceived usefulness	the extent to which elderly think smart health monitoring technology helps health management and quality of life	Health management support	In the process of health management, AI intelligent health monitoring provides various assistance, encouragement and resources to help individuals achieve and maintain health goals. .
		The effect of improving life confidence	Will the use of AI intelligent health monitoring increase the confidence of elderly in controlling their health, thereby improving their ability to live independently?
		Personalized service	AI intelligent health monitoring provides exclusive suggestions based on individual health needs, which can increase the willingness of elderly to use it and make them feel valued and cared for.
Attitude	Assessing the positive or negative attitudes of elderly towards smart health monitoring technology.	Technology acceptance	Do elderly actively accept AI smart health monitoring and consider it a useful resource in life?
		social support	The support of family and community can promote the positive attitude of elderly towards new technologies and encourage them to actively use AI intelligent health monitoring.
		Willingness to learn new technologies	Are elderly willing to invest time and energy in learning AI intelligent health monitoring and adapting to new health management technologies?
Behavioral intention	To explore the willingness and motivation of elderly to use smart health monitoring technology in the future.	Willingness to continue using	Are elderly willing to use AI smart health monitoring devices for a long time?
		Trust and security	Whether concerns about data privacy and security affect willingness to use.
		Willingness to explore new features	Are elderly willing to try the new features of their devices to further improve their quality of life?
Actual use	Analyze the actual use of smart health monitoring technology by elderly, including frequency of use and functional scope.	Device usage frequency	How often do elderly use AI smart health monitoring every day or every week, and whether they have formed stable usage habits.
		Functional diversity use	Do elderly use the various functions of AI intelligent health monitoring, such as physiological monitoring, emergency contact and health advice?
		degree of integration in daily life	Evaluate whether AI intelligent health monitoring has been integrated into daily life and provides obvious support for daily activities.

3.2. *Research Subjects*

In order to investigate the use of AI technology to integrate service innovation by Taiwanese elderly industry players, this paper takes the elderly who are using AI intelligent health monitoring in Tainan, Taiwan as the research object, and uses purposive sampling as the subjects of the FAHP questionnaire. In addition, before conducting the FAHP questionnaire survey in this paper, the researcher first explained the purpose of this paper to these FAHP questionnaire subjects, and conducted the FAHP questionnaire survey on the elderly who were willing to assist in the FAHP questionnaire survey.

4. Empirical Result of AI-Integrated Service Innovation Models

This paper further multiplies the five evaluation indicators with their respective dimensions, and the resulting value is the comprehensive weight.

Table 2 is the comprehensive weight of AI intelligent health monitoring acceptance.

Table 2. The comprehensive weight of AI intelligent health monitoring acceptance

Facets	weight	Evaluation indicators	weight	Comprehensive weight
perceived ease of use	0.209	Ease of operation	0.572	0.120
		Interface friendliness	0.263	0.055
		Technical support availability	0.165	0.034
perceived usefulness	0.302	Health management support	0.559	0.169
		The effect of improving life confidence	0.253	0.076
		Personalized service	0.188	0.057
Attitude	0.232	Technology acceptance	0.501	0.116
		Social support	0.315	0.073
		Willingness to learn new technologies	0.184	0.043
behavioral intention	0.141	Willingness to continue using	0.511	0.072
		Trust and security	0.307	0.043
		Willingness to explore new features	0.182	0.026
Actual use	0.116	Device usage frequency	0.472	0.055
		Functional diversity use	0.329	0.038
		Degree of integration in daily life	0.199	0.023

CI<0.1, CR<0.1

The most important evaluation indicators recognized by experts in AI intelligent detection, health sports and elderly research are "Supportability health management" (the comprehensive weight is 0.169) and "Ease of operation" (the comprehensive weight is 0.120 ranking Second), the third is " Technology acceptance " (the comprehensive weight is 0.116), the fourth is "The effect of improving life confidence" (the comprehensive weight is 0.076), and the fifth is "Social support" (the comprehensive weight is 0.073), and the evaluation indicators ranked sixth to fifteenth are: "willingness to continue using " (the comprehensive weight is 0.072), "Personalized service" (the comprehensive weight is 0.057), "Interface friendliness" (the comprehensive weight is 0.055), "Device usage frequency" (the comprehensive weight is 0.055), "Trust and security" (the comprehensive weight is 0.043), "Willingness to learn new technologies" (the comprehensive weight is 0.043), "Functional diversity use" (The comprehensive weight is 0.038), "Technical support availability" (the comprehensive weight is 0.034), "Willingness to explore new features " (the comprehensive weight is 0.026), and "Degree of integration in daily life " (the comprehensive weight is 0.023).

5. Conclusions and Recommendation

This study found that through the Technology Acceptance Model (TAM) and FAHP questionnaire survey, the acceptance of AI intelligent health monitoring among elderly can indeed effectively improve the quality of life and social participation. AI intelligent health monitoring can not only continuously monitor the physiological conditions of elderly, but also provide various management suggestions based on different health conditions, so that elderly are willing to improve their physical health, enhance their quality of life and participate in social activities. However, in order to benefit more elderly, strategies to promote and increase the acceptance of AI smart health monitoring should be strengthened in the future to further promote healthy aging.

Therefore, in order to promote the popularization of AI intelligent health monitoring, this study recommends starting from the following dimensions:

5.1 Health management support

Because the elderly can understand their physiological status in real time through AI intelligent health monitoring, and cooperate with their own personal health management plans, they can manage their health goals more accurately.

Therefore, health management support should focus on the accuracy of data analysis and immediate response to abnormal physiological phenomena to ensure that elderly can take preventive health measures, enhance their confidence and control over health management, reduce the number of medical visits, and improve Take care of yourself. For example, conduct pattern analysis based on historical data, set personalized warning ranges, and stimulate motivation to participate in health management by setting personalized daily step goals or diet plans.

5.2 Ease of operation

The ease of use of AI intelligent health monitoring technology is an important factor influencing the use of it by elderly. Research shows that when the operating interface is clear, simple and easy to operate, it can reduce the user's learning curve and make it easier to operate new technologies. Therefore, for elderly, a more intuitive user interface and appropriate technical support can effectively reduce their rejection of AI smart health monitoring and increase their intention to use it. For example, set up a voice assistance system that allows users to operate in natural language or present it with clear icons (such as a heart-shaped icon representing heart rate data) to simplify text descriptions.

5.3 Technology acceptance

According to the Technology Acceptance Model (TAM), perceived usefulness and perceived ease of use are the core of technology acceptance. Therefore, the acceptance of AI smart health monitoring by elderly is closely related to their understanding of the system’s practicality and ease of operation. In addition, in recent years, the elderly have paid more and more attention to health management, especially chronic disease management and daily health monitoring.

Therefore, providing accurate real-time data and corresponding treatment suggestions through AI intelligent health monitoring can make the elderly more accepting of AI intelligence. Health monitoring and continuous health management for yourself. For example, a simple mode is set up to gradually guide elderly to become familiar with core functions and provide various teachings. In addition, interactive lectures and demonstrations are used to showcase technologies on how to improve health management.

5.4 The effect of improving life confidence

As age increases, various health functions gradually decline. For elderly, effective AI intelligent health monitoring technology can reduce elderly' anxiety about their health and promote self-confidence. Establish.

Through the complete analysis and positive feedback of AI intelligent health monitoring data, in addition to creating personalized intelligent health promotion plans, it can also enable elderly to better understand their physical conditions and maintain autonomy and enthusiasm in their daily lives. For example, set small goals and feedback, set achievable health goals in stages, and provide positive feedback or use AI to simulate professional health consultants to provide real-time answers and suggestions.

5.5 Social support

Through the promotion of AI smart health monitoring related devices and government and community policies, elderly can be encouraged to participate in online and offline community activities. In addition to reducing the loneliness caused by aging, they can also gain support from their peers. participation, thereby enhancing the sense of belonging of the elderly. For example, design activities such as "daily step challenge" or "walking competition", where participants can track their progress and compare with others, or hold special lectures to provide knowledge on chronic disease management, exercise guidance, etc., and set up real-time interactive Q&A sessions.

6. References

1. Y.K. Dwivedi et al. (2021) “Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy ” International Journal of Information Management 57 (2021) 101994.
2. S.-H. Tan et al. (2024) “Informal caregivers' perception of assistive robots in eldercare ” Journal of Open Innovation: Technology, Market, and Complexity 10 (2024) 100234.
3. W. Badawy, M. Shaban (2025) “Exploring geriatric nurses' perspectives on the adoption of AI in elderly care a qualitative study ” GeriatricNursing61(2025)41-49.
4. Sestino and A. D'Angelo (2024) “Elderly patients' reactions to gamification-based digital therapeutics (DTx): The relevance of socialization tendency seeking ” Technological Forecasting & Social Change 205 (2024) 123526.
5. Fred D. Davis, Richard P. Bagozzi & Paul R. Warshaw (1989), “User acceptance of computer technology: A Comparison of two theoretical models, ” Management Science, Vol. 35, No. 8, pp. 982-997.
6. Kokila et al. (2024) “Determinants of Adoption of Mobile Health Applications: A Machine Learning Approach ” Procedia Computer Science 235 (2024) 1568–1576.
7. W. Guo et al. (2022) “Does modified mobile government satisfy elders' needs? An empirical study of China's Zhejiang and Jiangxi provinces ” Government Information Quarterly 39 (2022) 101676.
8. M. Yan et al. (2021) “I Mobile apps for healthy living: Factors influencing continuance intention for health apps ” Technological Forecasting & Social Change 166 (2021) 120644.

Authors Introduction

Dr. Li-Min Chuang

He received the Ph.D. degree from the National Cheng Kung University. He is currently an associate professor and department head at the Department of International Business at Chang Jung Christian University, Taiwan. His research interests focus on strategic management, service innovation, and innovation management.

Mr. Zong-Sheng Li

He received his Master's degree from the MBA, Kun Shan University, Taiwan in 2018. He is currently a Doctoral course student in Chang Jung Christian University, Taiwan. His research interests include the application of information technology to healthy exercise industry.

Reliability Analysis and Optimization of Distribution Network with Distributed Generation

Peng Wang
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222， China

Mengyuan Hu
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222， China
E-mail: autowangpeng@tust.edu.cn,1090465299@qq.com

Abstract

With the access of distributed generation, the distribution network is facing the problem of voltage quality degradation and cost increase. In this paper, Markov Monte Carlo simulation method (MCMC) is used to verify the influence of distributed generation access on distribution network. Secondly, the model based on particle swarm optimization (PSO) and K-means clustering algorithm is used to minimize the cost. Finally, the improved IEEE-33 node system is used to verify the example.

Keywords: Reliability, Distribution network optimization, Energy management, PSO

1. Background

Due to the increasing scarcity of energy and the increasing demand for electricity, the penetration rate of distributed energy in the distribution network has increased significantly, reducing the dependence on fossil fuels and environmental pollution. However, this trend also brings challenges such as unstable voltage quality and rising costs. As an important part of the power system, the stability and economy of the distribution network directly affect the reliability of the power supply. Therefore, the research on the influence and planning of distributed generation on distribution network has become an important topic in power system research.

In this paper, the MCMC method was used to simulate the normal operation and fault state of the distribution network to verify the influence of the distributed power supply on the distribution network. Secondly, combined with PSO and the K-means clustering algorithm, an operation-planning joint two-layer configuration model was established: the upper layer was the location and capacity model of photovoltaic and energy storage, and the lower layer considered the optimal scheduling of abandoned light and energy storage output. Taking the IEEE33 node as an example, the particle swarm optimization algorithm was used to solve the multi-objective model of the operating cost and voltage offset of the lower layer, and the Pareto front solution set was obtained by the multi-objective particle swarm optimization algorithm. The best result was brought into the upper model to realize the solution and iterative optimization of the upper and lower models. The model not only considered the impact of distributed generation (DG) on the distribution network but also improved the economy by optimizing the allocation of resources, provided a scientific basis for power system decision-making, and promoted the efficient use of renewable energy. In summary, this paper aimed to provide theoretical support and practical guidance for the access of distributed generation through the analysis of the distribution network.

The rest of this article was organized as follows. The second section introduced the photovoltaic power generation system and energy storage system. In the third part, the algorithm principles of Markov Monte Carlo, K-means, and PSO were discussed. In the fourth section, two models were constructed. The fifth section provided two simulation examples to verify the availability of the designed model. The sixth part summarized the main content of this paper.

2. Systematic Introduction

2.1. Photovoltaic power generation system

Photovoltaic power generation uses photovoltaic effect. Semiconductor materials excite photons to generate holes under light, and convert solar radiation energy into low-voltage direct current. Subsequently, through the converter and DC/AC inverter, the low-voltage DC is converted into AC power suitable for load use, or transmitted to the appropriate voltage level of the power grid or equipment. The power conversion process of the grid-connected photovoltaic power generation system is shown in Fig.1.

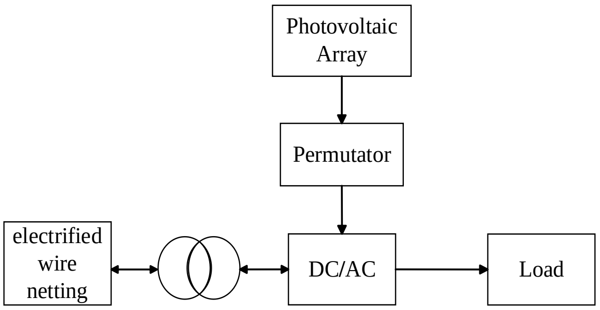


Fig.1 Grid-connected pv system

Multiple photovoltaic modules are usually combined to form a complete distributed photovoltaic power supply. In order to improve the output power of the system, multiple photovoltaic cell units can be combined into larger photovoltaic modules in series or in parallel through inverters. The schematic diagram of the structure is shown in Fig.2.In this system, the power generation performance and equivalent circuit model of photovoltaic cells can be expressed as Eq.(1).

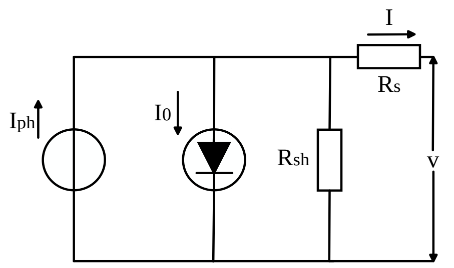


Fig.2 Solar photovoltaic cell equivalent circuit

$$I = I_{ph} - I_0\left\{\exp\left[\frac{q}{BKT}(V + IR_s)\right] - 1\right\} - \frac{V + IR_s}{R_{sh}} \quad (1)$$

Where V is the output current and output voltage;I_{ph} and I_0 are the photogenerated current and reverse saturation current; q is the charge constant; K is the Pultzman constant; T is the absolute temperature; B is the diode factor; R_s and R_{sh} are the series and parallel equivalent resistances.

2.2. Distributed energy storage system

The capacity of distributed energy storage system usually does not exceed 10 MWh, and the output power is between several kilowatts and several megawatts, which is mainly used in the position near the load end. The energy storage system stores excess power when the distributed generation exceeds the load and releases it at high load. At the same time, the energy storage system can quickly respond to power fluctuations, accurately adjust current and voltage, effectively suppress load fluctuations, and solve the problem of peak load and power fluctuations in the power grid[1].

The state of charge (SOC) of the energy storage battery, as a key indicator to measure the energy utilization efficiency of the energy storage system, can directly reflect the current remaining energy of the battery. That is, in the case of following specific charging and discharging rules, the ratio between the actual stored energy inside the battery and its rated capacity after charging and discharging cycles is shown as Eq. (2).

$$SOC = \frac{E(t)}{E_r} \quad (2)$$

The SOC is the value of [0,1], E(t) is the remaining capacity of the energy storage system at time t, and Er is the rated capacity of the energy storage[2].

3. Algorithm Principle

3.1. Markov monte carlo simulation (MCMC)

MCMC can effectively capture the complex randomness and uncertainty in the system. By comparing the performance of the power system before and after the access to the distributed power supply, the specific impact of the distributed power supply on the operation of the distribution network is revealed. MCMC obtains samples from complex probability distributions by constructing a Markov chain and using random sampling, and then performs statistical inference. The detailed steps of the simulation loop are as follows :

- Initialize and read data : set basic parameters, read line and load data.
- Random fault simulation:Randomly select a component in the power grid to fail. According to the characteristics of the line and the set distribution, time to failure(TTF) and time to repair(TTR) are generated. Two sets of Gamma distribution parameters are used to determine the service time or fault repair time.

Among them, the TTF adopts a mixed distribution model, including Pareto distribution and exponential distribution are expressed as Eq.(3).

$$TTF = \begin{cases} gprnd(k,\sigma,0), rand < pf \\ -\frac{1}{\Lambda|i|}\log(rand), others \end{cases} \quad (3)$$

Where gprnd is the fault time to generate Pareto distribution, k is the shape parameter, σ is the scale parameter, Λ is the failure rate and i is the failure rate determined by the line length and failure rate.

TTR was calculated by exponential distribution model,expressed as Eq. (4).

$$TTR = -\Gamma \log(rand) \quad (4)$$

Γ depends on the repair time parameter of the line.

- Update Fidelity Time:Update the total simulation time to include the time of fault occurrence and repair. Record fault events and evaluate the load affected by the fault.
- Network reconstruction:According to the fault location and DG location, the network configuration and load power supply status are updated. When the DG is set to island mode, the system implements a reconfiguration strategy by calling the reconfigure _ network function to optimize load distribution and restore power supply. The function removes the fault line and uses the remaining DG to find the optimal path for power supply. Subsequently, the update _ network _ load _ status function is used to update the power supply status of each load, reflecting the impact of the new grid structure on the load power supply, and confirming whether each load point is covered by DG and getting power supply. During fault recovery, according to the real-time power and position generated by DG, the load point that DG can restore power supply is calculated, and the shortest path algorithm is used to check the influence of fault on the load point covered by DG.
- Calculation of reliability index:Calculate the reliability index of each power system, and analyze the specific impact of faults on network performance, such as power outage frequency and duration.

3.2. Particle swarm optimization(PSO)

The particle swarm optimization algorithm simulates the behavior of birds flying for food, and achieves the optimal purpose through the collective cooperation between birds. In the PSO system, each alternative solution is called a particle, and multiple particles coexist and cooperate to optimize. Each particle flies to a better position in the problem space according to its own experience and the best experience of the particle swarm to search for the optimal solution. The PSO algorithm is mathematically expressed as follows: Let the search space be D-dimensional, and the total number of particles is n. The vector $X_i=(x_{i1},x_{i2},\dots,x_{iD})$ is the position of the i th particle ; $P_i=(P_{i1}, P_{i2},\dots, P_{iD})$ is the optimal position in the flight history of the i th particle (that is, the position corresponds to the optimal solution), where the historical optimal position P_g of the g th particle is the optimal position in all $P_i(i=1,\dots,n)$; the vector $V_i=(V_{i1},V_{i2},\dots,V_{iD})$ is the position change rate of the i th particle.

The particle velocity update formula is expressed as Eq.(5).The particle position update formula is expressed as Eq.(6).

$$V_{i+1} = w \times v_i + c_1 \times r_1 \times [P_{best} - x_i] + c_2 \times r_2 \times [g_{best} - x_i] \quad (5)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (6)$$

Where c_1,c_2 are two learning factors, r_1,r_2 are random numbers, w is an inertial factor. Larger w is suitable for large-scale exploration, and smaller w is suitable for small-scale excavation. x_i is the current particle position, v_i is the current particle velocity, x_{i+1} is the updated particle position, v_{i+1} is the updated particle velocity.The initial position and velocity of the particle swarm are randomly generated. In each iteration, the particle updates the speed and position by tracking two extreme values:the individual extreme value (Pbest) is the optimal solution of the particle itself, and the global extreme value (gbest) is the optimal solution of the group[3].

3.3. K-means

The K-means algorithm divides the data into K different clusters by iteration, and minimizes the sum of the distances between each data point and the centroid of its cluster. The execution process of the K-means algorithm usually includes the following steps : First, randomly select K data points as the initial cluster centroid ; secondly, according to the distance between each data point and the centroid of each cluster, it is assigned to the nearest cluster. Then, the centroid of each cluster is recalculated, that is, the average value of all data points in the cluster is taken as the new centroid ; finally, repeat the above allocation and update steps until a certain termination condition is met.

4. Model Construction

4.1. Establishment of optimal model

The improved particle swarm optimization algorithm to solve the bi-level optimization model is shown in Fig.3.

- Input the distribution network parameters, and used K-means to process the annual historical data of photovoltaic.
- Initialize the particle position and velocity, the location and capacity of the planning layer flexibility resources, as the input of the running layer.

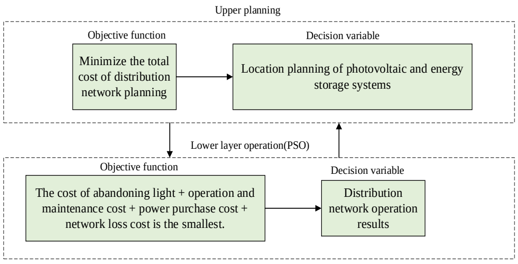


Fig.3 The upper and lower relationship of the joint model

This paper taked the system to reduce the investment cost as the optimization goal, considered the cost, benefit, environmental impact and related operating indicators.

The loan annuity factor u,investment costs,discarding penalty cost,Photovoltaic power generation,energy storage costs and total cost are as Eq.(7),Eq.(8),Eq.(9)，Eq.(10) , Eq.(11) and Eq.(12).

$$u=\frac{r(1+r)^{year}}{(1+r)^{year}-1} \quad (7)$$

$$C_{investment}=u\times(c_{pv}\times cap_{pv1}\times s_{pv}+c_{ess}\times cap_{cn}\times s_{cn}) \quad (8)$$

$$C_{loss}=c_{loss}\times \text{sum of power losses} \quad (9)$$

$$C_{pv}=cpv\times cap_{pv1} \quad (10)$$

$$C_{ess}=cess\times cap_{cn} \quad (11)$$

$$C_{total}=C_{investment}+C_{loss}+C_{pv}+C_{ess} \quad (12)$$

Where r is the discount rate, year is the service life of the equipment, c_{pv} is the investment cost of unit capacity photovoltaic, cap_{pv1} is the photovoltaic capacity, s_{pv} is the single photovoltaic capacity, c_{ess} is the investment cost of unit capacity energy storage, cap_{cn} is the energy storage capacity, s_{cn} is the rated capacity of a single group of energy storage, c_{loss} is the network loss price, and sum of power losses is the total power loss.

4.2. Constraint conditions

Abandoning light punishment, energy storage charge and discharge cost,power grid purchase cost,grid loss cost,and voltage deviation cost are as Eq.(13),Eq.(14)，Eq.(15),Eq.(16) and Eq.(17).

$$C_q=c_{qpv}\sum(gailv(i)\times(cap_{pv1}\times center(i,:)-pv1_s(i))) \quad (13)$$

$$C_y=cpvy\times\sum(gailv\times\sum(pv1_s))+cessy\times\sum(|cn|) \quad (14)$$

$$C_{gc}=\sum(C_{buy}\times P_{gen}) \quad (15)$$

$$C_{loss}=c_{loss}\times\sum(branch\_losses) \quad (16)$$

$$V_{pc}=\sum(|V_{bus}-V_{mean}|) \quad (17)$$

Where C_{qpv} is the penalty cost of discarding light, and gailv(i) is the scene probability. Cap_{pv1} is the photovoltaic capacity,center(i,:)is the scene center, $pv1_{s(i)}$ is the photovoltaic power output, cpvy is the unit capacity photovoltaic operation cost, gailv is the scene probability, cessy is the unit capacity energy storage operation cost, cn is the energy storage charge and discharge volume, C_{buy} is the main network purchase price, P_{gen} is the power generation.

5. Example Analysis

5.1. The impact of distributed power access to the distribution network is verified.

In this section, the IEEE Bus6 F4 test system is used as the system architecture diagram of the example, and on this basis, the improvement is made. The distributed power supply is set at 13 and 25 nodes. The power generation type is photovoltaic power generation, and the power generation power is 10 MW. The operation mode is island operation. The MCMC algorithm is used to analyze it. The simulation period is 20 years. The influence of DG and no DG on the distribution network is compared.

Through simulation, the system average interruption frequency index(SAIFI) and the system average interruption duration index (SAIDI) of the system with and without DG, the customer average interruption duration index (CAIDI) of the load, and the expected energy not served (EENS) are as Eq.(18),Eq.(19)，Eq.(20)，Eq.(21).The comparison results are shown in Fig.4.

$$SAIFI=\frac{\sum(load\_fault\times load\_users)}{\sum load\_users} \quad (18)$$

$$SAIDI=\frac{\sum(load\_fault\_time\times load\_users)}{\sum load\_users} \quad (19)$$

$$CAIDI=\frac{SAIDI}{SAIFI} \quad (20)$$

$$EENS=\frac{\sum(load\_power\times load\_users)}{\sum load\_user\times N} \quad (21)$$

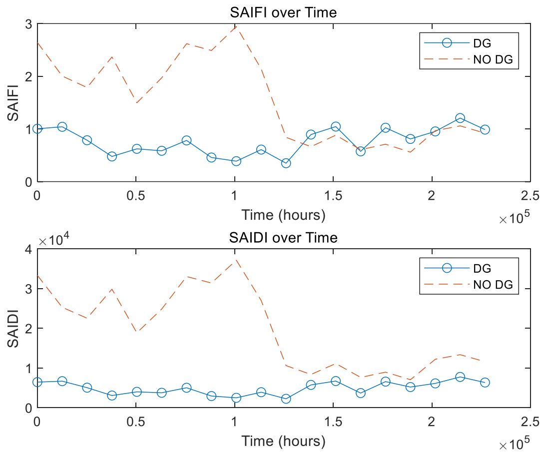


Fig.4 SAIFI and SAIDI with and without DG

It can be seen from the figure that although there are fluctuations at some time points, the average number of power outages and fault time of the system with DG are smaller than those without DG as a whole. The reliability indexes are obtained by comprehensive calculation,which is shown in Table 1.

Table 1 Reliability index of DG and no DG

Reliability index	DG	No DG
SAIFI	0.031693 times/ (year·household)	0.031741times/ (year·household)
SAIDI	1.0455hours/ (year·household)	1.5225hours/ (year·household)
CAIDI	38.807hours/ (year·household)	41.0893hours/ (year·household)
EENS	1.4106MW (year·household)	1.5166MW (year·household)
ASAI	99.9881%	99.9826%

Through simulation calculation, it is concluded that distributed generation has a positive impact on the reliability index of distribution network, which is manifested in the reduction of fault frequency, the shortening of fault recovery time, the reduction of power loss and the improvement of system availability. These results show that distributed generation plays an important role in improving the overall reliability of the distribution network.

5.2. Research on double-layer optimal configuration of photovoltaic and energy storage access to distribution network

This section used the IEEE-33 node distribution network model as a case to explore the optimal configuration strategy for photovoltaic power and energy storage. The topology of the system is shown in Fig.5. The voltage level is 10kV. The photovoltaic nodes that were selected were 2, 3, 4, 6, 8, 10, 12, 15, 18, 20, 26, 27, 29, 30, 31, and 32. The energy storage nodes that were selected were 3, 6, 9, 14, 18, 20, 28, 29, 31, and 33. The single photovoltaic capacity was 50 kW, the maximum number of photovoltaics was 15, and the minimum number of photovoltaics was 4. The rated capacity of a single group of energy storage was 10 kW, the maximum number of energy storage groups was 20, and the minimum number of energy storage groups was 8. The main network purchase price was 0.6 RMB, the network loss price was 0.4 RMB, the penalty cost of abandoning light was 0.6 RMB, the discount rate was 0.08, the service life of photovoltaic and energy storage equipment was 10 years, the investment cost of unit capacity energy storage was 8000 RMB, the operation cost of unit capacity energy storage was 0.05 RMB, the investment cost of unit capacity photovoltaic was 5000 RMB, and the operation cost of unit capacity photovoltaic was 0.1 RMB. The charging and discharging efficiency of energy storage was 90%, the minimum state of charge of energy storage equipment was 10%, and the maximum state of charge of energy storage equipment was 90%. When using the particle swarm optimization algorithm, the maximum number of iterations was 50 times, and the population size was 150.

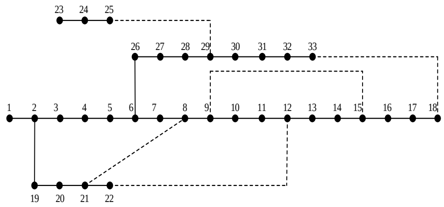


Fig.5 IEEE-33

Firstly, K-means clustering analysis is used to process the photovoltaic data, and the power generation characteristics of different time periods in a day can be classified into four scenarios, as shown in Fig.6.

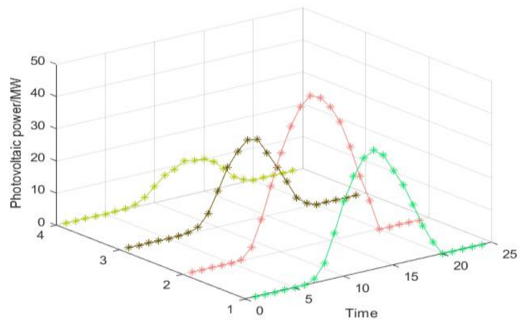


Fig.6 Photovoltaic cluster analysis results

The photovoltaic and energy storage location and capacity model is the upper model of the program, and its main goal is to realize the optimal configuration of photovoltaic and energy storage systems. Particle swarm optimization algorithm is used as the solution method. The optimal location and capacity configuration scheme of photovoltaic and energy storage system is obtained. Through the particle swarm optimization algorithm, we can get multiple pareto frontier solution sets, select the best results, and bring them into the lower model.

The optimal scheduling of light curtailment and energy storage output is the lower model of the program, which mainly considers the operating cost and voltage offset of photovoltaic and energy storage systems. In the model design, the multi-objective particle swarm optimization algorithm is used to solve the problem. The multi-objective particle swarm optimization algorithm can consider multiple optimization objectives at the same time, so as to obtain more feasible solutions, and can obtain the pareto frontier solution set by adjusting the weight. Through the multi-objective particle swarm optimization algorithm, the scheduling scheme of photovoltaic and energy storage systems is obtained, and compared with the best results in the upper model, so as to realize the iterative optimization of the whole model. The fitness curve of the algorithm,as shown in Fig.7.

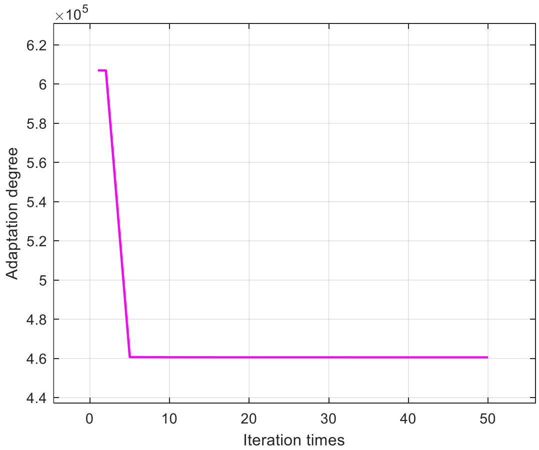


Fig.7 Algorithm fitness curve

This method not only considered the location and capacity configuration of photovoltaic and energy storage systems but also considered the optimal scheduling of light curtailment and energy storage output. Through the joint solution and iterative optimization of the upper and lower models, the optimal configuration scheme of the photovoltaic energy storage system was finally obtained, thereby realizing the flexible resource allocation and operation of the distribution network.

The load power supply is shown in Fig.8,It could be seen from the figure that the power had obvious fluctuations in different time periods, especially during the day and night. The power of photovoltaic power generation was higher during the day, while the power supply mainly depended on the power grid at night. The histogram of charging and discharging showed the operation of the energy storage system. Charging was usually carried out when the power demand was low, while discharging provided additional power support during peak demand. The power output of photovoltaic power generation during the day was significantly higher than that at night, reflecting the characteristics of solar power generation. With the change of sunshine intensity, the power of photovoltaic power generation also changed. The changes in the original load lines showed the system's demand for electricity, which usually peaked during the day and evening and was relatively low at night.

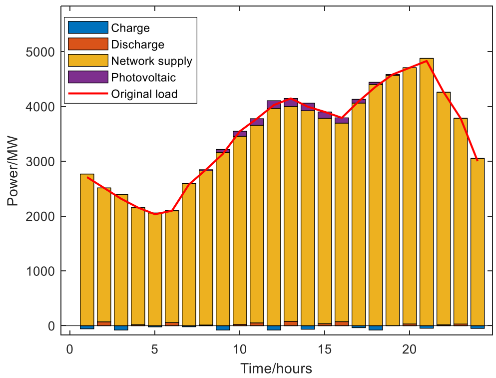


Fig.8 Load power supply situation

The results of equipment location planning are shown in Fig.9. The results show that the highest cost is 607039.544 RMB. After multiple iterations of the particle swarm optimization algorithm, the final cost of energy storage and photovoltaic is the lowest at 18 and 9 nodes, respectively. The cost is 460564.2716RMB, which is reduced by 24.16 %.

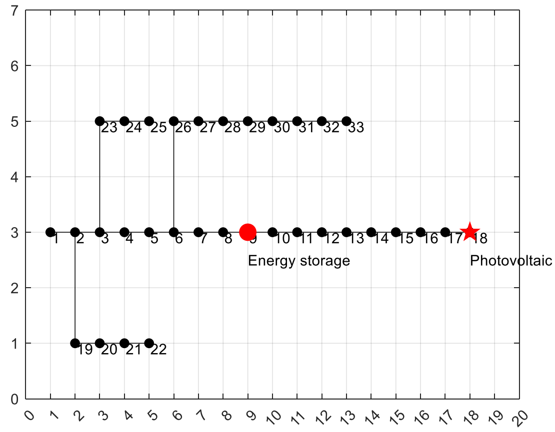


Fig.9 Equipment location planning results

6. Conclusion

This paper first used MCMC to verify the impact of distributed power access on the distribution network. The importance of optimization simulation in the power system to improve the efficiency and reliability of the power grid was expounded. The PSO and K-means clustering algorithms and their specific applications in research were introduced. Secondly, the optimization model was established with the minimum cost as the optimization objective. Finally, the test system (33-node system) was used for simulation analysis to obtain the optimal location of the photovoltaic distributed power supply and energy storage system.

Acknowledgements

The research is supported by the major special project of carbon neutralization technology in Tianjin, China in 2024 (24ZXTKSN00070).

References

1. MEHMET E,IBRAHIM A,GULFEM T. Acomprehensive review of the impacts of energy storageon power markets. Journal of Energy Storage 2024,91: 111935.
2. MEMARZADEH G,KEYNIA F.A new hybridCBSA-GA optimization method and MRMI-LSTMforecastingalgorithm for PV-EsS planning indistribution networks. *Journal of EnergyStorage2023*,72:108582.
3. Kennedy J,Eberhart R. *Particle swam optimization. Proceedings ofthe lEEE International Conference on Neural Networks*. Piscataway,NIEEE Press. Perth, Australia, 1995.

Authors Introduction

Ms. Peng Wang

She is a postgraduate tutor of Tianjin University of Science and Technology.In 2014, she received a doctorate from North China Electric Power University.The research direction is the functional safety assessment of safety instrumented systems.

Ms. Mengyuan Hu

In 2023, she received her Bachelor of Engineering degree from the School of Electronic Information and Automation, Tianjin University of Science and Technology, China.She is pursuing a master's degree in engineering from Tianjin University of Science and Technology.

Research on Improved PPLCNet Classification Network Based on CBAM Attention Mode

Peng Wang*

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

Shengfeng Wang

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

Qikun Wang

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

Yuting Zhou

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

*E-mail: *2324365941@qq.com, autowangpeng@tust.edu.cn, globeen@163.com,*

2994383067@qq.com

www.tust.edu.cn

(Tianjin University of Science and Technology, Tianjin, China)

Abstract

This paper studies pedestrian attribute recognition based on the pplcnet network because it is of great significance in the field of traffic security. Firstly, the research status of pedestrian attribute recognition and common deep learning models are introduced. Secondly, considering that convolutional block attention module contains both spatial attention module and channel attention module, we add this attention model to pplcnet to improve performance. Finally, this paper verifies the model through the pa100k dataset and obtains good results.

Keywords: Neural networks, Pedestrian attributes,CBAM, Deep learning

1. Introduction

In the current context of rapid technological advancement, the significance of attribute recognition technology is increasingly prominent [1]. It plays a crucial role in key sectors such as healthcare, public security, and intelligent furniture design. This has drawn extensive attention from academia and industry alike.With the deep integration of deep learning algorithms , vast amounts of data can be effectively processed. Computer systems can utilize these data via continuous self-learning.Notably, the person attribute recognition technology based on deep learning stands out. Through convolutional and pooling neural networks, it can precisely extract feature data from pedestrian images.Via complex classification procedures, it can acquire multi-dimensional pedestrian attribute information. The extraction of such attribute information enriches our understanding of individual characteristics. It also lays a practical foundation for applications in smart city construction and other crucial areas like military security. This promotes technological innovation and development transformation in related fields [2].

The rest of this article is organized as follows. The second section of the article introduces the deep separable convolutional network, pplcnet and data sets. In the third chapter, this paper compares the performance of the network by focusing on the location of the network. Finally, in Chapter 4, we summarize the full text.

2. Neural Networks and the Principle of Attention

In the field of computer vision, deep learning has become a basic tool. In this study, a large number of deep learning concepts will be used, and deep learning neural networks will be used to identify pedestrian attributes. Finally, results will be obtained in experiments.

2.1. Depthwise separable convolution

Depthwise Separable Convolution (DepthSepConv) is an efficient convolution operation method, which plays an important role in the field of deep learning. Especially in the architectures of deep convolutional neural networks.

Structurally, it consists of two crucial steps: depthwise convolution and pointwise convolution. In the depthwise convolution stage, for the input data. Such as a feature map with the shape of $H \times W \times C$ (where H represents the height, W represents the width, and C represents the number of channels). Each convolution kernel processes the data of a single channel independently.

In the pointwise convolution stage, a 1×1 convolution operation is used to fuse the features of each channel obtained in the depthwise convolution stage. This operation can change the number of output channels, enabling the network to adjust the dimension of the feature representation .

In terms of performance advantages, DepthSepConv performs excellently in terms of the amount of computation and the number of parameters. Compared with traditional convolution, its amount of computation and the number of parameters are significantly reduced. Taking traditional convolution as an example, the amount of computation is approximately $H \times W \times C \times K \times K \times N$, and the number of parameters is approximately $C \times K \times K \times N$. For DepthSepConv, the amount of computation is approximately $H \times W \times C \times (K \times K + N)$, and the number of parameters is approximately $C \times (K \times K + N)$.

2.2. Practical plain - based lightweight convolutional neural network

PPLCNet (Practical Plain-based Lightweight Convolutional Neural Network) is a lightweight convolutional neural network model that has drawn significant attention in the field of deep learning. It is designed to strike a balance between model accuracy and computational efficiency, while the storage requirements are minimized. This makes it highly suitable to be deployed in resource-constrained devices and applications with strict real-time constraints.

The network architecture of PPLCNet is characterized by several key features. A specialized depthwise convolution structure is employed. Compared with traditional convolutional layers, the number of parameters can be reduced by performing convolutions on each channel independently in the depthwise convolution and then integrating the channel information via pointwise convolutions (1×1 convolutions). In this way, the fine-grained local features, such as the texture and edge details in image data, can be extracted. Additionally, residual connections are incorporated into the model. The vanishing gradient problem can be effectively mitigated as the information is allowed to flow directly from shallower to deeper layers by these connections. More effective feature representations can be learned by the network during training through facilitating the backpropagation of gradients. In terms of performance, PPLCNet has distinct advantages. High computational efficiency is exhibited, and the data can be processed rapidly.

2.3. Convolutional block attention module

CBAM (Convolutional Block Attention Module) is an attention model in the field of deep learning.It is designed to adaptively adjust the feature weights.It has the importance of input features learned.So that the key features can be enhanced and the irrelevant ones can be weakened.Thus the overall performance of the neural network can be improved. It can be seamlessly embedded into various convolutional neural network architectures and is utilized to handle data with spatial structures such as images.

Its structure is composed of a channel attention module and a spatial attention module. The channel attention module is focused on the channel dimension of the feature map. Descriptors are obtained through global average pooling and max pooling being carried out, and then they are processed by a shared multilayer perceptron. After that, the channel attention weights are generated through the Sigmoid activation function being applied, and the channels are weighted accordingly. The spatial attention module is centered on the spatial dimension.A spatial attention map is generated through channel concatenation and convolution operations being performed.Then the weights are obtained through the Sigmoid activation function being activated.Finally, the spatial positions are weighted.

Based on the attention mechanism, during the forward propagation, CBAM is processed by these two modules in sequence. The key information in the channel and spatial dimensions can be focused on. Compared with traditional CNNs, a stronger feature representation ability is possessed by CBAM. It is widely applied in computer vision tasks such as image classification, object detection, and semantic segmentation.

2.4. Pedestrian attribute (PA) 100K dataset

PA100K is a dataset in the field of pedestrian attribute recognition. It is large-scale and highly influential. It occupies an important position in the relevant research and application process. It is carefully collected from outdoor surveillance cameras and contains as many as 100,000 pedestrian images. In the current scope of pedestrian attribute recognition data resources, its scale is outstanding. The internal structure of the dataset is meticulously planned. Among them, 80,000 images are designated for training models. The models build effective recognition capabilities through learning from a large amount of data. 10,000 images are used as validation images. These validation images are used to finely adjust and optimize the model parameters during the training process. Another 10,000 images serve as test images. These test images are used to accurately evaluate the performance and accuracy of the model in practical application scenarios.

It is particularly noteworthy that each image is accompanied by 26 detailed and highly representative attribute annotation information. In terms of clothing and accessories, it distinguishes whether a hat or glasses are worn in detail. The types of tops are subdivided into short-sleeved, long-sleeved, striped, logo-printed or patterned, color-blocked, plaid shirts and other styles. The lower garments also cover categories such as striped, printed, long coats, long pants, short pants, skirts or dresses. At the same time, it clearly annotates the presence or absence of shoes and the specific types of bags. Such as handbags, shoulder bags, backpacks, and even whether there are handheld items in front of the body. From the perspective of character characteristics, the age is divided into three segments: over 60 years old, 18 - 60 years old, and under 18 years old. The gender information is clearly defined, and the orientation of the human body is precisely defined as facing forward, sideways or backward. These annotation dimensions are rich and comprehensive.They provide an extremely detailed and accurate data foundation for pedestrian attribute recognition research.They strongly promote the in-depth development and innovative application of technologies in this field.

3. Experimental Design

3.1. Model training

In the research work of this chapter, we will focus on the ppclenet network architecture for in-depth exploration and experimentation. Among them, the depthwise separable convolution module plays an extremely crucial role in the ppclenet network. Through its unique structural design, it effectively reduces the computational load of the network and improves the operational efficiency. To further optimize the performance of this network, we plan to introduce the CBAM attention mechanism into the depthwise separable convolution module respectively. The CBAM attention model can perform adaptive weighting processing on the feature information in the channel dimension. The CBAM attention model can also perform adaptive weighting processing on the feature information in the spatial dimension. It enables the network to more accurately focus on the regions and features in the image that are of critical significance for the task. It significantly enhances the network's ability to extract and represent features.

We employ the PA100K dataset to conduct this experiment. The dataset contains 100,000 pedestrian images collected from outdoor surveillance cameras . And it is one of the relatively large and highly valuable datasets in the field of pedestrian attribute recognition. It is randomly divided into 80,000 training images, 10,000 validation images and 10,000 test images. And each image is annotated with 26 commonly used attributes covering multiple dimensions from clothing to posture. During the specific experimental operation process, we will systematically add the CBAM attention model in different positions of the depthwise separable convolution module in turn. When added in the pre-stage of depthwise separable convolution, observe its effect on attention guidance in the initial feature extraction process. When added in the middle position, explore its influence on the feature transformation and transfer process. When added in the post-stage, analyze its role in the final feature integration and output stage. The specific experiments are shown in Table 1. A meticulous and comprehensive experimental setup is carried out. The PA100K dataset is utilized. A rigorous evaluation and analysis of the models trained under each addition position is conducted in multiple evaluation metrics. The optimal position of the CBAM attention model in the depthwise separable convolution module of the ppclenet network is determined. The improvement of network performance is maximized. A solid and efficient network model foundation for subsequent applications and research in related fields is laid. The development and progress of the entire technical system in related tasks such as image recognition and object detection is promoted.

Table 1. Experimental Setup

Experimental	CBAM Positions
Experiment One	The Second Module
Experiment Two	The Third Module
Experiment Three	The Fourth Module
Experiment Four	The Fifth Module
Experiment Five	The Sixth Module

3.2. Evaluation results

In this experiment, the position of the CBAM attention model within the depthwise separable convolution module

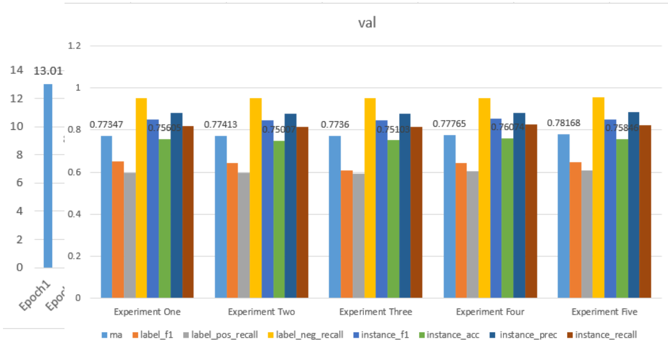


Fig. 5 Experiment Five

Fig. 6 val

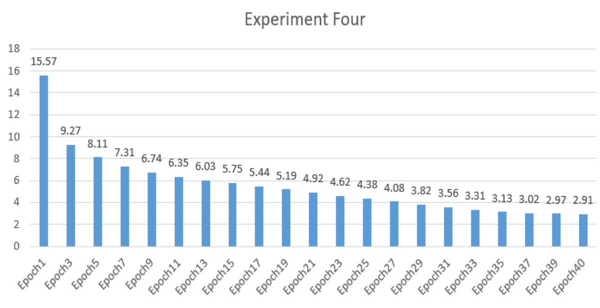


Fig. 4 Experiment Four

was explored intensively. Then, the position was adjusted. Based on this, the training work was carried out. Also based on this, the validation work was carried out. The epoch was set as 40, and the learning rate was precisely set at 0.01. Meanwhile, the momentum parameter was taken as 0.9 to ensure the stability and effectiveness of the training process. During the training process, the changes in loss values under different position settings were closely monitored, and corresponding charts were drawn.

A detailed analysis of the charts is conducted. It enables the clear discernment of performance differences of the CBAM attention model at various positions within the depthwise separable convolution module..Fig. 1, Fig. 2, Fig. 3, Fig. 4 and Fig. 5 respectively display the scenarios where the loss declines as epochs increase for different positions of CBAM in the five experiments.

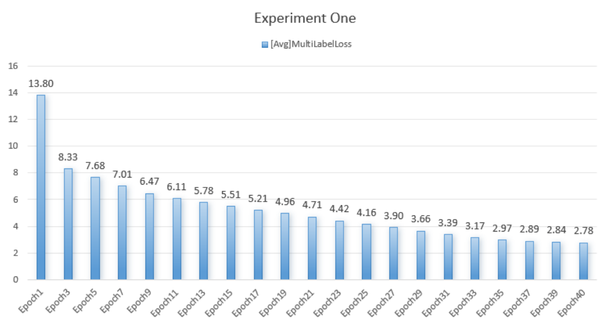


Fig. 1 Experiment One

It can be observed from the aforementioned five figures that after 35 epochs, the rate at which the loss is decreased with the increment of epochs is extremely slow. In Experiment One, a loss value of 2.78 is achieved. In Experiment Two, a reduction of the loss to 2.8 is attained. In Experiment Three, the loss is brought down to 2.77. In Experiment Four, the loss is lowered to 2.91. And in Experiment Five, the loss is diminished to 2.94.

The following Fig. 6 presents the performance outcomes of each group of experiments during the evaluation phase. Our primary focus lies in the mean average (ma) and instance accuracy (instance_acc). Among the five groups of experiments, in terms of ma, Experiment Five exhibits the most outstanding performance with a value of 0.78168. Experiment Four follows closely behind, with a ma value of 0.77765. Regarding instance_acc, Experiment Four shows the most remarkable result, reaching a value of 0.76074. Experiment Five is the runner-up, with an instance_acc value of 0.75846. This indicates that different experimental setups have varying degrees of influence on these two key evaluation metrics. And a comprehensive analysis of such differences is crucial for a more in-depth understanding and optimization of the experimental model and its associated techniques.

4. Conclusion

In this article, the depthwise separable convolutional neural network, the CBAM attention model and the dataset pa100k were first introduced. Then, the attention model was added into different modules of pplcnet for training and evaluation. Through the comparative analysis of evaluation model data, it was discovered that the performance in Experiment Four and Experiment Five was

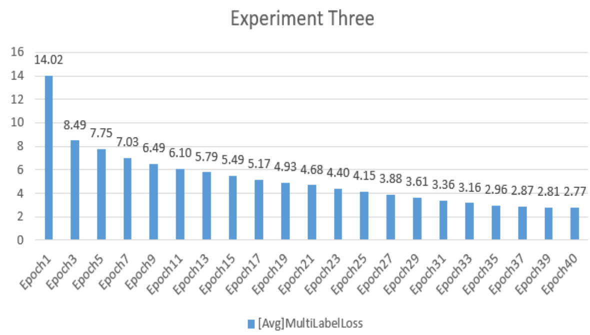


Fig. 3 Experiment Three

better. Namely, when the CBAM attention was placed in the last two modules, a higher model accuracy was obtained. The depthwise separable convolutional neural network served as the base framework. With its unique structural features, it can maintain certain feature extraction ability while reducing computational complexity and has drawn much attention. The CBAM attention model, owing to its dual attention mechanism in channel and spatial dimensions, can enhance the network's ability to capture key information. The dataset pa100k was used to supply data for model training and evaluation.Subsequently, the attention model was inserted into pplcnet modules for training and assessment, involving parameter adjustment and model optimization. A strict experimental process and a comprehensive evaluation were carried out. From the data comparison of evaluation models, it was clearly observed. Experiment Four and Five exhibited significant indicator advantages. Specifically, with the CBAM attention in the last two modules, a higher model precision could be attained. This finding is anticipated to offer valuable reference and guidance for related model architecture design and optimization and promote the further development of research in this area.

Acknowledgements

The research is supported by the major special project of carbon neutralization technology in Tianjin, China in 2024 (24ZXTKSN00070).

References

1. Wu Peishu, Zidong Wang, Han Li et al., KD-PAR: A Knowledge Distillation-Based Pedestrian Attribute Recognition Model with Multi-Label Mixed Feature Learning Network. *Expert Systems with Applications* ,2023.121305.
2. Ji Zhong, Zhenfei Hu, Erlu He, et al., Pedestrian Attribute Recognition Based on Multiple Time Steps Attention. *Pattern Recognition Letters* 138.2020.07.018.

Authors Introduction

Ms. Peng Wang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2014, she received a doctorate from North China Electric Power University. The research direction is the functional safety assessment of safety instrumented systems.

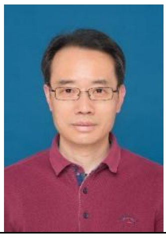
Mr. Shengfeng Wang

In 2023, he received his Bachelor of Engineering degree from the School of Electronic Information and Automation, Tianjin University of Science and Technology, China.He is pursuing a master's degree in engineering from Tianjin University of Science and Technology.

Ms.Yuting Zhou

In 2023, she entered Tianjin University of Science and Technology.She is pursuing a Bachelor of Engineering degree in the School of Electronic Information and Automation.

Mr. Qikun Wang



In 1996, he received his Bachelor of Engineering degree from the School of Electronic Information and Automation, Tianjin University of Science and Technology, China. He is a senior

Optimizing Microgrid Power Dispatch with Integrated Ground Source Heat Pumps Using Cellular Automata

Peng Wang
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

Shunqi Yang
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China
E-mail: autowangpeng@tust.edu.cn, yangshunqitust@163.com

Abstract

As a new type of energy supply and management system, this paper improves a simulation method based on cellular automata (CA) to optimize power dispatching in microgrids, especially for the ground source heat pump (GHP) system. Firstly, this paper simulates the dynamic behavior and interaction of each unit in the microgrid using cellular automata, addressing uncertainties on both the supply and demand sides. Secondly, this paper uses a two-level optimization method to enhance the maximization of self-consumption and optimize energy flow in the grid. The results show that this method can effectively improve microgrid energy utilization efficiency, reduce operating costs, and enhance system reliability and resilience.

Keywords: Microgrid, Cellular automata, Power dispatching, Sustainable energy sharing, Ground source heat pump

1. Introduction

With the growth of global energy demand and the emphasis on environmental protection, the traditional centralized energy supply mode faces the challenges of large energy transmission loss and high carbon emissions. As a distributed energy solution, microgrid has attracted much attention, which can integrate renewable energy, improve supply reliability and flexibility, and reduce transmission loss.

Ground source heat pump is the key heating and cooling equipment of microgrid. It uses shallow geothermal resources to save energy efficiently, reduce carbon emissions, and significantly reduce electricity consumption in summer and winter. In autumn, solar panels store excess power to the battery, and supply power during the peak power consumption in winter to ensure demand. When electricity is surplus, it can also supply neighbors and realize power allocation and sharing.

However, when the ground source heat pump is integrated into the microgrid, its power consumption fluctuation and collaborative work problems bring new challenges to power scheduling [1]. Cellular automata can accurately capture the dynamic interaction of components through modeling analysis and realize power scheduling optimization, which is of great significance for improving the reliability of microgrid, reducing costs and promoting sustainable energy development.

The rest of this article is organized as follows. The second section introduces the three key equipment used in the system. In the third part, the cellular automata model is used to model the neighborhood power exchange system. In the fourth part, a single house is taken as an example to analyze and calculate the propagation probability between cellular automata, and the specific modeling analysis is carried out. The fifth part summarizes the main content of this paper.

2. Key Technology Introduction

The micro-grid is composed of multiple buildings composed of such buildings, which are equipped with photovoltaic solar panels, ground source heat pump technology and energy storage batteries, as shown in Fig.1. The following is an introduction to these three technologies.

Fig.1 House schematic diagram

2.1. Photovoltaic solar power generation technology

Monocrystalline silicon solar panels are laid on the roof of a 200 square meter house and the wall with the best orientation. The inclination angle is determined according to the latitude to fit the solar radiation angle and improve the capture efficiency.

The installed power (P_{total}) is given by the following formula:

$$P_{total} = n \times P_s \tag{1}$$

The n represents the number of photovoltaic solar panels, and P_s represents a single power.

Through the photovoltaic effect, the solar energy is converted into direct current, and the inverter (conversion efficiency (η_{inv}) is about 95 % -98 %) is converted into alternating current for use in the house. The power generation P_{pv} (t) fluctuates in real time with the sunshine intensity I(t) and temperature T(t), following the formula:

$$P_{pv}(t) = P_{total} \times \frac{I(t)}{I_{STC}} \times [1 + \partial \times (T(t) - T_{STC})] \quad (2)$$

I_{STC} is the standard test light intensity (1000W / m2), T_{STC} is the standard test temperature (25℃), and α is the power temperature coefficient (about-0.4%/℃ for monocrystalline silicon)[2].According to the meteorological data, the power generation is calculated hourly and the power generation curve is constructed as shown in Fig.2.

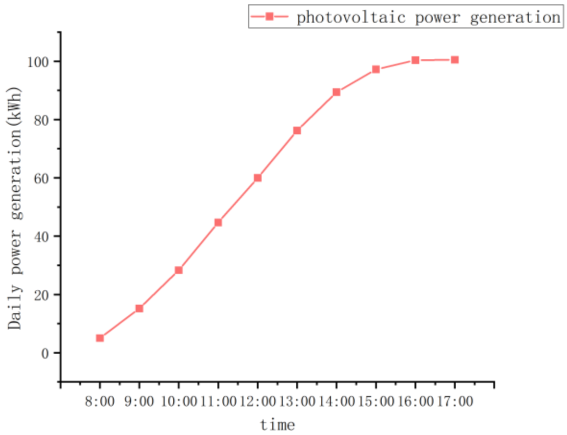


Fig.2 Solar panel daily power generation curve

2.2. Ground source heat pump

The ground source heat pump uses the principle of Carnot cycle and inverse Carnot cycle to transfer cold and heat. In winter, the heat pump unit absorbs the heat in the soil through the underground buried pipe system, and transmits the heat to the room through the heat exchanger to achieve heating. In summer, the heat pump unit releases the indoor heat into the soil through the underground buried pipe system to achieve refrigeration. At the same time, the ground source heat pump also utilizes the huge heat storage capacity of the underground soil to form an annual cold and heat cycle. The schematic diagram of ground source heat pump is shown in Fig.3.

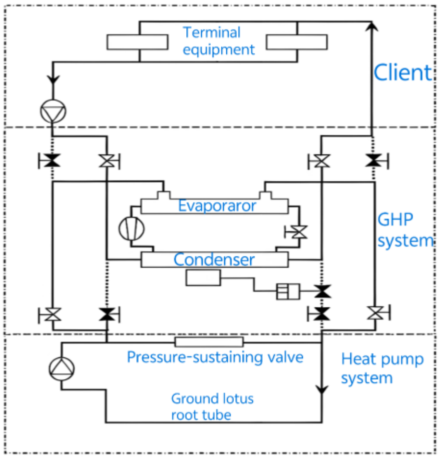


Fig.3 Ground source heat pump schematic diagram

The main power consumption of ground source heat pump is compressor operation, heat exchanger operation and waterway circulation.

The heating performance coefficient of ground source heat pump can be expressed as:

$$\beta = \frac{Q_h}{W} = \frac{(Q_e + W)}{W} = \frac{Q_e}{W} + 1 \quad (3)$$

Among them, Q_h represents the heat provided by the heat pump system to the user,W represents the electrical energy consumed by the system, and Q_e represents the heat absorbed from low-temperature heat sources (such as water). The heating performance coefficient β is always greater than 1, which indicates that the geothermal heat pump heating can save high grade energy compared with the geothermal heat pump driven energy direct heating.

2.3. Battery energy storage system

The lithium-ion battery pack is selected, and the capacity is estimated based on the surplus electricity in spring and autumn and the deficit electricity in summer and winter. Battery rated capacity C_{bat} (unit is kWh). The voltage V_{bat}, the state of charge SOC(t) characterizes the proportion of remaining electricity, the update formula is:

$$SOC(t) = SOC(t-1) + \frac{\eta_c \times P_c(t)}{C_{bat}} - \frac{P_d}{\eta_d \times C_{bat}} \quad (4)$$

The η_c and η_d are the charge-discharge efficiency (charge about 93 % -95 %, discharge about 92 % -94 %), and the charge-discharge power. In the spring and autumn, when the solar power surplus, according to the control strategy to the appropriate current and voltage charging ; in summer and winter, the peak discharge of electricity is used to supplement energy, stabilize load fluctuations, and ensure continuous power supply.The lithium electronic battery is used to store the power, and the charge and discharge efficiency and self-discharge efficiency of the battery are calculated by the formula.

$$Round-TripEfficiency=(\frac{Discharged\ \ Energy}{Charged\ \ Energy})\times 100\% \tag{5}$$

$$Self-Discharge\ Rate=(\frac{Initial\,Capacity-Final\,Capacity}{Initial\,Capacity})\times 100\% \tag{6}$$

3. Model Construction

The second - order five - state probabilistic cellular automata model is used to conduct in - depth modeling and analysis for a specific case, that is, a scene composed of a single 200 square meters of housing buildings. In this model, the house integrated with ground source heat pump and solar photovoltaic power generation technology is treated as an independent producer - retailer unit, and a self - sufficient energy sharing system is constructed for each such unit[3]. This structure is named ' smart node ' by us, which can achieve power supply stability by intelligently adjusting the upper and lower limits of battery energy storage capacity as shown in Fig. 4.

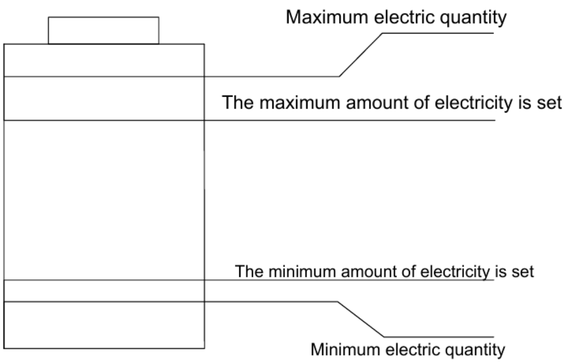


Fig.4 Battery power setting

3.1. Calculation of electricity generated and consumed by users

Statistics are made on the power generation of photovoltaic solar panels and the power consumption of ground source heat pumps as the main power consumption of pilot houses from 2021 to 2023, as shown in Fig. 5 and Fig. 6.

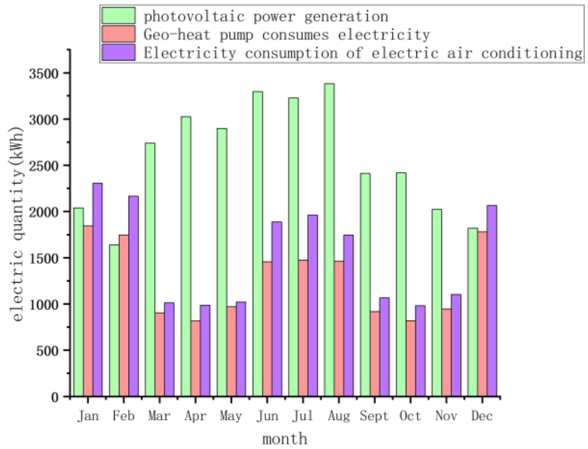


Fig.5 Monthly electricity consumption statistics

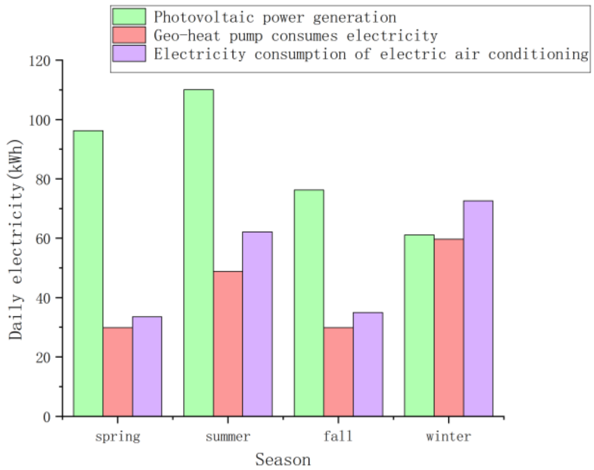


Fig.6 Seasonal electricity consumption statistics

It can be seen that the power consumption is significantly higher than the production period, which is concentrated in winter, so the cellular automata model is constructed in winter as an example.

3.2. Construction of cellular automata model

In our power system, we have built a microgrid architecture that consists of a number of interconnected components that can exchange energy. These components can be either pure electricity consumers or prosumers with both consumption and production capacity, and they usually represent residential user units [4]. These users are equipped with distributed power generation facilities, such as photovoltaic solar panels. Under this framework, energy consumers and prosumers are orderly arranged in a two-dimensional grid, which together constitute the basic unit of this complex and efficient energy network.

The cell space is expressed as

$$L=\{C_{ij};i,j\in 1,2...,n\ and\ \ j=1,2...,m\} \tag{7}$$

where n and m are the number of nodal elements along the vertical axis and the horizontal axis, respectively.

All building buildings are connected to neighbors through physical and information layers. Each house unit is used as a node to link to its own index (i, j) in the community virtual grid. The cellular automata model adopts Moore neighborhood as Fig.7, and the neighborhood equation is:

$$N(c_{ij})=\{c_{kl};|k-i|\le R,|l-j|\le R\},c_{ij}\in L \quad (8)$$

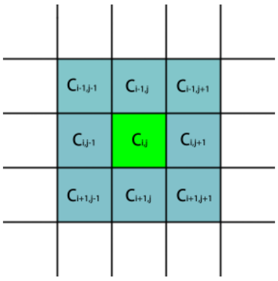


Fig.7 Moore neighborhood

Five cell states are set, and the state transition rules are shown in Fig.8.

State 1 : Solar energy can meet its own needs and charge the battery.

State 2 : Photovoltaic solar power generation can not meet its own needs, and the storage capacity of the battery can not meet its own needs. For the disadvantage state.

State 3 : The power generated by the photovoltaic solar panel can meet its own power demand and will be stored more than the power. Power supply state for neighbors.

State 4 : Power outage status.

State 5 : The state of power supply to the power grid.

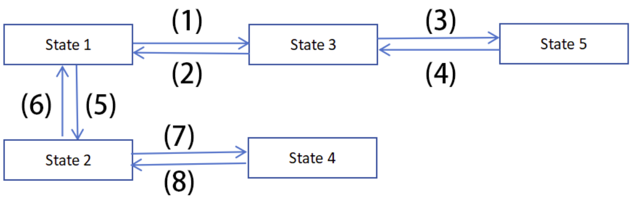


Fig.8 State transition diagram

(1) After three iterations of state 1, it is converted to state 3. That is, after a period of charging, the storage capacity of the battery is higher than the set maximum value ;

(2) The number of neighbors in the surrounding state 2 is n2, and the probability of n_2p_2 is converted to state 1. That is, when there is a state of lack of electricity around, the state of lack of electricity 2 is converted into a charging state 1 after power supply ;

(3) When $n_2 = 0$ and there is no state 2 around, it is converted to state 5 after three iterations. That is, when there is no neighbor in the state of power shortage, it will continue to be in the state of power supply to the neighbor, and after a certain period of time, it will be converted into a state of power supply to the power grid ;

(4) State 5 after an iteration, it is converted to state 3. That is, after supplying power to the grid, it is converted into state 3 ;

(5) When the solar power generation is lower than its own power consumption, the charging state 1 will be converted to the power shortage state 2 with the probability of P_3 ;

(6) n_1 is the number of state 3 that can be powered to neighbors. When power is supplied to state 2, state 3 is transformed into state 1 with the probability of n_1p_1 ;

(7) n_1 is the number of surrounding states 3, and the probability of $p_5 = p_4 / n_1$ is converted to state 4. That is to say, the smaller the number of surrounding power supply states 3, the greater the probability of power shortage state 2 transforming into power outage state 4 ;

(8) After the power supply of the power grid, it is converted into state 2 after 2 iterations ;

Under the micro-grid combined with ground source heat pump system.The simulation results are shown in Fig.9 when the battery power is not set. When the battery power is set, the simulation results are shown in Fig.10.

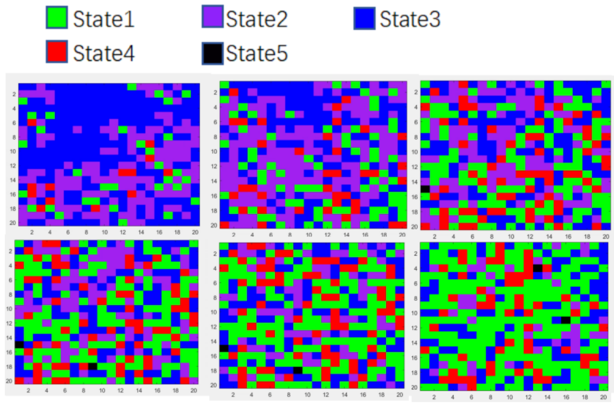


Fig.9 Simulation without setting the battery power

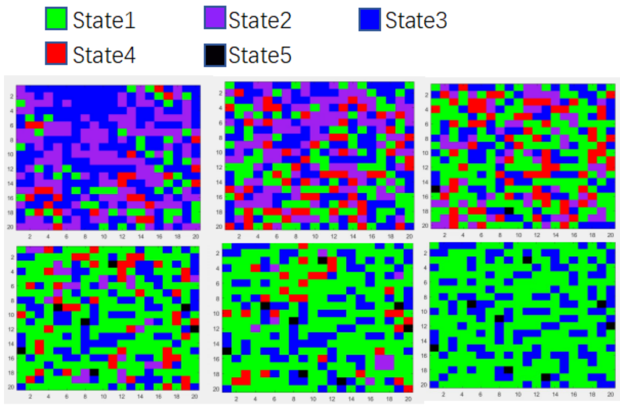


Fig.10 Simulation after the battery power setting

It can be seen from Fig.9 that when there is no electricity setting, at the end of the day, the number of power-deficient users is 47, an increase of 11.75 % compared to Fig.10. It can be seen that in winter days, the user 's power consumption state in the microgrid system is converted from an unstable state without the battery power setting to a stable state with the battery power setting ,the power setting of the battery can effectively improve the power stability of the user group.

4. Conclusion

This study explores and implements a two-level optimal storage strategy for battery power, explores the excellent low energy consumption and low carbon emission characteristics of ground source heat pump compared with electric heating system, and skillfully realizes the precise regulation and balance of power demand in summer and winter, that is, peak shaving and valley filling effect.On this basis, we introduce the cellular automata theory and construct an advanced model of microgrid energy sharing.

The technical scheme not only shows high popularization and application potential, but also can achieve efficient and intelligent configuration of energy while greatly reducing carbon emissions. Moreover, it can significantly reduce the dependence on the main power grid during the summer peak period, and has the ability to reverse power supply to the main power grid, thus effectively alleviating the operating pressure of the main power grid during the summer peak period and improving the stability and reliability of the entire power system.

Acknowledgements

The research is supported by the major special project of carbon neutralization technology in Tianjin, China in 2024 (24ZXTKSN00070).

References

1. Xing Wang,Tao Li,Yingying Yu,et al.(2024).Comprehensive analysis of a novel sustainable photovoltaic/thermal assisted ground source heat pump system with energy storage.*Journal of Energy Storage(PB)*,114161-114161.
2. H. Camblong, O. Curea, J. Ugartemendia, et al, Photovoltaic energy sharing: Implementation and tests on a real collective self-consumption system, Heliyon 9 (12) (2023) e22252.
3. L. Ali, S. Muyeen, H. Bizhani, et al, A peer-to-peer energy trading for a clustered microgrid – Game theoretical approach, Int. J. Electr. *Power Energy Syst*. 133 (2021) 107307.
4. I. Abdennour, M. Ouardouz, A.S. Bernoussi, Blockchain smart contract for cellular automata-based energy sharing, in: *International Conference on Cellular Automata for Research and Industry*, Springer, 2021, pp. 103–112.

Authors Introduction

Ms. Peng Wang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2014, she received a doctorate from North China Electric Power University. The research direction is the functional safety assessment of safety instrumented systems.

Mr. Shunqi Yang

He received his Bachelor's degree in 2023 from the School of Electronic Information and Automation, Tianjin University of Science and Technology. Currently, he is a master's student at the same university. His research area is microgrid reliability.

Indoor Personnel Thermal Comfort Monitoring System Design Based on Mobile Robots

PengWang,
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222,China

Zihang Zhou
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222,China
E-mail: autowangpeng@tust.edu.com , 1322615821@qq.com

Abstract

With the development of smart buildings, people's requirements for indoor environmental comfort are becoming increasingly higher. Studies have shown that non-invasive infraredography (IRT) technology can effectively predict thermal comfort. This paper explores the use of mobile robots to monitor the thermal comfort of indoor occupants. Mobile robots can collectant information from multiple angles, and can locate and estimate the thermal comfort of occupants in real time. This paper first illustrates the importance of thermal comfort to the indoor and how environmental factors affect comfort. It then studies the design scheme of the robot system. Finally, experiments are conducted to evaluate the accuracy and reliability of the data by the robot, and the experimental results are analyzed.

Keywords: Dynamic hermal comfort, Intelligent building, Robot system, Infrared thermal imager

1.Introduction

In modern times, technology is rapidly advancing, and the field of architecture has also seen significant progress, with smart buildings becoming a prominent trend, the way we interact with and manage indoor spaces. The core of smart buildings lies in their ability to closely monitor and control the indoor environment, where thermal comfort is key factor that directly affects the health and productivity of occupants.

The traditional approach to assessing thermal comfort in indoor environments has relied on placing fixed sensors at specific locations. Although these sensors can serve their purpose, their fixed positions limit the range of data collection, often resulting in incomplete coverage of the entire indoor space. Additionally, they lack the adaptability needed to account for the dynamic changes in the room's occupancy patterns and the distribution of occupants.

To address these issues, mobile robots can be used to monitor indoor thermal comfort. With their mobility and flexibility, mobile robots offer a novel solution to the challenges faced by traditional sensing methods[1]. They are equipped with thermal imaging cameras and environmental sensors that enable data collection from multiple vantage points within the indoor environment. The thermal imaging camera, with its high resolution and wide field of view, can capture detailed thermal images of occupants, providing information about their surface body temperatures. Meanwhile, the environmental sensors, which measure parameters such as temperature, humidity, and air speed, offer real-time data about the environmental conditions prevalent in the room.

Combining thermal imaging data with environmental data can more accurately evaluate thermal comfort and calculate the Predicted Mean Vote (PMV) thermal comfort index, is widely recognized and used in the field of indoor environmental quality assessment.

The remainder of this paper is organized as follows. Section 2 provided a detailed description of the hardware aspects of the system design. Section 3 focuses on the functional modules of the robot system, showing the operation of each software part in the entire system. Section 4 presents preliminary experiments and analyzes the experimental to explore the impact of the collected data on understanding thermal comfort and its determinants. Finally, Section 5 concludes the main findings and outlines potential directions for future in the field.

2.System Design

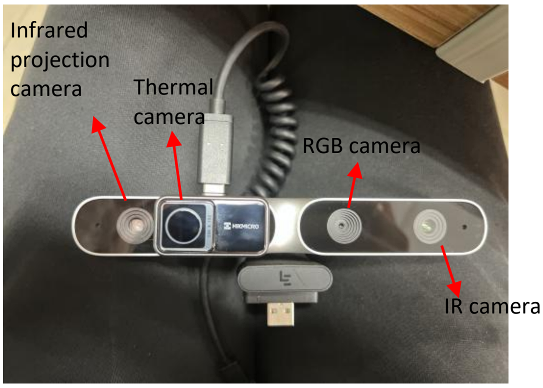
The proposed system is meticulously crafted with a comprehensive and in-depth consideration of a multitude of factors, all of which are essential to guarantee its optimal effectiveness and unwavering reliability in the crucial task of monitoring the thermal comfort of indoor personnel. This undertaking involves a holistic approach that takes into account not only the technical aspects but also the practical requirements and challenges associated with operating within an indoor environment.

The mobile robot, for instance, is not just a simple vehicle but a sophisticated platform that combines mobility with a suite of sensing capabilities. Its design is optimized to balance the need for compactness, which allows it to navigate through tight spaces and around obstacles commonly found in indoor settings, with the requirement for housing and powering a range of sensors and processing units.

This system not only provides real-time data on the thermal comfort levels of occupants but also offers actionable insights that can be used to optimize the indoor environment, enhance occupant well-being, and improve energy efficiency.

2.1. Mobile robot platform

The mobile robot used in this study is a platform with a compact design. It is equipped with a high-resolution infrared thermal imaging camera and a suite of environmental sensors, including temperature, humidity, and air velocity sensors. The robot is controlled by an onboard microcontroller, which enables it to navigate autonomously within the indoor environment. The infrared camera has a wide field of view and can capture detailed thermal images of occupants. The environmental sensors provide real-time data on the ambient conditions, which are used in conjunction with the thermal images to assess thermal comfort. The hardware part for robots to acquire external information is shown in Fig.1.



a.The visual components of the robot

b. Mobile base with temperature and humidity ,wind speed detection module

Fig. 1. The hardware part for robots to acquire external information.

2.2. Data acquisition and processing

The data collected by the robot is transmitted to a central server for processing, via both wired and wireless means. The server uses image processing algorithms extract thermal information from the images, such as the surface temperature of the occupant. Environmental data is also integrated into the analysis, to account for the impact of temperature, humidity, and wind speed on thermal comfort. The processed data is used to calculate the thermal comfort index, or Predicted Mean Vote (PMV)[2]. As shown in Table1, the thermal sensation labels are represented in seven levels.

Table 1
Thermal sensation labels.

Thermal sensation	Label
Uncomfortable warm	3
Warm	2
Slightly warm	1
Comfortable	0
Slightly cold	-1
Cold	-2
Uncomfortable cold	-3

3.Introduction to System Modules

With the help of these functional modules, we can achieve personnel thermal comfort monitoring.

3.1. Multi-passenger facial recognition

To avoid the problem of duplicate calculations during the robot's room patrol, a method for multi-occupant facial matching and recording was.

The facial detection algorithm Multi-task convolutional neural networks(MTCNN) identified the facial regions in the image frame. MTCNN is a multi-task cascaded deep learning network that integrates bounding regression and feature point selection to achieve facial recording and matching. Before performing facial matching, facial embedding is an essential step that allows the matching algorithm to focus on the features of the image. Facenet is an extraordinary embedding algorithm, especially in facial matching. It can directly map the face to a high-dimensional vector in Euclidean. Matching vectors in Euclidean space to determine if they are from the same person. In general, vectors from the same person are very close in Euclidean space while vectors from different people are far apart. The method of calculating the Euclidean distance of vectors is:

$$\text{vectorX} = [x1, x2, \ldots, xn] \tag{1}$$

$$\text{vectorY} = [y1, y2, \ldots, yn] \tag{2}$$

$$\text{EuclideanDistanceXY} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{3}$$

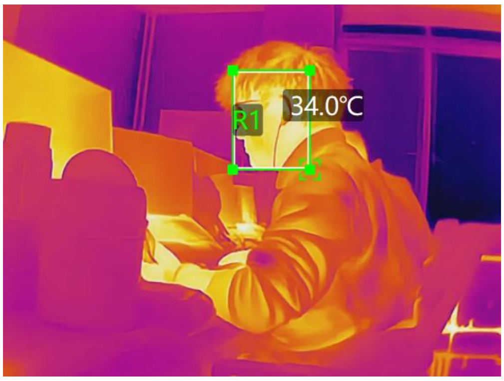
Here we set the threshold to 1.242 according to the Euclidean distance computation criterion. If the Euclidean distance between the two vectors is less than 1.24, then the two face bounding boxes are considered to be from the same person, and the face match is successful. Otherwise, it means they may come from people, and the faces do not match.

If the facial image matches the observed human face, the ID is labeled with the unique number of this person. If the facial image does not, a new number is registered in the dataset, and the facial image is labeled with that number. Finally, the detected facial image is added to the dataset for matching. However, when we record the detected images with a mobile robot, the dataset grows too large to match. To match quickly, we only record two facial with the closest Euclidean distance for each person.

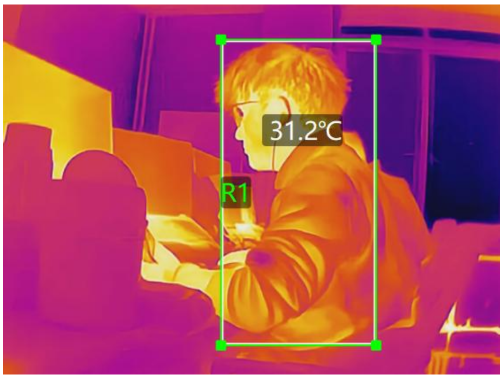
3.2. Infrared temperature measurement module

To obtain human thermal comfort, body temperature is indispensable. We used the YOLOv8 human recognition model in combination with an infrared imager to measure the temperature of the human body trunk and face respectively. The experimental results show that using facial temperature instead of body temperature is more accurate and better the thermal sensation of the person. The contrastive experiment is shown in Fig.2

Therefore, the system uses the YOLOv8 deep learning model for facial recognition, and obtains thermal image data in real through a thermal imaging camera[3]. The obtained thermal images are preprocessed to meet the input requirements of the trained model, and then the images are fed into the model inference. The model outputs the bounding box location and temperature prediction value of the face, and the system uses this information to perform visual annotation on the image, showing location of the face and the predicted temperature. The facial temperature measurement of the thermal image is shown in Fig. 2(a)



a.Facial thermal image



b.Body thermal image

Fig. 2. Infrared thermal image

3.3. *Human body attribute recognition module*

The robot system is embedded with a series of convolutional neural networks to recognize human characteristics, including age, gender, pose, and. The BodyPix model is applied to segment the human body from the background, and then convolutional neural networks are used to identify the characteristics of the occupant The model is based on the Yolov8 object detection network and is trained with the Adience dataset, which contains 29,000 images It can simultaneously recognize information about a person's gender, age, and body pose. However, different people can have significant differences in appearance at the same age, it difficult to predict exact age solely by analyzing a person's image. Therefore, it only predicts possible age ranges in five groups (14-17,18-26, 27-35, 36-42, and over 42). The clothing recognition model is based the Yolo-v5 object detection network and is trained with the DeepFashion2 dataset, which contains images of various clothing. Clothing is categorized from thin to thick seven levels, including short-sleeve shirts, shorts, long-sleeve shirts, long pants, sweaters, coats, and down jackets. These models are and can be embedded in the robot's system for real-time computation.The test results are shown in Fig.3.

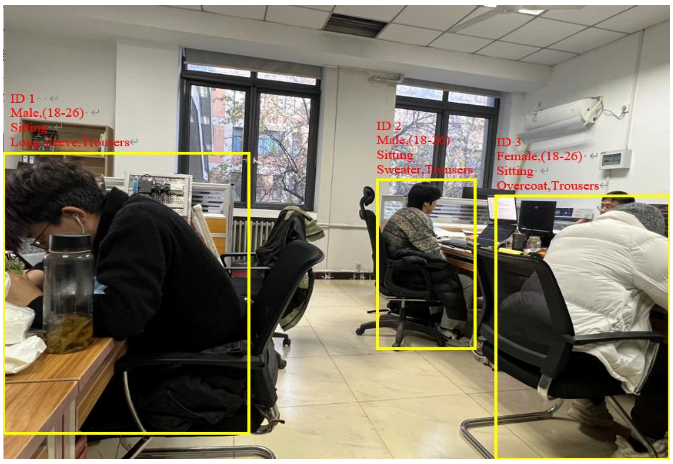


Fig 3. Human body feature detection

3.4. Robot path tracking and thermal comfort distribution

There are two parts working together in this module. One is the global positioning and tracking of the robot, and the other is the positioning the human. The former detects surrounding objects and spatial layout by emitting laser pulses from the LiDAR and measuring the time it takes for these pulses to bounce back This information is then used to build a three-dimensional map of the environment. With the environmental map and its own positioning, the robot can use the path planning Dijkstra algorithm to calculate the optimal path from its current position to the target position. At present, we mark five points on the main road in the room for the robot move in a loop, achieving comprehensive detection of the room environment. The latter is to estimate the position of the human body relative to the robot. In this part, RGB images and depth are collected synchronously to show the distance between the person and the camera, which is used to build a three-dimensional point cloud of objects in the camera view and three-dimensional coordinate $(x_c, \ y_c, \ z_c)$ in the robot view. The body relative coordinates are then converted to global coordinates (x_g , y_g , z_g) through the rotation matrix R and the translation vector t as follows:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R_{3\times3} & 1_{3\times1} \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x_g \\ y_g \\ z_g \\ 1 \end{bmatrix}. \tag{4}$$

When a new view is received, the facial detection and localization function is triggered. As extracting features from the model is time-consuming, the feature function runs ten times a minute on the collected personnel image segmentation. The thermal comfort prediction function is triggered when the feature attributes have been fully collected or some attributes have updated. The room mapping function is triggered when the robot is on the move, displaying the predicted thermal comfort on the map in real-time.

4. Prediction of Thermal Comfort and Experiment

4.1 Thermal comfort prediction algorithm

Danish scholar Professor Fanger took into account several factors affecting human comfort, established the thermal comfort sensation evaluation index equation through relevant data analysis, and proposed the PMV representing human cold and hot sensation, its computational formula is:

$$\text{PMV} = [0.303\exp(-0.036\text{M}) + 0.0275]\text{TL} \tag{5}$$

$$\begin{aligned} TL = {} & (M - W) - 3.05[5.733 - 0.007(M - W) - P_a] \\ & - 0.42(M - W - 58.15) \\ & - 1.73 \times 10^{-2} M(5.867 - P_a) \\ & - 1.4 \times 10^{-3} M(34 - t_a) \\ & - 3.96 \times 10^{-8} \times f_{ct}[(t_{ct} + 273)^4 \\ & - (t_r + 273)^4] - f_{cl} h_c (t_{ct} - t_a) \end{aligned} \tag{6}$$

However, the mathematical expression of the PMV thermal comfort index he uses has the disadvantages of complex computation and inaccurate results, and cannot used as a control variable for real-time control of indoor equipment. Therefore, this project adopts a PMV prediction model based on BP neural network[4]. BP network is a multi-layer feedforward network that continuously trains samples to adjust the network's weights and thresholds, minimizing the error between the output value and the desired value The network mainly consists of an input layer, an intermediate layer, and an output layer, and it is trained through forward signal propagation and backward error propagation.

To improve the slow convergence speed of traditional BP neural networks, the bird flock algorithm is used to optimize the initial weights and thresholds of the BP neural network. The bird algorithm is a new type of optimization algorithm that not only has the advantages of the particle swarm optimization (PSO) algorithm but also effectively avoids premature convergence due to diversity. Therefore, this project uses the bird flock algorithm to optimize the BP neural network, and uses this model to predict the PMV of the indoor environment.

4.2. *Experiment and data analysis*

The experiment was conducted in a controlled indoor environment, such as an office or laboratory. The room was equipped with heating, ventilation, air conditioning (HVAC) to maintain a stable environmental temperature. Mobile robots were deployed in the room and programmed to move along a predetermined path, periodically collecting and environmental data. The thermal images were used to measure the surface temperature of the occupants, and the environmental data were used to calculate the thermal comfort index. The thermal comfort was visualized as different colors on a map, and finally, a thermal comfort distribution map of the room was generated, as shown in Fig.4 All occupants were dynamically added to the global map with color-coded estimated thermal comfort. We conducted a questionnaire survey to compare the optimized BP neural network prediction algorithm with traditional methods, asking whether the PMV values were consistent or close to the subjective thermal comfort ratings provided by the occupants. The survey results were summarized and Table2 after the experiment. The results show that the system is capable of accurately detecting changes in thermal comfort caused by changes in environmental conditions. Our experiments recorded the feature results of multiple indoor individuals, as shown in the Fig 5, proving the validity and accuracy of the model.

Table 2

Prediction accuracy.

Group number	BP algorithm	Traditional
1	92%	74%
2	88%	69%
3	96%	78%
4	92%	66%
5	86%	60%

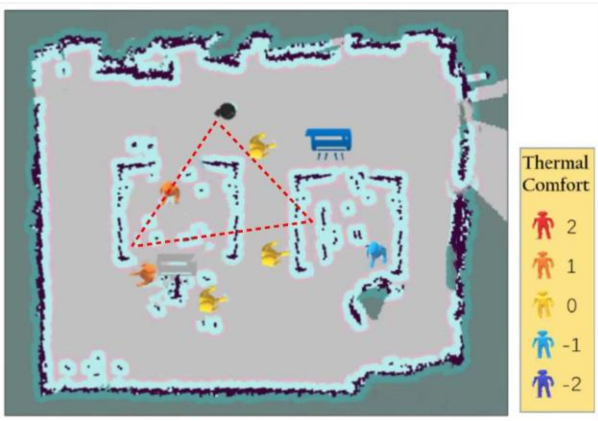


Fig 4. Thermal comfort distribution

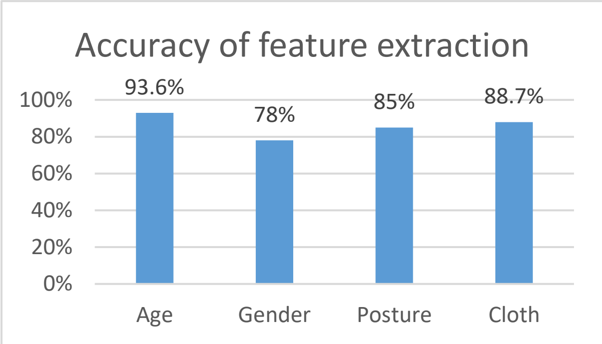


Fig 5. Results of human feature extraction.

4.3. *Future development direction*

The data collected by the mobile robot provides valuable insights into the thermal comfort of indoor occupants. By analyzing the relationship between the body surface temperature, environmental factors, and thermal comfort indices, it is possible to identify the factors that have the greatest impact on thermal comfort. This information can be used to optimize the HVAC system settings to improve occupant comfort and energy efficiency. For example, if the analysis shows that a particular area of the room is consistently experiencing discomfort due to high temperatures, the HVAC system can be adjusted to provide more cooling in that area.

5. Conclusion

This paper presents a system based on mobile robots for real-time monitoring of indoor occupants' thermal comfort. The experimental results show that the system can and accurately evaluate thermal comfort. Future work will focus on improving the accuracy and reliability of the system, as well as integrating it with other building management systems to achieve comprehensive environmental control, allowing the indoor environment to actively adjust to enhance occupant comfort and health.

Acknowledgements

The research is supported by the major special project of carbon neutralization technology in Tianjin, China in 2024 (24ZXTKSN00070).

References

1. D. Martinez, M. Teixidó, D. Font, et al, *Ambient intelligence application based on environmental measurements performed with an assistant mobile robot*, Sensors 14 (4) (2014) 6045–6055.
2. J. Miura, M. Demura, K. Nishi, et al, *Thermal comfort measurement using thermal-depth images for robotic monitoring*, Pattern Recognit. Lett. 137 (2020) 108–113.
3. A. Ghahramani, G. Castro, B. Becerik-Gerber, et al, *Infrared thermography of human face for monitoring thermoregulation performance and estimating personal thermal comfort*, Build. Environ. 109 (2016) 1–11.
4. BP neural network multi-module green roof thermal performance prediction model optimized based on sparrow search algorithm. 96 (2024) 110615.

Authors Introduction

PengWang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2014, she received a doctorate from North China Electric Power University. The research direction is the functional safety assessment of safety instrumented systems.

Zihang Zhou

He received his Bachelor's. degree from College of Tongji University Zhejiang College, China. He is currently a Master course student in Tianjin University of Science and Technology. His research area is about Internet of Things in Architecture.

Market Trading Strategy of Integrated Energy Park from the Perspective of Non-cooperative Game

Peng Wang
ollege of Electronic Information and Automation, Tianjin University of Science and Technology, 300222，China

Siyi Wang
ollege of Electronic Information and Automation, Tianjin University of Science and Technology, 300222，China

Liangyu Wang
ollege of Electronic Information and Automation, Tianjin University of Science and Technology, 300222，China

Chengkai Miao
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222，China

E-mail: autowangpeng@tust.edu.cn, 1363489829@qq.com, 2697909953@qq.com, miaock12345@qq.com

Abstract

This paper introduces a park trading framework including energy managers, distributed photovoltaic and wind power users and electric vehicle charging service providers, and establishes a non-cooperative game model in which three subjects pursue maximum benefits. Taking a typical winter day in a park as an example, the simulation results show that: in the game equilibrium, energy managers profit from energy supply, distributed photovoltaic and wind power users improve resource utilization and reduce costs through margin online sales, and electric vehicle charging service providers choose low-bid charging to reduce costs and assist users to absorb excess resources and reduce the load of distribution network.

Keywords: integrated energy market, electric automobile, photovoltaic power generation, energy scheduling

1. Background

With the aggravation of the global greenhouse effect and energy crisis, the contradiction between energy demand and the natural environment became increasingly obvious. Countries began to attach importance to a safe, efficient, low-carbon, and clean energy operation mode to promote energy supply-side reform. The integrated energy system (IES) is a system coupling power, natural gas, heating, and transportation. It became an important research direction for the efficient utilization of distributed renewable energy. The IES effectively reduced carbon emissions and helped achieve the "double carbon" goal by jointly scheduling multiple energy sources. In the park IES, it was of great practical significance to use renewable energy, improve the flexibility of demand-side scheduling. And realize multi-energy complementarity[1].

In the integrated energy park, multiple market entities such as system energy operators, distributed photovoltaic users. And electric vehicle charging agents needed to conduct energy transactions to achieve efficient use and optimal allocation of resources. However, the traditional cooperative trading methods faced problems such as information asymmetry and game strategies, which resulted in low transaction efficiency. To address this issue, this paper studied the integrated energy market mechanism in which different market players operated in coordination. And established a three-party model of park energy operators, distributed photovoltaic user clusters, and EV charging agents, including the CHP system. And discussed the game bidding model within the park IES. Finally, taking a business park IES as an example, the improved particle swarm optimization algorithm was used to verify the model.

The rest of this article was organized as follows. The second section introduced the integrated energy park scenario. In the third part, the trading strategy game model of the three party market players were discussed. In the fourth section, PSO algorithm principle was introduced. The fifth section provided an examples to verify the availability of the designed model. The sixth part summarized the main content of this paper.

2. Scene Analysis of Integrated Energy Park

The structure of the integrated energy park studied in this paper is shown in Fig.1, which mainly included the park energy trading center (ETC) and three market players:park energy operators, distributed photovoltaic users, and electric vehicle (EV) charging agents.The ETC of the park transmitted transaction information and scheduling instructions to these market players through their respective energy management systems (EMS). The EMS was responsible for formulating energy quotation strategies and managing energy demand. While the ETC collected, distributed, and calculated information according to the market trading mechanism. The electric energy produced by the park followed the principle of local consumption and did not sell electricity to the superior distribution network.

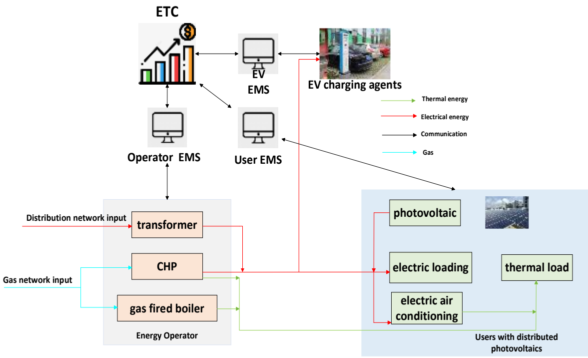


Fig.1 Structure of integrated energy park

2.1. Energy operator

As the leader of the market, energy operators are responsible for the supply of electric energy and heat energy in the park. Its main task is to formulate reasonable energy prices, connect the external energy supply network with the internal distribution network. And ensure that the energy needs of the park are met. Operators optimize the production and distribution of energy by scheduling their own energy equipment to maximize revenue. The pricing model is shown as Eq. (1) Eq. (2) and Eq. (3).

$$P_{opr} = [P_{e,opr}, P_{h,op}r] \tag{1}$$

$$P_{e,opr} = [P_{e,opr}(1), P_{e,opr}(2), \ldots P_{e,opr}(T)] \tag{2}$$

$$P_{h,opr} = [P_{h,opr}(1), P_{h,opr}(2), \ldots P_{h,opr}(T)] \tag{3}$$

Where $P_{e,opr}$ and $P_{h,opr}$ are the prices of electricity and heat sold by energy operators respectively; T is the total number of scheduling periods in a day.

2.2. Users with distributed photovoltaics

Users with distributed photovoltaics are both energy producers and consumers. The user is equipped with photovoltaic power generation facilities, mainly relying on photovoltaic power generation to meet their own needs. When photovoltaic power generation is insufficient, users purchase electricity from energy operators; in the case of excess power generation, users can sell excess power to EV charging agents to maximize revenue[2]. The electricity price model of electricity sales is shown as Eq.(4).

$$P_{user} = [P_{e,user}(1), P_{e,user}(2), \ldots P_{e,user}(T)] \tag{4}$$

Where P_{user} is the user 's electricity sales price.

2.3. EV charging agents

The EV charging agent is responsible for managing the charging demand of electric vehicles in the park. According to the market quotation, the charging agent selects the optimal power supplier and formulates the charging strategy to reduce the charging cost and promote the consumption of photovoltaic resources. Through interaction with users and energy operators, charging agents can effectively adjust the charging load and reduce the pressure on the distribution network. When choosing the power supplier, the charging agent will give priority to operators and users with lower electricity prices. The selection strategy is shown as Eq.(5).

$$P_{ch}(t) = \min\{P_{e,user}(t), p_{e,opr}(t)\} \tag{5}$$

The trading mechanism is shown in Fig.2.

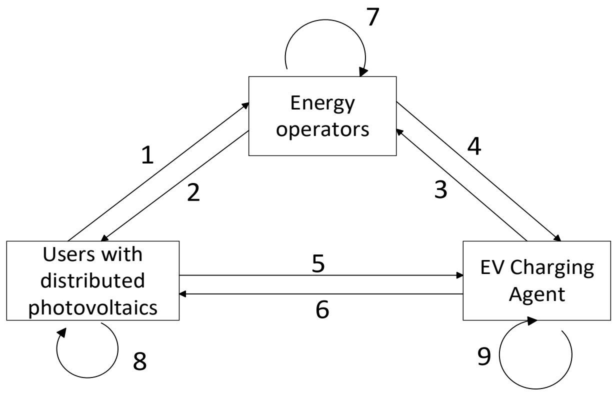


Fig.2 The trading mechanism

The explanation of the trading mechanism diagram is shown in Table 1.

Number	Interpretation
1	Report the adjusted energy consumption plan
2	Publish quotations $P_{e,opr}$ and $P_{h,opr}$
3	Reporting the energy required by the operator
4	Publish quotations
5	According to $P_{e,opr}$ specified quotation P_{user}
6	Users need to be reported
7	Adjust its own operating status and energy sales quotation
8	Adjust its own operating status and energy sales quotation
9	Select two quotations

Table 1. The explanation of trading mechanism diagram

As the leader of the internal market transaction of the park, the energy operator first publishes the initial energy supply quotation to the ETC of the park according to the historical data. And then passes the quotation to the EMS of the user and the EV charging agent. After receiving the quotation, users with distributed photovoltaics formulate a competitive photovoltaic on-grid price and feed it back to ETC. EV charging agents compare quotes, determine the optimal power consumption strategy and manage EV charging to maximize benefits. After the charging agent uploads the required power, the user 's EMS formulates. And uploads the energy consumption plan according to the power demand, photovoltaic power generation forecast. And the output of the operator 's electricity and heat price adjustment equipment. After receiving the energy plan, the operator 's EMS adjusts the energy equipment and energy sales quotation to increase revenue. The electricity price adjustment of operators may affect other subjects, prompting them to make corresponding adjustments. ETC will repeat this process until the three parties reach the best trading strategy. The analysis shows that the transaction strategies among energy operators, users with distributed photovoltaics and EV charging agents constitute a three-party non-cooperative game model. Through communication and interaction, the three parties finally reach Nash equilibrium and achieve the best trading state.

3. The Trading Strategy Game Model of the Three Party Market Players

This section introduced in detail the trading strategy game model of three types of market trading entities in the park: energy suppliers, users with distributed photovoltaics, and EV charging agents.

3.1. Park energy operator transaction model

The park 's energy operators are the main participants in market transactions, usually only one. As the main energy supplier, operators are responsible for providing continuous and stable power and thermal energy services. When energy operators optimize the allocation of energy, they aim to maximize their own net income, which is the difference between electricity sales revenue and operating costs. In the case of fixed energy demand of users and electric vehicle charging agents, operators can reduce operating costs by scheduling energy supply equipment. Or adjust electricity sales prices to obtain higher returns.

The revenue of energy operators refers to the difference between the revenue from selling energy to users and EV agents and their own operating costs is shown as Eq.(6).

$$R_{opr} = C_{opr}^{sale} - C_{opr}^{buy} - C_{opr}^{op} \quad (6)$$

Where P_{opr} is a gain for energy operators.

C_{opr}^{sale} refers to the energy sales revenue of energy operators, specifically from the sales of electricity and heat to other loads,which is shown as Eq.(7).

$$C_{opr}^{sale} = \sum_{t=1}^{T} (p_{e,opr}(t)P_{e,opr}(t) + p_{h,opr}(t)P_{h,opr}(t)\Delta t) \quad (7)$$

Where $P_{e,opr}$ and $P_{h,opr}$ are the power to supply electric energy and thermal energy to the park at each time ; δt is the scheduling time interval.

C_{opr}^{buy} refers to the cost of energy operators purchasing energy from power grid companies and natural gas companies,which is shown as Eq.(8).

$$C_{opr}^{buy} = \sum_{t=1}^{T} (P_{grid}(t)P_{grid}(t) + P_{gas}(t)q_{gas}(t))\Delta t \quad (8)$$

Where P_{grid} is the price of electricity sold by the grid company;p_{gas} is the price at which gas sells gas to gas companies;P_{grid} is the electric power injected into the grid side at any time;q_{gas} is the unit gas purchase at any scheduling time.

C_{opr}^{op} refers to the operation and maintenance cost of equipment sold by energy operators, which is composed of electricity sales operation and maintenance cost $C_{h,opr}^{op}$ and heat sales operation and maintenance cost $C_{h,opr}^{op}$,which is shown as Eq.(9).

$$\begin{cases} C_{opr}^{op} = C_{e,opr}^{op} + C_{h,opr}^{op} \\ C_{e,opr}^{op} = \sum_{t=1}^{T} (c_{e,opr}^{T} P_{e,opr}^{T}(t) + c_{e,opr}^{CHP} P_{e,opr}^{CHP}(t)) \Delta t \\ C_{h,opr}^{op} = \sum_{t=1}^{T} (c_{e,opr}^{CHP} P_{h,opr}^{CHP}(t) + c_{h,opr}^{GB} P_{h,opr}^{GB}(t)) \Delta t \end{cases} \quad (9)$$

In the formula, $C_{e,opr}^{T}$, $C_{e,opr}^{CHP}$, $C_{h,opr}^{CHP}$, $C_{h,opr}^{GB}$ are the unit capacity operation costs of energy operator transformers, CHP units, and gas boilers. $P_{e,opr}^{T}$, $P_{e,opr}^{CHP}$, $P_{h,opr}^{CHP}$, $P_{h,opr}^{GB}$ are the output values of energy operator transformers, CHP units, and gas boilers at each moment.

The power supplied by energy operators to the outside world should be balanced with the output power of their own production capacity equipment in real time,which is shown as Eq.(10).

$$\begin{cases} P_{e,opr}(t) = P_{e,opr}^{T}(t) + P_{e,opr}^{CHP}(t) \\ P_{h,opr}(t) = P_{h,opr}^{GB}(t) + P_{h,opr}^{CHP}(t) \end{cases} \quad (10)$$

Energy operators cannot adjust their quotations without limit in the pricing process, and they need to formulate within a certain market constraint range,which is shown as Eq.(11).

$$\begin{cases} P_{e,opr}^{\min} \le P_{e,opr}(t) \le P_{e,opr}^{\max} \\ P_{h,opr}^{\min} \le P_{h,opr}(t) \le P_{h,opr}^{\max} \end{cases} \quad (11)$$

Where, $P_{e,opr}^{\min}$, $P_{e,opr}^{\max}$ represents the upper and lower limit constraints of the electricity selling price; $P_{h,opr}^{\min}$, $P_{h,opr}^{\min}$ represents the upper and lower limit constraints of heating price.

3.2. User transaction model with distributed photovoltaic

When users with distributed photovoltaics formulate energy use strategies, the goal is to maximize revenue. Users can increase revenue by adjusting the electricity price and equipment operation status. When photovoltaic power generation meets its own needs and has a surplus, users can formulate competitive quotations to win the power sales right of electric vehicle charging agents, so as to obtain income. In addition, after receiving the charging demand, the user 's energy management system (EMS) will schedule the equipment according to the photovoltaic power generation forecast. And the electricity sales quotation to reduce the operating cost. The user 's income with distributed PV refers to the difference between the income from selling electricity to EV power agents and the operating cost is shown as Eq.(12).

$$R_{user} = C_{user}^{sale} - C_{user}^{buy} - C_{user}^{op} \quad (12)$$

Where R_{user} is user revenue.

C_{user}^{sale} refers to the electricity sales revenue of users with distributed photovoltaics, including two parts. Part of the revenue comes from selling electricity to EV charging agents, and the other part comes from government subsidies for full-power PV access,which is shown as Eq.(13).

$$C_{user}^{sale} = \sum_{t=1}^{T} (P_{e,user}(t) P_{e,user}^{ch}(t) + p_{alow} P_{e,user}^{PV}(t)) \Delta t \quad (13)$$

Where $P_{e,user}^{ch}$ is the power supplied to the EV charging agent at any time, and P_{alow} is the government 's subsidy price for photovoltaic power generation.

C_{user}^{buy} refers to the cost of users with distributed photovoltaics purchasing additional electrical and thermal energy from energy operators,which is shown as Eq.(14).

$$C_{user}^{buy} = \sum_{t=1}^{T} (P_{e,opt}(t) P_{e,opr}^{user}(t) + p_{h,opr} P_{h,opr}(t)) \Delta t \quad (14)$$

Where $P_{e,opr}^{user}$ is the power that energy operators sell to users at any time.

C_{user}^{op} refers to the operation and maintenance cost of user equipment with distributed photovoltaic,which is shown as Eq.(15).

$$C_{e,user}^{op} = \sum_{t=1}^{T} (C_{e,user}^{pv}(t) P_{e,user}^{PV}(t) + C_{e,user}^{AC} P_{e,users}^{AC}(t)) \Delta t \quad (15)$$

Where $C_{\mathrm{e,user}}^{PV}$ and $C_{e,user}^{AC}$ are the unit capacity operation cost of photovoltaic power generation and electric air conditioning respectively; $P_{e,user}^{PV}$ and $P_{e,users}^{AC}$ are the output values of photovoltaic power generation and electric air conditioning at any time.

3.3. EV charging agent transaction model

The electric vehicle charging agent is responsible for charging the electric vehicle to meet the needs of the owner. The energy management system (EMS) of the agent selects the power supplier with a lower quotation. If the power supply is insufficient, it is supplemented by the other party with a higher quotation. Energy operators can usually provide sufficient energy and meet all needs during low-cost periods. While users with distributed photovoltaics preferentially meet their own needs. And only sell electricity to charging agents when photovoltaic power generation is surplus. If the demand for electric vehicles is large, even if the photovoltaic power generation is fully supplied, it may still be unable to meet the demand. At this time, it is necessary to purchase the gap power from energy operators at a higher price.

C_{ch}^{buy} indicates the charging cost of EV charging agents, which is the sum of the cost of purchasing electricity from energy operators and users,which is shown as Eq.(16).

$$C_{ch}^{buy} = \sum_{t=1}^{T} (P_{e,opt}(t) P_{e,opr}^{ch}(t) + p_{e,user} P_{e,user}^{ch}(t)) \Delta t \quad (16)$$

Because the goal of the game is to maximize the utility of the game, the utility function of the EV charging agent is defined as Eq.(17).

$$R_{ch} = -C_{ch}^{buy}(t) \quad (17)$$

The power value of the input agent at any time should be the sum of the electric power provided by the energy operator and the user,which is shown as Eq.(18).

$$P_{e,opr}^{ch}(t) + P_{e,user}^{ch}(t) = \sum_{i=1}^{N} P_{ch}^{i}(t) \quad (18)$$

Where $P_{ch}^{i}(t)$ represents the charging power of the i th EV at that time, and N represents the total number of EVs managed by the agent.

Each EV charging needs to reach the preset power to meet the traffic demand of the owner,which is shown as Eq.(19).

$$Ei(t_{i,dept}) = E_{i,set} = E_i(t_{i,arr}) + \eta_{ch}^{i} \Delta t \cdot \sum_{t=1}^{T} P_{ch}^{i}(t) \quad (19)$$

Where E represents the power function of EV, and Ei, set represents the final charging power required by the owner;$t_{i,arr}$ and $t_{i,dept}$ indicate the arrival and departure time of the i th EV; η_{ch}^{i} indicates the vehicle charging efficiency of the i th EV.

4. Algorithm Principle

The three market trading entities constitute a three-party non-cooperative game model. The three goals are to rationally pursue their own maximum returns. This paper uses particle swarm optimization (PSO) to solve the problem.

PSO is a classical swarm intelligence algorithm inspired by the flight and foraging behavior of birds. Birds find the global optimal solution through information interaction between individuals. PSO regards the problem to be optimized as a flock of birds, the solution space is regarded as the flight space of the flock of birds. And the position of each bird represents a solution. In this paper, the PSO algorithm simplifies the original 1×5T dimensional problem into 1×3T dimensional, which significantly reduces the difficulty of optimal particle search in the iterative process. The particle swarm is composed of n particles, and the position of each particle is 3T dimension vector a, which represents the potential game strategy set. The speed is 3T dimension vector; the local optimal strategy is 3T-dimensional vector pbest, and the global optimal strategy is 3T-dimensional vector gbest. In the process of algorithm evolution, particles track the optimal position of individual history and the optimal position of population history.

The individual position change of particle swarm optimization algorithm is based on two basic formulas,which are shown as Eq.(20) and Eq.(21).

$$v_{id}^{t+1} = \omega v_{id}^{t} + c_1 r_1 (p_{id}^{t} - x_{id}^{t}) + c_2 r_2 (p_{gd}^{t} - x_{id}^{t}) \quad (20)$$

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \quad (21)$$

Where r1 and r2 are random numbers between (0,1), c1 and c2 represent learning factors, and the value is generally c1 = c2 = 2.

In order to solve the problem of premature convergence in the traditional particle swarm optimization algorithm, this paper adopts a linear decreasing strategy for the inertia weight. In the early stage of iteration, the larger inertia weight helps to enhance the global search ability and jump out of the local optimum. At the end of the iteration, the reduced inertia weight is conducive to accurate search near the global optimum, thus promoting the convergence of the algorithm. In addition, the learning factor adopts a nonlinear inverse cosine acceleration strategy, so that the particles mainly refer to their own historical information in the early stage. And focus more on group information in the later stage to avoid falling into local convergence. The value of inertia weight and learning factor c1, c2 is shown as Eq.(22).

$$\begin{cases} \omega = \omega_s - \dfrac{k}{k_{\max}}(\omega_s - \omega_e) \\ c_1 = c_{1e} + (c_{1s} - c_{1e})[1 - \dfrac{\arccos(-2k/k_{\max}+1)}{\pi}] \\ c_2 = c_{2e} + (c_{2s} - c_{2e})[1 - \dfrac{\arccos(-2k/k_{\max}+1)}{\pi}] \end{cases} \quad (22)$$

In the formula:k_{max} is the maximum number of iterations, ω_s and ω_e represent the initial and final values of the iteration of the inertia weight respectively.c_s and c_e represent the initial and final values of the iteration of the learning factor respectively. The fitness function in the PSO algorithm of the three-party game is shown as Eq.(23).

$$fitness(\varphi) = \max\{|\Delta R_{opr}|, 0\} + \max\{|\Delta R_{user}|, 0\} \quad (23)$$

The algorithm flow diagram of solving the equilibrium solution of the game model in this paper is shown in Fig.3.

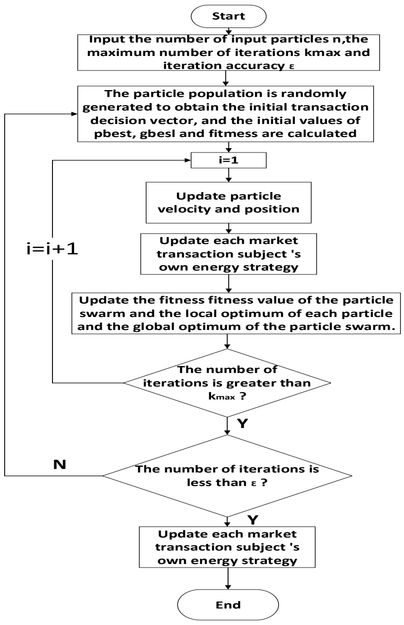


Fig.3 Algorithm flow chart

5. Example Analysis

5.1. Parameter setting

In this paper, the IES of a business park shown in fig.1 is taken as the object for simulation calculation. The park has an energy operator, a user cluster with photovoltaics and an electric vehicle charging agent. The PV-containing user cluster includes five commercial buildings with distributed PV panels. The equipment parameters of the integrated energy park are shown in Table 2.

Trade subject	Equipment name	Num -ber	Parameter	Value
Energy operator	CHP unit	1	Maximum power supply/kW	500
			Maximum heating power/kW	650
			ratio of heat to electricity	1.3
			transmission efficiency/%	35
			heating efficiency/%	45.5
			maintenance cost/(RMB/kW·h)	0.070
			Uphill and downhill climb/(kW/min)	5
	Gas fired boiler	1	Maximum heating power/kW	500
			heating efficiency/%	90
			maintenance cost/(RMB/kW·h)	0.002
			Uphill and downhill climb/(kW/min)	5
	Distributio -n transforme -r	1	Maximum power supply/kW	500
			maintenance cost/(RMB/kW·h)	0.002
Users with distribut -ed photovo -ltaics	Photovolta -ic panel	5	Maximum power supply/kW	200
			energy generation subsidy/(RMB/kW·h)	0.3
			maintenance cost/(RMB/kW·h)	0.006
	Heating air conditioni- ng	5	Maximum heating power/kW	108
			energy efficiency ratio	2.7
			maintenance cost/(RMB/kW·h)	0.002
EV chargin- g agents	EV	40	Maximum charging power/kW	7
			charge efficiency/%	90
			Power battery capacity(kW·h)	24
			Maximum state of charge	0.95

Table 2. Equipment parameters of integrated energy park

5.2. Simulation results and analysis

The fitness curve of the particle swarm optimization algorithm is shown in Fig4.

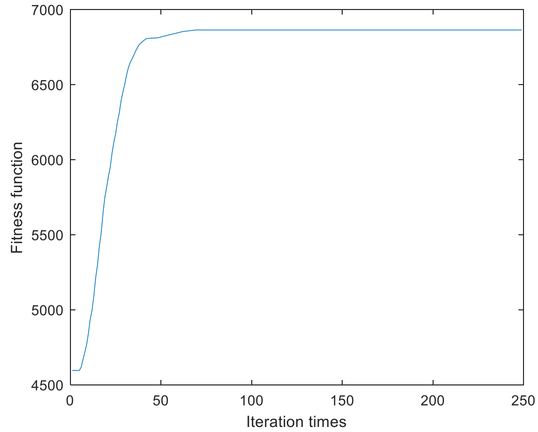


Fig.4 Particle swarm algorithm fitness curve

The electric load, heat load and photovoltaic output of the park are shown in fig.5.

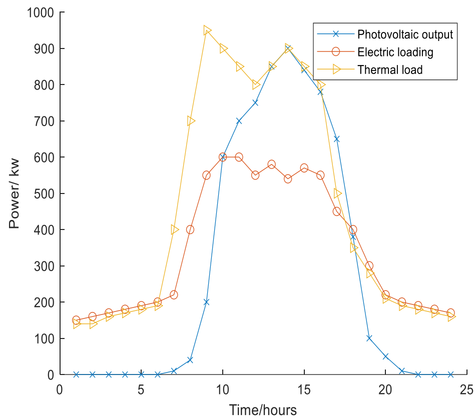


Fig.5 The park's electric load, heat load, and photovoltaic output.

The output of photovoltaic power generation reaches its peak during the daytime (about 09 : 00 to 17 : 00). And the electrical load coincides with the output of photovoltaic power generation during the peak period of the daytime (about 12 : 00 to 15 : 00). Indicating that users have a higher demand for electricity during this period. The change of thermal load during the day is relatively stable. And the overall level is lower than that of electric load and photovoltaic output, indicating that the thermal energy demand is low during this period.

In the game equilibrium state, the power and heat sales of energy operators adopt the time-of-use pricing strategy. And the power sales of user groups adopt the time-of-use electricity price. The bidding results of energy operators and users are shown in Fig.6.

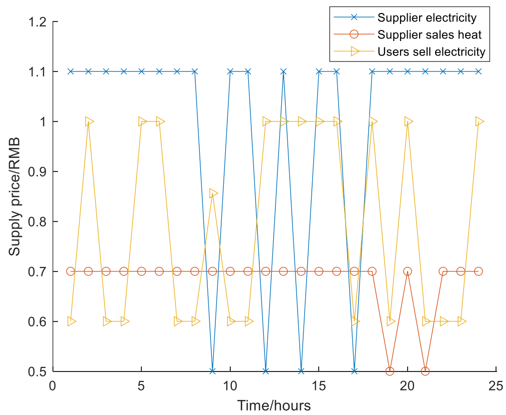


Fig.6.The bidding results of energy operators and users

From the diagram, it can be seen that in most of the time period, the user 's electricity sales quotation is generally lower than the operator 's quotation. Due to the low electricity quotation of photovoltaic users, EV charging agents are more inclined to choose the photovoltaic power sold by users to reduce costs. If users do not take price incentives during the high incidence of photovoltaics, it may lead to a situation of ' oversupply '. Resulting in waste of photovoltaic power generation, thus affecting their own income. Therefore, the user 's bidding strategy should be adjusted in the direction of improving their own income. In addition, the heat quotation of energy operators is relatively low. This is because if the heat price is too high, users will be more inclined to use their own photovoltaic power generation to drive the electric air conditioning for heating. Thereby reducing the demand for heat energy from energy operators. This situation may reduce the potential heat sales revenue of operators. Therefore, they will adjust the heat bidding strategy to keep the heat price at a low level during this period[3].

When the game is balanced, the output of the energy operator 's power supply and heating system is shown in Fig.7.

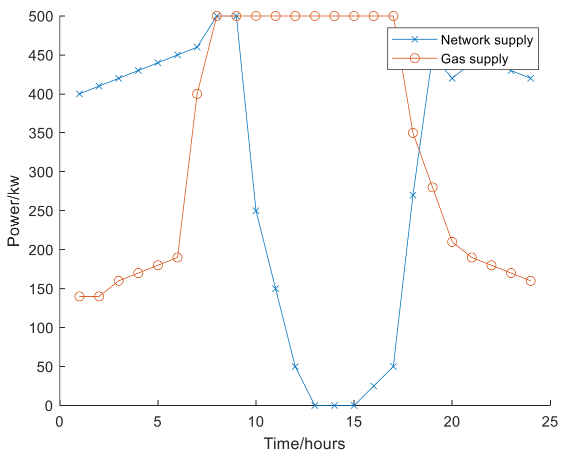


Fig.7 The output of the energy operator 's power supply and heating system

The power supply of the power grid remains relatively stable for most of the time. And the gas heating fluctuates greatly during the day, especially in the morning and evening, and the power drops to near zero at the trough. The two can complement each other to meet the different needs of users.

The power supply of the power grid to the charging station and the power supply of the user to the charging station are shown in Fig.8.

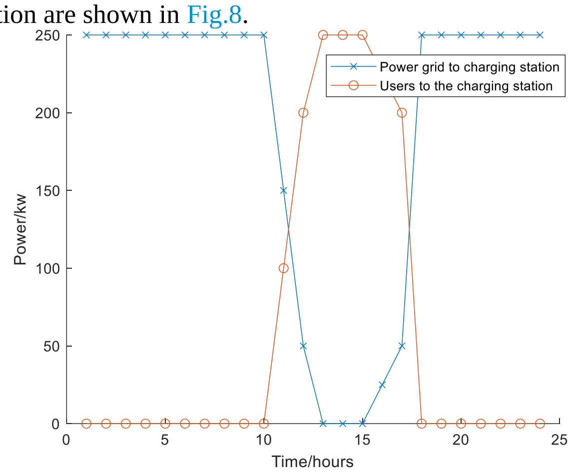


Fig.8 The power grid and users supply energy to the charging station.

From the diagram, it can be seen that during the period of large photovoltaic output, the power of the charging station is mainly provided by the user, and the rest of the time is mainly provided by the power grid.

The load of the grid supply charging station, the user 's electric load and the heat load of the heating network supply user are shown in Fig9.

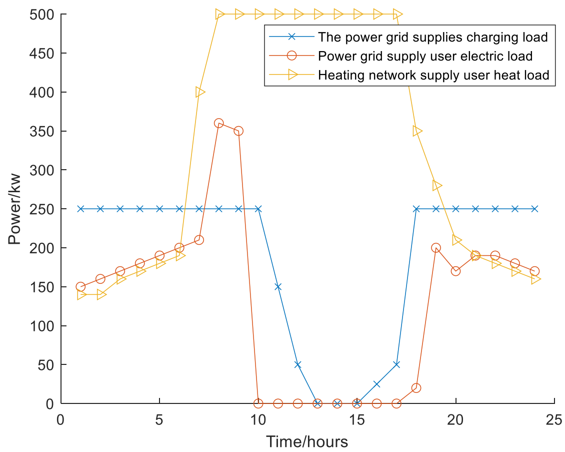


Fig.9 The grid and users supply each other's loads

The charging load of the power grid supply peaked during the day (about 09 : 00 to 18 : 00), indicating that the charging demand was higher during this period. The user 's electrical load has obvious peaks in the morning and evening. Especially at 08 : 00 and 18 : 00, indicating that the user has a greater demand for electricity in these two periods. The heat load is relatively low throughout the period, and the change is not large, indicating that the heat demand is relatively stable during this period.

Finally, when the game is balanced, the operator 's energy sales income is 15265.5RMB. Minus the cost of equipment operation and maintenance and the cost of purchasing energy from the distribution network and the gas network. The full-day net income is 1075.38RMB. The user 's energy sales income is 3268.69 RMB, and the user 's energy cost is 10840.8RMB. For EV charging agents, the all-day charging cost is 224.47RMB. If EVs do not participate in the market game, they are all connected to the distribution network in a disorderly charging state. According to the calculation, the charging cost is 480.54RMB;if all the power of the energy operator is used, the charging cost is 325.16RMB.

6. Conclusion

This paper studied the market trading framework of the integrated energy park, including energy operators, distributed photovoltaic users, and EV charging agents. And established a three-party non-cooperative game model. The studies showed that integrated energy parks could significantly reduce energy costs and improve the flexibility of clean energy consumption. The three-party game model promoted the balance of benefits among all parties, facilitated the effective utilization of photovoltaic resources, and reduced user energy costs. Energy operators profited by selling electricity and heat, distributed PV users profited by supplying energy to EV agents, and receiving PV subsidies. While EV agents reduced charging costs and helped users absorb excess PV resources.

References

1. YANG Zheng, PENG Sicheng, LIAO Qingfen, et al. Non-cooperative Trading Method for Three Market Entities in Integrated Community Energy System. *Automation of Electric Power Systems*, 2018, 42(14):32-39.
2. WANG Cheng,LIU Nian,CHENG Minyang,et al.Stackelberg game based optimal pricing model for photovoltaicprosumer cluster. *Automation of Electric Power Systems*,2017, 41(12):146-153.
3. YU M, HONG S H, A realtime demand-response algorithmfor smart grids: a Stackelberg game approach. *IEEETransactions on Smart Grid*,2017,7(2):879-888.

Authors Introduction

Ms. Peng Wang

She is a postgraduate tutor of Tianjin University of Science and Technology.In 2014, she received a doctorate from North China Electric Power University.The research direction is the functional safety assessment of safety instrumented systems.

Ms. Siyi Wang

She is studying Automation at Tianjin University of Science and Technology and is currently pursuing a Master's degree. Her research area is about deep learning.

Mr. Liangyu Wang

He is currently pursuing a Bachelor of Engineering degree in the College of Electronic Information and Automation at Tianjin University of Science and Technology.

Mr. Chengkai Miao

He is a sophomore in the School of Electronic Information and Automation at Tianjin University of Science and Technology. He was admitted to Tianjin University of Science and Technology in 2023. He is currently engaged in his undergraduate studies in this university.

Research on the Sensitivity of Thermal Comfort Using Sensitivity Algorithms Based on Variance and Stochastic Expansion

Peng Wang*, Yuting Zhou, Liangyu Wang, Chengkai Miao, Qikun Wang

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

*E-mail: *2324365941@qq.com*

www.tust.edu.cn
(Tianjin University of Science and Technology, Tianjin, China)

Abstract

In the research field of modern architectural environment, the sensitivity research of human thermal comfort factors is of crucial significance. In this paper, first of all, common algorithms in the field of sensitivity analysis and data sets for thermal comfort research are elaborated in detail. Secondly. The variance method is taken into consideration. It has the capacity to reflect the fluctuation of the influence that different factors exert on the results. Meanwhile, the stochastic expansion method is also regarded. It is capable of handling complex non-linear relationships. A decision is made to combine these two methods. And the combined methods will be applied to conduct the sensitivity analysis of thermal comfort factors. Finally, the most critical factors for thermal comfort are successfully identified, providing an important basis for the construction and optimization of the thermal comfort prediction model.

Keywords: Sensitivity analysis, variance, stochastic expansion, thermal comfort

1. Introduction

With the continuous improvement of people 's requirements for the quality of building environment, human thermal comfort has become the key content of building environment research[1]. The physical and mental health of users is not only affected by thermal comfort, but it is also closely related to building energy consumption. Identifying the key factors affecting human thermal comfort accurately is of great significance for optimizing architectural design and improving energy efficiency. Sensitivity analysis, as an effective tool, is capable of quantifying the impact of various factors on thermal comfort, while the variance method and the random expansion method each have their own advantages in dealing with complex data relationships. The combination of these two methods is expected to provide more accurate results for the analysis of thermal comfort factors.

The rest of this article is organized as follows. The second chapter introduces the specific algorithms and data sets used. In the third chapter, we conduct sensitivity analysis on the data sets using two algorithms respectively and draw conclusions. In the last chapter, we summarize the full text.

2. Related Algorithms and Data Sets in Thermal Comfort Research

In the field of statistics, there are numerous studies on sensitivity algorithms. In this paper, the global sensitivity analysis based on variance and the sensitivity analysis based on the stochastic expansion method are mainly adopted.

2.1. Variance-based sobol global sensitivity analysis

The Sobol global sensitivity analysis was proposed by Ilya M. Sobol. It has been widely applied in various fields. For instance, in environmental science, it is used to evaluate the impacts of meteorological factors and pollution source emission parameters on the spatial distribution of pollutant concentrations; in engineering system design, it helps to identify crucial design parameters; in financial risk assessment, it is employed to analyze the effects of economic factors on the returns of investment portfolios.

The main indices of this method include the Sobol total effect index (ST), which measures the contribution of an input factor (including itself and all its interactions with other factors) to the total variance of the output variable, and the Sobol first-order effect index (S1), which reflects the contribution of the change of a single input factor itself to the variance of the output variable.

The advantages of the Sobol global sensitivity analysis are as follows: Firstly, it has strong comprehensiveness as it can consider the entire range of values of input factors rather than just local changes, thus comprehensively assessing the impacts of factors on the output results. Secondly, it can effectively quantify the impacts of interactions among input factors on the output results. In many complex systems, the interactions among factors may have significant influences on the system performance, and this method is capable of quantifying these influences. Thirdly, it is based on the variance decomposition theory, possessing a solid theoretical foundation, and the results are highly interpretable. That is to say, researchers can accurately judge the importance of each factor in the system according to the calculated results of the Sobol indices.

2.2. Sensitivity analysis based on stochastic xxpansion method (PCE Model)

The sensitivity analysis based on the stochastic expansion method (PCE model) has been gradually developed and refined by scholars in relevant research. It is widely applied in various fields such as engineering and physics. For instance, it is used to evaluate the impacts of uncertain factors in structural reliability analysis.The main indicators include the contribution degrees of various coefficients to the results. Its advantages lie in the fact that it can effectively handle complex nonlinear relationships and approximate complex models with relatively fewer computational resources. It can also take into account multiple uncertain factors and their interactions.However, for high-dimensional and strongly nonlinear problems, there may be insufficient accuracy. There are certain difficulties in determining appropriate basis functions and solving coefficients when constructing the model. Moreover, it has certain requirements on the data distribution, and the accuracy of the results may be affected when the data characteristics are not favorable.

2.3. Introduction to data sets

The ASHRAE Global Thermal Comfort Database II (ASHRAE GTDB-II) is sourced from the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE)[2]. This database is a global thermal comfort database that has collected data on human perception of ambient temperature on a global scale. It is designed to provide a large amount of data regarding human perception of ambient temperature to support thermal comfort research and development. The ASHRAE GTDB-II database encompasses different environmental factors such as temperature, humidity, wind speed, illuminance, etc., as well as human evaluations of the environment. These data can be utilized to assess and optimize the thermal comfort performance of buildings, improve building design, and develop new thermal comfort assessment tools and techniques. The ASHRAE GTDB-II database is an important resource for thermal comfort research and development and is one of the significant tools in the fields of building design, system selection, and operational assessment. It is of great significance for ensuring the comfort and energy efficiency of buildings and helps to enhance the energy-saving efficiency of buildings and the satisfaction of occupants.

The ASHRAE GTDB-II encompasses climate data on a global scale, including data from North America, South America, Europe, Asia, Africa, Australia, and Antarctica. These regions were selected by ASHRAE for data collection because they cover a wide range of climate types, such as temperate, tropical, and subtropical climates. The extensiveness and richness of the climate data from these regions can better meet the requirements of thermal comfort prediction. Additionally, factors such as building types, building heights, and building materials in these regions all have an impact on thermal comfort prediction, and the diversity of the data from these regions can better illustrate the influence of different influencing factors on thermal comfort.

3. Experimental Design for Sensitivity Analysis

In this chapter, we first preprocessed and separated the data. Secondly, we analyzed the preprocessed data by using two sensitivity methods.

3.1. Data preprocessing

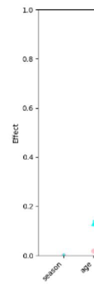
In the research field of building environment and human thermal comfort, considering that significant differences exist in the age structure, activity characteristics, and specific thermal comfort requirements among occupants in different building types, a detailed and systematic data processing procedure was carried out in this study. Firstly, the data were divided into three categories, namely office, classroom, and multi-family residential, based on building types, thereby constructing three independent data subsets. After the classification was successfully completed, null value filling was immediately performed on each sub-table. For data columns closely related to climate types, such as temperature, humidity, and clothing thermal resistance, when the season information could be clearly determined, the mode of the data in that season was used for filling to best match the typical characteristics of the season. If the season information could not be determined, the mean value was used for filling to ensure the relative integrity and rationality of the data. For data columns that have no direct correlation with climate seasons, such as age, height, weight, and metabolic rate, a unified mean value filling strategy was adopted to compensate for the information gap caused by missing data. In addition, for columns marked with non-numeric labels, such as thermal sensation evaluations (e.g., cold, hot) and gender information, they were all converted into digital codes to facilitate subsequent data analysis and model construction. Through the above rigorous and comprehensive data processing procedures, a substantial and reliable data resource was finally obtained. Specifically, 77,260 records of office data, 11,704 records of classroom data, and 13,984 records of residential data are available, as shown in Table 1. A solid data foundation is laid for further exploring the characteristics related to thermal comfort in different building types.

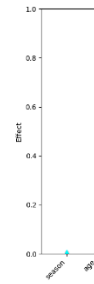
Table 1 data classification

type of construction	Num
office	77,260
classroom	11,704
multifamily houses	13,984

3.2. Sensitivity analysis based on variance

In this experiment, the machine learning model is first trained, and 18 input features such as season, age, gender,





ht, wt, ta, top, tr, tg, rh, vel, met, clo, t_out, rh_out, fan, window, door and thermal_sensation are selected as output features. The processed data is divided into training set and test set, and the proportion of 80 %-20 % is divided. Then, the training set data is used to train the Cubist model, and the parameters of the model are adjusted to optimize the performance of the model. The optimal parameter combination is determined by cross-validation, which makes the model have better fitting effect on the training set and avoids over-fitting. The trained Cubist model is evaluated using the test set data, and the mean square error (MSE) of the model is calculated as the evaluation index, as shown in Table 2.

Table 2 Cubist model indicators

type of construction	Mean Squared Error
office	1.03288586766953
classroom	1.06793067981422
multifamily houses	0.789635208548701

Secondly, the Sobol global sensitivity analysis is carried out by using the constructed Cubist model. In this paper, Sobol sequence sampling is used to generate a large number of input feature sample combinations, which will be used for subsequent sensitivity analysis and calculation. For the input feature sample combination obtained by each sampling, the corresponding output prediction value is calculated by the Cubist model. According to the variance decomposition theory, the main effect index (S1) and the total effect index (ST) of each input feature are calculated. The main effect index measures the contribution of a single input feature 's own change to the output variance, and the full effect index takes into account the contribution of the feature and all its interactions with other features to the total output variance. The main effect and full effect index results of each input feature in three different data sets are sorted out and recorded, so as to analyze and discuss the results in the future, and further understand the influence degree and interaction relationship of each input feature on the thermal _ sensation output feature. Fig. 1, Fig. 2 and Fig. 3 show the main effects and full effects of the three types of data input features such as office, classroom, and multi-family residential.

The data show that in the office building type, age, gender, ta, top, te, tg, clo, including t _ out, rh _ out and other factors have a greater impact on the results, and the basic total effect is greater than the main effect. Among the clssroom building types, only ta, top, tr, rh and t _ out have a greater impact on thermal sensation. In the multifamily houses building type, season, age, ta, top and t _ out have a greater impact. On the whole, the influence of temperature factors on thermal sensation in all buildings is relatively large, while for other factors, different building types have different sensitivities.

3.3. Sensitivity analysis ased on the PCE model

In this experiment, season, age, gender, ht, wt, ta, top, tr, tg, rh, vel, met, clo, t _ out, rh _ out, fan, window, door and other input features and thermal _ sensation are still selected as output features.

Firstly, the PCE model is constructed. Since we are dealing with nonlinear problems, we choose the second-order model, and the basis function is the Legendre polynomial. By substituting the sampled data into the expression of the PCE model, a linear equation group is obtained. The coefficients of the PCE model are obtained by solving the linear equations by the least square method. The objective of the least squares method is to minimize the sum of squared errors between the model predictions and the actual observations, namely,

$$\min \sum_{k=1}^{n} (y_k - \hat{y}_k)^2 \tag{1}$$

where y_k is the actual observed value, $\hat{y}_k$ is the predicted value of the model, and n is the number of sample points. Independent validation data sets were used to evaluate the accuracy of the PCE model. The verification data is

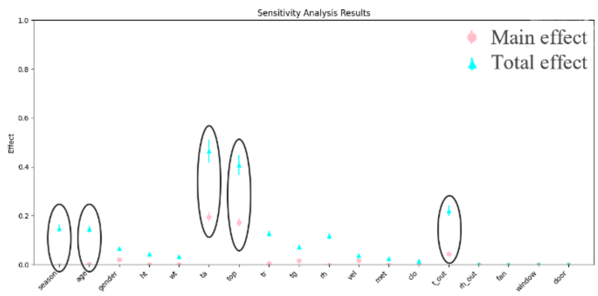


Fig. 3 multifamily houses- Sensitivity Analysis

substituted into the PCE model to calculate the prediction error.

Secondly, we conduct sensitivity analysis. The first-order sensitivity index (S1) measures the contribution of a single input variable to the variance of the output variable.

$$S1_i = \frac{V\left[E\left(y \middle| x_i\right)\right]}{V[y]} \tag{2}$$

Where $V\left[E\left(y|x_i\right)\right]$ is the variance of the conditional y expection under the given conditions of x_i, and $V[y]$ is the total variance of y, The total sensitivity index (ST) measures the contribution of the input variable and all its interactions with other variables to the variance of the output variable.

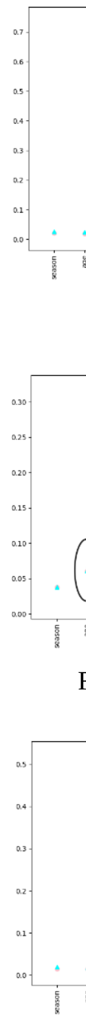


Fig. 6

$$ST_i = 1 - \frac{V\left[E\left(y|x_{-i}\right)\right]}{V[y]} \quad (3)$$

Where $V\left[E\left(y|x_{-i}\right)\right]$ is the conditional expectation variance of y under all other variable conditions except x_i. For the PCE model, S1 can be calculated by mathematical derivation of the coefficients and basis functions. According to the calculated S1 and ST indicators, the input variables are sorted. Variables with larger S1 and ST values have a greater impact on the output variables and are key sensitive factors. Compare the values of S1 and ST to understand the intensity of interaction between variables. If ST is much larger than S1, it shows that there is a strong interaction between the variable and other variables. The impact on the output variable is not only its own individual effect, but also the synergistic effect with other variables.

Fig4, Fig5 and Fig6 show the main effects and full effects of the three types of data input features based on the PCE model, such as office, classroom, and multi-family residential.

Through a detailed analysis of the data, we have obtained some important findings. Among all building types, the factor of "ta" has a relatively significant impact on the final results. Further observation reveals that the difference between its total effect and main effect is not obvious. When focusing on the "classroom" building type, factors such as age, gender, "top" and "tr" also have a considerable influence on the results. Among them, "age" may have an impact on the results related to the building due to the differences in usage habits among people of different age groups. "Gender" may affect the results because of the differences in spatial needs and usage preferences between men and women. The characteristics related to the top of the building represented by "top" and the specific building attributes or conditions represented by "tr" all play important roles among the factors influencing the results of the "classroom" building type.

4. Conclusion

This paper is mainly in the field of modern building environment research, and it is of great significance to study the sensitivity of human thermal comfort factors. Through research, it can be seen that common algorithms in the field of sensitivity analysis and data sets for thermal comfort research are crucial for in-depth exploration. In this study, in view of the fact that the variance method can show the fluctuation of the influence of different factors on the results, and the random expansion method can deal with the complex nonlinear relationship, these two methods have achieved remarkable results in the sensitivity analysis of thermal comfort factors. Finally, the most critical factors for thermal comfort are successfully identified. These results provide an important basis for the construction and optimization of thermal comfort prediction models. It is helpful to consider the influence of key factors more accurately in the subsequent design and regulation of building environment, so as to effectively improve the thermal comfort experience of human body in building environment and promote the development of thermal comfort in building environment to a more scientific and reasonable direction.

Acknowledgements

The research is supported by the major special project of carbon neutralization technology in Tianjin, China in 2024 (24ZXTKSN00070).

References

1. Cheng Chenxi, Xiangtian Deng, Xiaoyong Zhao, et al., Multi-Occupant Dynamic Thermal Comfort Monitoring Robot System. *Building and Environment* 234: 110137.
2. Földváry Ličina, Veronika, Toby Cheung, et al.Thomas Parkinson, Edward Arens, Chungyoon Chun, Development of the ASHRAE Global Thermal Comfort Database II. *Building and Environment* 142:502–12.

Authors Introduction

Ms. Peng Wang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2014, she received a doctorate from North China Electric Power University. The research direction is the functional safety assessment of safety instrumented systems.

Ms.Yuting Zhou

In 2023, she entered Tianjin University of Science and Technology.She is pursuing a Bachelor of Engineering degree in the School of Electronic Information and Automation.

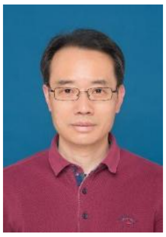
Mr. Liangyu Wang

He is currently pursuing a Bachelor of Engineering degree in the College of Electronic Information and Automation at Tianjin University of Science and Technology.

Mr. Chengkai Miao

He is a sophomore in the School of Electronic Information and Automation at Tianjin University of Science and Technology. He was admitted to Tianjin University of Science and Technology in 2023. He is currently engaged in his undergraduate studies in this university.

Mr. Qikun Wang



In 1996, he received his Bachelor of Engineering degree from the School of Electronic Information and Automation, Tianjin University of Science and Technology, China. He is a senior engineer, research direction is

Deep Learning Based Infant and Child Monitoring System

Peng Wang

College of Electronic Information and Automation, Tianjin University of Science and Technology, Tianjin, China

Jiale Jia

College of Electronic Information and Automation, Tianjin University of Science and Technology, Tianjin, China

E-mail: autowangpeng@tust.edu.cn,15127223153@163.com

Abstract

This paper focuses on a baby monitoring system based on computer vision and multi-branch convolutional neural network, firstly, the collected photos are processed, and then the algorithm is implemented using openCV library to train to get the baby's facial target detection model, and secondly, based on the opencv algorithm and yolov8 algorithm technology to achieve the tracking and analysis of baby's behavioral trajectory and facial detection. Finally, we realized the functions of target tracking, night lightening, image segmentation and baby face detection, and the detection achieved good results.

Keywords: Deep learning, Neural network, Attribute recognition, Yolov8

1. Introduction

With the fast pace of modern life and changes in family structure, more and more parents need to take care of their babies as part of their daily routine. However, it is difficult to remain vigilant at all times while taking care of their babies, especially when they need to fulfill other tasks, which may lead to injuries or mood swings that can have a negative impact on the family. Therefore, the development of a baby safety monitoring system can help parents identify potential problems and take necessary measures in a timely manner, contributing to the safety and emotional stability of their babies.

This project aims to develop a baby monitoring system based on computer vision and multi-branch convolutional neural networks. The system can reduce the blind spot of people's monitoring of infants and young children, and detect the safety factor of infants and young children to help parents monitor the safety and emotional state of their babies.

The rest of this paper is organized as follows: The second section introduces the research on infant facial recognition both domestically and internationally; the third section discusses the technical solutions for detecting infant facial expressions; the fourth section presents the experimental results; and the fifth section summarizes the main content of this paper.

2. Domestic and International Research

Infant facial recognition detection system is a technology that combines computer vision, artificial intelligence and biometrics for recognizing, analyzing and verifying facial features of infants and children. Such technologies have applications in a variety of fields, including child health, behavioral monitoring, and safety and security.

In China, with the rapid development of artificial intelligence technology, infant facial recognition technology has gradually gained attention and application. The research in China mainly focuses on the following areas: infant facial recognition accuracy, biometric identification applications, intelligent security and child protection. In foreign countries, the application of infant facial recognition technology is also growing, mainly focusing on the following areas: facial expression recognition and emotion calculation, medical and health monitoring, and intelligent parenting devices.

3. Technical Program

3.1. Establishment of a database

Crawler technique is used to collect enough baby expressions from multiple data sources to build a baby expression database [1]. Then used the sprite annotation assistant to label the inductive regions of the infant pictures in the database, which are mainly classified into three categories: POSITIVE, PEACE, and PASSIVE. Furthermore,I used rectangular boxes to crop out all the infant facial regions in the database, as a way to reduce the interference of other environmental factors on facial recognition. It is also necessary to categorize all the data images, due to the simplicity of the infant facial expressions, after all the images are labeled, they are exported to XML format. Then it is transferred to the format used by YOLO, through pre-processing, including data enhancement and data division. Data enhancement can be done by random cropping, rotating, scaling, flipping, etc. to expand the data volume and enhance the generalization ability of the model. Data partitioning can divide the dataset into training, validation and test sets for model training and evaluation.

3.2. Training models

Model training was performed using YOLOV8. Using the YOLO algorithm library, after collecting sufficient quantity and quality of data, it is fed into the model to be trained using a number of techniques such as deep neural networks, CNN (convolutional neural networks) [2], data augmentation, optimizers, and GPU acceleration, to build a face recognition model with good generalization performance. After the model is trained, it can be applied to real-time video streams to determine whether infants are in a positive, negative or calm emotional state by analyzing their facial expressions.

3.3. Model evaluation

In target detection, the important metrics for evaluating the performance of a model are Accuracy, Precision, Recall and mean average precision (mAP).

First, calculate the accuracy (Accuracy)

$$Acc = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

Afterwards the accuracy is calculated (Precision):

$$Precision = TP / (TP + FP) \quad (2)$$

Recall is then calculated:

$$Recall = TP / (TP + FN) \quad (3)$$

Where TP is positive class determined as positive, FP is negative class determined as positive, FN is positive class determined as negative, and TN is negative class determined as negative. The higher the recall, the more the model is able to detect the target correctly.

For each category,the Precision and Recall are calculated for different thresholds and plotted on the Precision-Recall curve.On the basis of the Precision-Recall curve,a formal evaluation is obtained by calculating the average of the Precision values corresponding to each Recall value:

$$AP = \sum_{i=1}^{n-1} (r_{i+1} - r_i) P_{inter}(r_i + 1) \quad (4)$$

where r1,r2...rn are the Recall values corresponding to the first interpolation at the first interpolation of the Precison interpolation segment in ascending order. The AP of all categories is the mAP:

$$mAP = \sum_{i=1}^{k} AP_i / k \quad (5)$$

From Fig.1. it can be seen that the higher the average accuracy, the better the model detects under different categories.

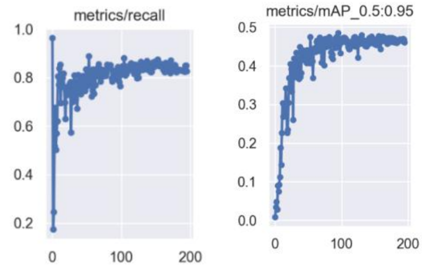


Fig.1 Precision-Recall Curve

3.4. Trace detection

By processing the video stream from the camera to obtain the relevant information about the person and locking it, using the Mean Shift algorithm, the user feature values can be extracted in the subsequent processing of the image to avoid following the target confusion.

The Mean Shift algorithm is a non-parametric clustering algorithm based on kernel density. The algorithm assumes that the datasets of different clusters follow different probability density distributions, finds the direction of the fastest density growth at any local point , and finds the region with high sample density corresponding to the maximum value of the distribution.

Therefore, the Mean Shift algorithm flow is:

(1) Calculate the mean drift vector for each sample: $m_h(x)$

(2) For each sample point with $m_h(x)$ Perform a translation, :

$$x_i = x_i + m_h(x_i) \quad (6)$$

(3) Repeat (1)(2) until the sample points converge:

$$m_h(x) = 0 \quad (7)$$

(4) Samples that converge to the same point are considered members of the same cluster class.

4. Presentation of Results

4.1. Training results

First we simply constructed a confusion matrix as in Fig.2 that facilitates better analysis of important indicators of model performance later.

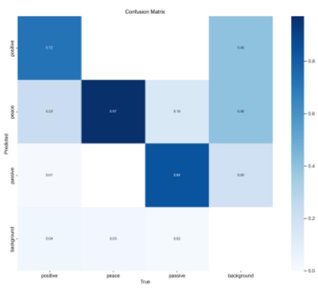


Fig.2 Confusion matrix

After completing the training of the model, we then analyzed the accuracy, precision, and recall. Fig.3 is the F1 curve, positive, peace, and passive three curves are basically the same, with the growth of confidence firstly increased, then stable and finally decreased. Fig.4 and Fig.5 are the curves of precision and recall, respectively.

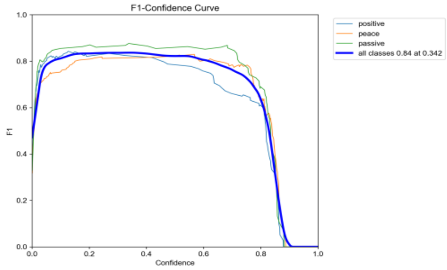


Fig.3 F1 curve

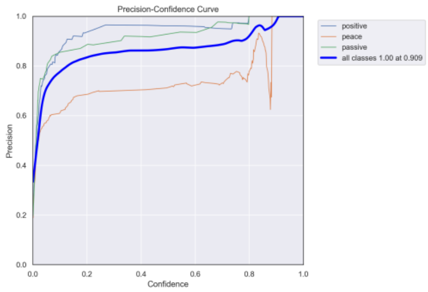


Fig.4 Accuracy curve

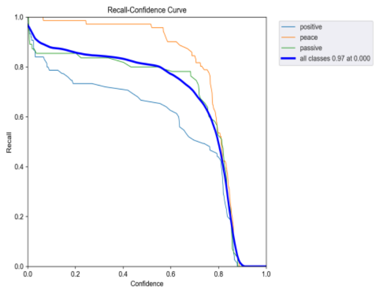


Fig.5 Recall curve

After that the trained model can analyze the expression of each baby photo as in Fig.6 in which each photo indicates the state as well as the degree of the moment.

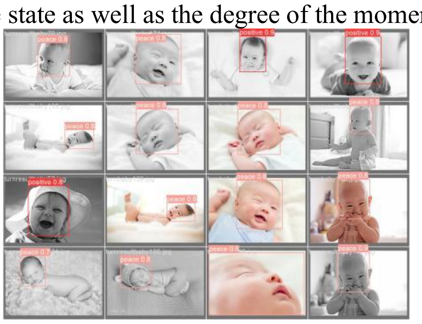


Fig.6 Partial identification

4.2. Functional realization

The full range of core functions of a baby monitoring system, including nighttime light boosting, image segmentation, and target tracking.

(1) Nighttime Light Enhancement

On the premise of the original clarity, the product traverses the pixels to brighten their gray values, adding a nighttime brightening effect to make the picture brighter and more stable in low-light conditions. As Fig.7 shown.

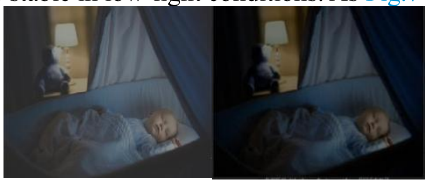


Fig.7 Lightening treatment

(2) Image segmentation

Gradient flow field based image segmentation is accomplished with watershed function to improve the accuracy of face recognition. Using image segmentation to extract the local features of the face can reduce the interference of background information on the face recognition algorithm and improve the accuracy of recognition. As Fig.8 shown.

Fig.8 Segmentation Processing

(3) Goal tracking

The Mean-shift algorithm is mainly used to mean shift the image to track the position of some regions in the video frame.

The facial recognition system continuously tracks and recognizes the baby's face in a video stream or sequence of images and can more accurately locate facial features, thereby improving the accuracy of facial recognition. As shown in Fig.9 .

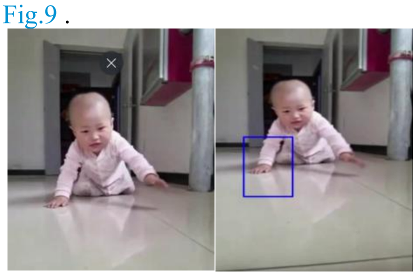


Fig.9 Tracking Processing

5. Conclusion

In this paper, a baby monitoring system is successfully built by establishing a database, training the model and applying the yolo algorithm program, and finally realizing the functions of nighttime light enhancement, image segmentation, target tracking, and baby face detection.

References

1. Simeng Yan, Wenming Zheng, Chuangao Tang, et al, "ARL-IL CNN for Automatic Facial Expression Recognition of Infants under 24 Months of Age." *Journal of Physics: Conference Series*, vol. 1518, 2020, p. 012027
2. Yue Sun, Caifeng Shan, Tao Tan, et al, "Detecting Discomfort in Infants through Facial Expressions." *Physiological Measurement*, vol. 40, no. 11, 2019, p . 115006

Authors Introduction

Ms. Peng Wang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2014, she received a doctorate from North China Electric Power University The research direction is the functional safety assessment of safety instrumented systems.

Mr. Jiale Jia

He was admitted to the School of Electronic Information and Automation at Tianjin University of Science and Technology in 2022. He is currently pursuing his undergraduate studies at Tianjin University of Science and Technology.

Fuzzy-Controlled Multi-Valve Pneumatic Soprano Recorder Auto-Playing and Score Recognition System

Chun-Chieh Wang*
Department of Electrical Engineering, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.

Chung-Wen Hung
Department of Electrical Engineering, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.

Kuo-Hsien Hsia
Bachelor Program in Intelligent Robotics of Electrical Engineering, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.

Chian-Cheng Ho
Department of Electrical Engineering, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.

Ying-Yuan Yao
Graduate School of Intelligent Data Science of Electrical Engineering, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.

E-mail: jasonccw@yuntech.edu.tw, wenhung@yuntech.edu.tw, khhsia@yuntech.edu.tw, futureho@yuntech.edu.tw, anderler899@gmail.com*

Abstract

This study introduces advanced improvements to previous methods for automated score recognition and soprano recorder performance, addressing limitations in playback speed, accuracy, and sound quality. The score recognition system now employs an enhanced method for identifying and removing staff lines, which improves the distinction of musical symbols. In addition, a refined approach is used to classify notes by their placement on the staff, supporting accurate pitch assignment and musical sequencing. The playback mechanism has also been upgraded by replacing traditional motor-driven finger actuators with a solenoid-based multi-valve system, which optimizes performance speed and reduces mechanical noise. The recorder's pitch control now incorporates a 9-valve configuration, offering more precise air pressure regulation across an extended range from bass to treble, thereby addressing pitch stability and tonal accuracy issues. Further, automated tuning is achieved through advanced fuzzy control, which significantly enhances tuning speed and fidelity. Experimental results confirm that these refinements enable substantial improvements in tonal precision and overall performance quality for automated soprano recorder systems.

Keywords: Fuzzy-controlled multi-valve pneumatic, Soprano recorder auto-playing, Enhanced music score recognition, Automated tuning

1. Introduction

With the continuous advancement of robotic technology, its applications have gradually expanded from industrial and military fields to music and arts. The development of automated performance technology, particularly in automatic piano playing, has reached a level comparable to human performance, demonstrating the potential of robots in the arts and music domains. This study aims to develop a multi-valve pneumatic recorder automatic performance and music score recognition system based on fuzzy control to enhance the performance of the existing platform. In the system design, traditional pen-type pneumatic cylinders were replaced with reciprocating electromagnets to reduce noise in the pressing mechanism. Simultaneously, the number of pneumatic valves was increased from three to nine, integrated with fuzzy control technology for automatic adjustment. Furthermore, the system introduced more precise air pressure regulation and multi-phase note classification techniques to address challenges in pitch stability and tuning accuracy, aiming to achieve human-level automation in recorder performance.

In recent years, the application of robotic systems in music has continued to grow. In 2007, Lee W.C. and colleagues proposed a dual-wheel robot capable of autonomously reading music scores and singing using voice, although its score recognition was limited to Arabic numeral formats [1]. In 2016, Hong M.J. developed a mixed media technique applying hand-drawn styles to portrait generation; however, the use of only five colors led to a color-matching error of about 10% [2]. In 2015, Zhang Yonghui et al. adopted Optical Music Recognition (OMR) technology, which, although capable of identifying music symbols, relied on a symbol database, making it prone to errors from minor discrepancies [3]. In 2019, Xiao Zhe studied a real-time OMR system for dulcimer-playing robots, but the decomposition of notes into only three basic elements limited the accuracy of note recognition [4]. Additionally, in 2014, Wang B.R. proposed an iOS-based music score recognition system that required manual assistance, making full automation unachievable [5].

Addressing the shortcomings of the aforementioned studies, C.C. Wang guided students to develop a recorder-playing robot. Initial efforts focused on designing a mechanism capable of replacing human fingers for hole pressing. After completing the initial design, the development of an automatic music score recognition program followed, along with improvements to the pressing mechanism. Lastly, a controller for the blowing component was planned and designed to ensure the machine could produce soft and precise airflow, thereby playing melodious music. However, the first-generation recorder-playing robot encountered issues with insufficient pressing speed and air leakage. Servo motors could not switch pressing actions quickly enough to meet the performance rhythm, while silicone pressing components had inadequate coverage, leading to air leaks due to tolerance issues.

To address these deficiencies, C.C. Wang proposed a pen-type pneumatic cylinder-based pressing mechanism in 2020, incorporating sponge and electric cloth as cushioning elements to effectively reduce air leakage and pressing speed issues [6]. However, the noise generated during pressing by pen-type pneumatic cylinders continued to interfere with performance. Additionally, although the scale classification pneumatic valves were improved to low, medium, and high sections, inaccuracies in scale airflow still resulted in tone-breaking issues. In 2022, C.C. Wang introduced a nine-valve pneumatic system based on fuzzy control to further enhance the precision of scale control, although there remained room for optimization in the fuzzy controller design [7].

Building on the above background and issues, this study, referencing related works [7][8][9], proposes a novel automatic tuning fuzzy control inference for the pneumatic valve control of recorder-playing robots. The goal is to improve the overall performance of automated recorder playing and provide new directions and technical support for future research in music robotics.

2. Soprano recorder auto-playing vehicle

2.1. *Reciprocating electromagnet pressing system*

This study replaces the pen-type pneumatic cylinder with a reciprocating electromagnet, significantly enhancing air pressure stability, reducing air leakage noise, and improving music quality. Additionally, the module's more compact size contributes to streamlined system design and overall stability. Figure 1 shows the soprano recorder auto-playing vehicle, and Figure 2 depicts the reciprocating electromagnet pressing system.

Fig. 1. Soprano recorder auto-playing vehicle

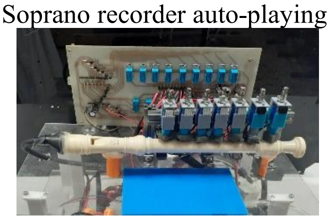


Fig. 2. Reciprocating electromagnet pressing system

2.2. *Scale-classified air valve*

In the section on pitch-classified air valves, the number of pneumatic solenoid valves was increased, resulting in more precise air pressure, reduced risk of note detuning, and enhanced system stability. However, this also increased module space requirements and tuning time, making the operation slightly more complex. Figure 3 illustrates the connection diagram of the third-generation tuning pneumatic valves.

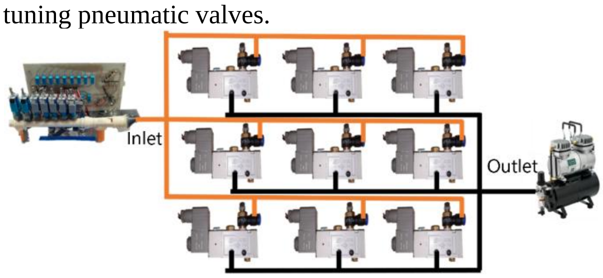


Fig. 3. The connection diagram of the third-generation tuning pneumatic valves

2.3. *Pneumatic valve automatic tuning system*

In the improvement of the pneumatic valve automatic tuning system, 3D-printed connectors were used to integrate the manual pressure adjustment valve with the SAVOX SC-0251 MG servo motor, enabling automatic control and precise adjustments. This enhancement significantly improves the system's accuracy and operational efficiency, while simplifying the user interface. As shown in Figures 4 and 5, the diagrams illustrate the structure of the pneumatic valve automatic tuning system and the playing range divided into nine intervals.

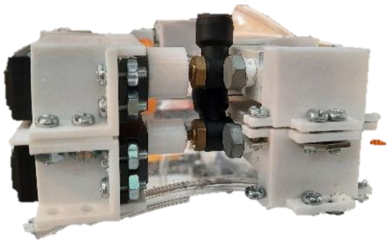


Fig. 4. Pneumatic valve automatic tuning system

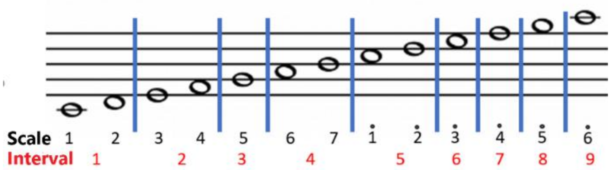


Fig. 5. The playing range divided into nine intervals

2.4. *Manual tuning*

To ensure the accuracy of the automatic tuning, manual tuning is first performed to determine the frequency values corresponding to each scale. Using the Acquire Sound module in LabVIEW, the audio is converted into waveforms, and frequency and amplitude data are obtained through the Tone Measurements module. Figure 6 shows the human-machine interface for sound recognition. This process provides crucial frequency reference data, ensuring the accuracy of the pitch in the automatic flute-playing system, which is essential for maintaining high music quality.

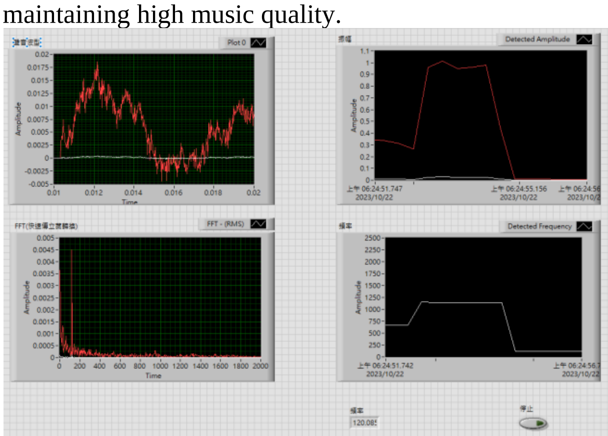


Fig 6. Human-machine interface for sound recognition.

3. Fuzzy-controlled automatic scale tuning method for multi-valve pneumatic systems

3.1. *Fuzzy control*

To enhance the accuracy of musical performance, this work utilizes nine electromagnetic pneumatic valves, each equipped with independent control of the opening degree to meet the airflow demands of different pitch ranges. Each valve is assigned a dedicated fuzzy controller, which takes pitch frequency as input and outputs a correction value for adjusting the valve's opening degree. The fuzzy controller infers the corrected angle, combines it with the preset opening angle, and determines the final opening degree of the valve. For example, if the preset opening angle is 60 degrees and the correction angle is -8 degrees, the final angle would be 52 degrees. This project defines terms such as pitch frequency (e.g., Do = R1) and corrected angle (CorAngl) and designs specific fuzzy rule bases for each pitch range to improve valve control accuracy and musical performance. Figure 7 shows the fuzzy inference for range i, where $i = 1\sim9$. The variables and fuzzy rule base for pitch interval 1 are shown in Figures 8 and 9, while the variables and fuzzy rule bases for other intervals follow the same methodology.

If R_i=NB, Then CorAngl=NB

If R_i=NS, Then CorAngl=Z

If R_i=Z, Then CorAngl=NS

If R_i=PS, Then CorAngl=PB

If R_i=PB, Then CorAngl=PS

Fig 7. Fuzzy inference

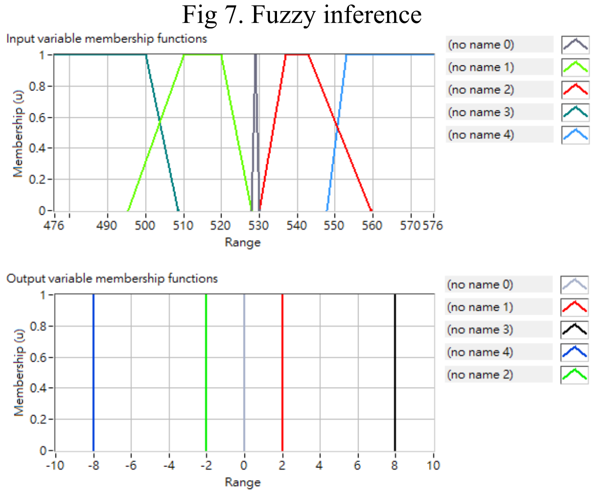


Fig 8. Variables for Range 1

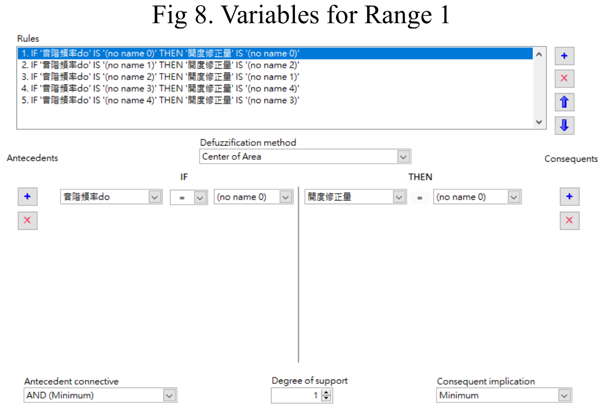
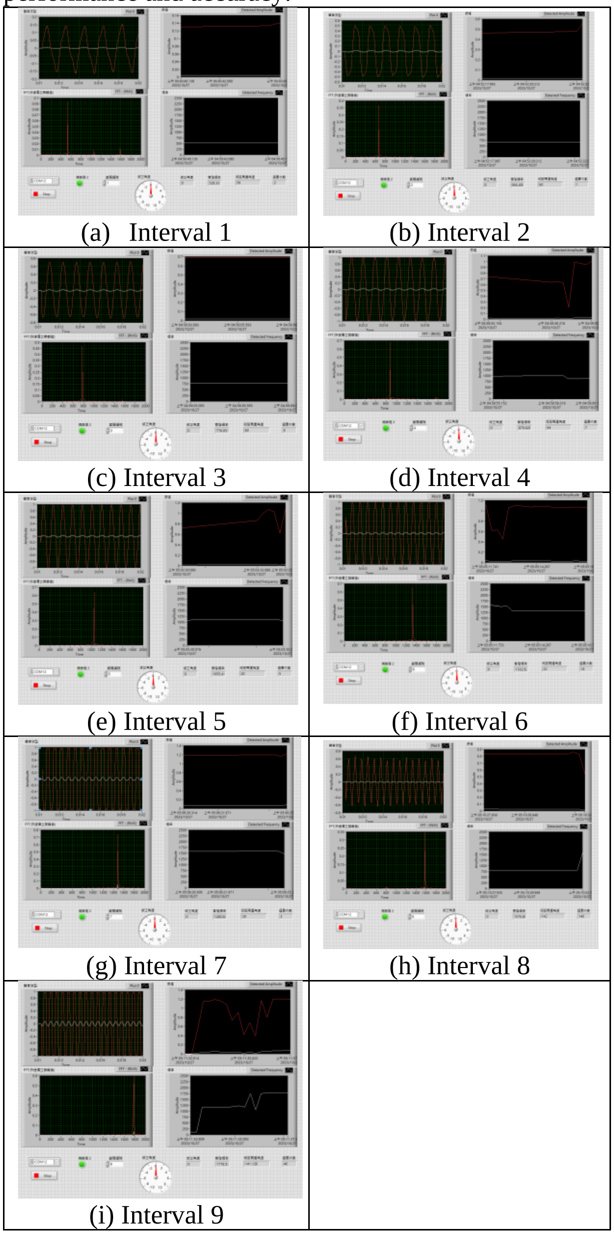


Fig 9. Fuzzy Rule Base for Range 1

3.2. *Automatic scale tuning method*

In the automatic tuning experiment, the tuning tests for nine pneumatic valves were successfully completed, with the results shown in Figure 10. Taking Section 1 as an example, the frequency range for the note was set to 528–530 Hz, and the measured frequency was 528.33 Hz. The servo motor's final angle was 58 degrees, with a correction angle of 0 degrees. Other sections similarly achieved the required frequency range with precision. The experimental results demonstrate that the automatic tuning function performs excellently across all note sections, significantly enhancing the system's musical performance and accuracy.



(a) Interval 1 (b) Interval 2
(c) Interval 3 (d) Interval 4
(e) Interval 5 (f) Interval 6
(g) Interval 7 (h) Interval 8
(i) Interval 9

Fig 10. Automatic tuning results of the 9 pneumatic valves

4. Experimental results

Since the topic of this research is "Fuzzy-Controlled Multi-Valve Pneumatic Soprano Recorder Auto-Playing and Score Recognition System," the experimental results will be presented in the form of a video, demonstrating an improved playing performance compared to previous systems.

5. Conclusion

This study improves the automated score recognition and soprano recorder performance system, enhancing playback speed, accuracy, and sound quality. Optimized recognition and classification methods improve pitch assignment. The solenoid-based multi-valve system enhances performance, while the 9-valve setup and fuzzy control significantly improve pitch stability and tuning speed. Experimental results confirm that these refinements greatly enhance tonal precision and overall system performance.

Acknowledgment

This work was supported by grants from the National Science and Technology Council, Taiwan, R.O.C. (MOST 111-2637-E-270-001).

References

1. Lee W.C., Gu H.Y., Chung K.L. et., "The realization of a music reading and singing two-wheeled robot", 2007 IEEE Workshop on Advanced Robotics and Its Social Impacts, pp.1-6, 2007.
2. Hong M.J., "Robotics Artistic Colorful Picture Drawing and Painting Using Visual Feedback Control System", Master Thesis of National Taiwan University, 2016.
3. Chang Y.H., Peng Z.X., Jeng L.D., "Automatic Music Score Recognition System Using Digital Image Processing", International Journal of Computer and Information Engineering, Vol.9, No.7, pp.1804-1810, 2015.
4. Zhe Xiao, Xin Chen, Li Zhou, "Real-Time Optical Music Recognition System for Dulcimer Musical Robot", Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 23, No. 4, pp. 782-790, 2019.
5. Wang B.R., Chen C.Y., "Development of an image processing based sheet music recognition system for iOS devices", 2014 IEEE International Conference on Consumer Electronics, Taiwan, May 26-28, pp.223-224, 2014.
6. C.C. Wang, "Fuzzy Theory-Based Air Valve Control for Auto-Score-Recognition Soprano Recorder Machines," Journal of Robotics, Networking and Artificial Life (JRNAL), Vol.8, Issue 4, pp. 278-283, March, 2022.
7. Željko Šitum, Danko Ćorić, "Position Control of a Pneumatic Drive Using a Fuzzy Controller with an Analytic Activation Function", Sensors, 22, 1004, 2022.
8. Grzegorz Filo, "A Review of Fuzzy Logic Method Development in Hydraulic and Pneumatic Systems", Energies, 16, 7584, 2023.
9. Jony Javorski Eckert, Ícaro Pavani Teodoro, Matheus Valente Lopes, Qing Wu, Auteliano A. Santos, "Multi-objective optimization of electro-pneumatic braking process with fuzzy logic control for heavy haul railway applications", International journal of rail transportation, DOI: 10.1080/23248378.2024.2338837, 2024.

Authors Introduction

Prof. Chun-Chieh Wang

He is a contract-based professor of Department of Electrical Engineering of National Yunlin University of Science and Technology. His areas of research interest include robotics, image detection, electromechanical integration, innovative inventions, long-term care aids, and application of control theory. He is now a permanent member of Chinese Automatic Control Society (CACS) and Taiwan Society of Robotics (TSR). He is also a member of Robot Artificial Life Society.

Prof. Chung-Wen Hung

He received the PhD degree in Electrical Engineering from National Taiwan University in 2006. Currently, he is a professor at National Yunlin University of Science and Technology. His research interests include IoT, IIoT, power electronics, motor control, and AI application.

Prof. Kuo-Hsien Hsia

He received the Ph.D. degree in electrical engineering from the National Sun Yat-Sen University, Taiwan, in 1994. He is currently an Associate Professor of National Yunlin University of Science and Technology in Taiwan. His research interests are in the area of mobile robotics, fuzzy control and image processing.

Prof. Chian-Cheng Ho

He was born in Taipei in 1973 and received his Ph.D. in electrical engineering from National Chung Cheng University in 2000. He was a visiting Ph.D. student at UCLA in 2000 and worked at ChungHwa Telecom Labs on VoIP equipment development until 2005. Since then, he has been an associate professor at National Yunlin University of Science and Technology, focusing on embedded processor firmware, real-time operating systems, and visual robotics applications.

Mr. Ying-Yuan Yao

He is currently pursuing a master's degree in the Graduate School of Intelligent Data Science, Department of Electrical Engineering, at National Yunlin University of Science and Technology. His research interests encompass robotics and electromechanical integration.

Development of agricultural robots based on ROS

Jr-Hung Guo*
Department of Electrical Engineering, National Yunlin University of Science & Technology.
123 University Road, Section 3,Douliou, Yunlin 64002 , Taiwan, R.O.C

Kuo-Hsien Hsia
College of Future National Yunlin University of Science and Technology
123 University Road, Section 3,Douliou, Yunlin 64002 , Taiwan, R.O.C
Email: jrhung@yuntech.edu.tw, khhsia@yuntech.edu.tw

Abstract

The agricultural industry has become a major issue in many countries due to drastic weather changes and a decrease in the number of people willing to engage in agriculture. Therefore, this study attempts to develop a platform for agricultural robots so that robots can assist people in agricultural work. This study uses Robot Operating System (ROS) as the software foundation. Through this convenient software foundation, different agricultural robots can be developed quickly. This study uses this architecture to develop a lawn mower and leaf sweeper, which can be easily converted to different agricultural applications in the future. In this study, the control system architecture we used is different from the commonly used single-board computer or microcontroller architecture. Instead, we used an ultra-small x86 mini-PC platform, mainly because the x86 platform used to be larger in size, more power consumption and other issues. However, the platform used in this study can be as small as 88mm*88mm*50mm, which is very close to the common Raspberry Pi or Jetson Nano, but the computing power, expansion capability and subsequent related development issues can be well developed. Finally, we are also trying to develop a web-based development and human-machine interface (HMI), hoping to make the subsequent development and use of different robots easier.

Keywords: Agricultural robots, Robot Operating System (ROS), x86 mini-PC, Web base HMI

1. Introduction

Because the agricultural industry is a relatively hard job, fewer and fewer people are willing to work in it. Among them, GILLER, Ken E., et al. [1] mentioned the problems and future development of global agricultural production. KASSIM, Mohamed Rawidean Mohd [2] mentioned the trends and advantages of IoT technology in future agricultural development. CHUNJIANG, Z. H. A. O., et al. [3] proposed an agricultural robot architecture, while OLIVEIRA, Luiz FP; MOREIRA, António P.; SILVA, Manuel F. [4] proposed the technological progress and future of agricultural robots. From the above related research, it is found that there are actually several key issues in the development of agricultural robots. (1) The structure of agricultural robots is complex. Since the structures of various agricultural uses vary greatly, it is impossible to use one structural platform to meet all agricultural needs. (2) Technical difficulties: agricultural environments vary greatly, so the entire control and related technologies are very complex. (3) High cost and difficult operation: agricultural robots are expensive because there is no universal platform. In addition, it cannot provide a simple development or operation interface, so farmers have many problems in using it. Due to the above problems, robots cannot be widely used in agriculture.

In addition, in terms of the robot's structure and control system, this study uses an existing remote-controlled agricultural robot platform. The main reason for doing so is that such a platform can already be used in agriculture, but it lacks automation and intelligence functions. . In addition, we use the x86 mini-PC architecture in the control system. Most of the current mobile robots use single-board computers such as Raspberry Pi or Jetson Nano as the main controller. However, the computing and expansion capabilities of these single-board computers are insufficient, and the size and cost advantages that these single-board computers previously had have also disappeared. The authors of this study have used mini-PCs to develop many platforms and related robot teaching systems in the past, so we hope to apply such experience and architecture to agriculture, hoping to provide some help for agricultural automation.

2. System architecture

2.1. Hardware and Mechanism

In the past, we developed mobile robot platforms for teaching purposes. Despite this, we still have some higher-level development and use such as image processing, but in the process, we found that the commonly used control boards such as Raspberry Pi or Jetson Nano are relatively insufficient in computing power or scalability, so we Use x86 mini-PC as the control core of the robot. Fig. 1 is a block diagram of the robot system. Through this architecture, we make the development and modification of robots easier. This has always been our goal. We hope to use a similar platform as the control core of robots with different architectures.

Fig. 2(a) is the mobile robot teaching platform we developed. This robot is only 130mm*130mm*150mm, and its height is only 90mm without LiDAR (Light Detection and Ranging). After we develop this platform, we hope that it can be used not only in teaching but also in actual industry.

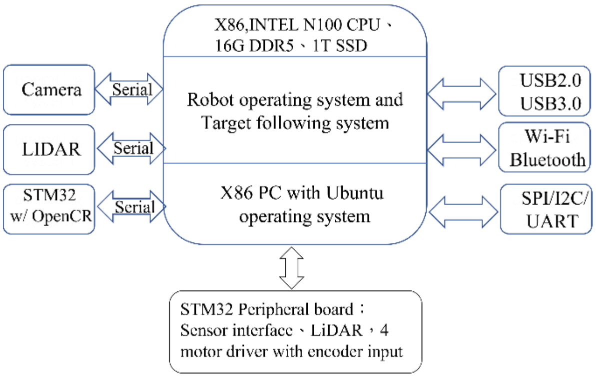
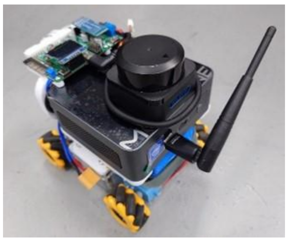


Fig.1 Block diagram of the robot control system.



a. Teaching robot.

b. Remote controlled lawn mower robot.

Fig. 2. Robotic platform.

There just happened to be a manufacturer of agricultural lawn mowers that wanted to transform their products from traditional remote-controlled to fully automatic lawn mowers. Fig. 2(b) shows the original remote control lawn mower. This original lawn mower uses a DC motor, but there is no encoder or related sensors. The lawn mower needs to be driven manually, so a considerable amount of work has been done on the mechanism and electronic control modification, Fig. 3 shows some of the modified areas. After the lawn mower modification was completed, the overall performance was very good after the commissioned manufacturer tested it. Therefore, the manufacturer commissioned us to modify the leaf sweeper. The general structure of this leaf sweeper is similar to that of the lawn mower, but the driving part uses 4 motors and wheels. Due to the systematization of the architecture and ROS parameters we use, the modification of the leaf sweeper was completed with only simple modifications. Fig. 4 shows the lawn mower and leaf sweeper after the modification. After completing the hardware modification of the two agricultural robots, we also developed the relevant human-machine interface and enabled the grass-cutting robot and the leaf-sweeping robot to cooperate with each other to complete the task. Next, we will explain the human-machine interface and software experience algorithm.

Fig.3 Block diagram of the robot control system.

a. Teaching robot.

b. Remote controlled lawn mower robot.

Fig. 4. Agricultural robots based on ROS.

2.2. *Group Task Algorithm*

Since ROS is a development system based on a single robot, it is relatively lacking in multi-robot interaction, data exchange, and task scheduling. We plan to adopt a sequential single-item auction [5]. We define each robot as R_i , and each job or path is defined as $W = \{W_1, W_2, \ldots\ldots, W_n\}$, and the time each W takes the robot to take is defined as Tn, then we can define the cost of each robot as $P(R_i, W_{i,n}), i = 1 \cdots k$. In addition, we define an $Q = \{Q_1, Q_1, \ldots\ldots, Q_n\}$ indicator to evaluate the efficiency of each W, on different robots, that is, different robots such as wheel type, carrying capacity, movement speed, etc. The higher the value of Q, the more suitable the robot is for this W. And the value of Q ranges from (1% to 100%), and the detailed definition of Q can be freely set according to the characteristics of the robot. Then we can rewrite the above robot cost definition as $P(R_i, W_n(R_i)Q_n)$. In addition, we also define a as the additional cost when the robot executes W under different conditions. Then we can define the time cost when the robot executes W without changing tasks midway and with changing tasks midway as Eq. (1);

$$\begin{cases} if\, change T\; P(R_i, W_n(R_i)Q_n) = P(R_i, W_n(R_i)Q_n) + E \\ if\, not\; change T\; P(R_i, w_n(R_i)Q_n) = P(R_i, w_n(R_i)Q_n) \end{cases} \quad (1)$$

Therefore, the evaluation function of the robot when performing work is:

$$MINSUM : \min_{W} \sum_{i=1}^{n} P(R_i, W_n(R_i)Q_n) \quad (2)$$

$$MINMAX : \min_{W} \max_{i} P(R_i, W_n(R_i)Q_n) \quad (3)$$

This function is also used to evaluate whether the robot should change or increase W during the execution of W. We define a collision space function $G = (C, N)$, where are all the task points, so $N = \{N_1, N_2, ..., N_n\}$, and we define an array S , which is the distance between two task points $S\{C_1, C_2, ..., C_n\}$. Then the optimal W allocation can be written as:

$$\min \sum_{i=1}^{n-1} C_i x_i \quad (4)$$

But subject to;

$$\sum_{i=1}^{n} x_i = 1 \quad (5)$$

$$x_i \in \{0,1\} \qquad (i = 1,2,......,n) \quad (6)$$

Because x_i is defined as whether each new W is the shortest path. Finally, we can rewrite the new MINSUM function as;

$$MINSUM : \min_{T} \sum_{i=1}^{n} P'(R_i, W_n(R_i)Q_n) \quad (7)$$

In this way, the various characteristics and paths of the robot can be taken into account, and the path or task can be updated quickly. Therefore, when necessary, a data center can be built to store the location of each robot in the environment, the environment map, the load The robot can update data at any time to monitor the status of the robot, and the data center can also transmit the calculation results to the robot. Through such a mechanism, each robot can complete all tasks quickly, and the task scheduling of agricultural robots can complete the tasks assigned by farmers to the robots in the simplest way.

2.3. *Human Machine Interface*

The agricultural robot in this study uses a web-based human-machine interface. The reasons for using this method are: (1) The operating interface is popular, and both WINDOWS, LINUX, and Mac OS have the ability to use web-based interfaces. Based on the application, and simple to use, almost everyone can do it. (2) 2D, 3D and multimedia technologies. In the past, due to the limitations of related web technologies, it was very difficult to display 3D images or multimedia and other complex visual effects on web interfaces. With the advancement of technology, the above All problems have been solved, so users can see the real-time working environment status, or simulate and operate in 3D simulation. (3) System integration is simple and has good security. Web-based is a standard network protocol application based on TCP/IP, so it is quite easy to connect with other systems. In terms of security, such as SSL (Secure Sockets Layer) , AES (Advanced Encryption Standard) can provide a very high level of data communication security protection, and is also widely used by government agencies, banks and other websites. Therefore, web-based is a pretty good platform.

This study uses Rosbridge API to develop the human-computer interface and converts graphical interface software such as rqt_graph and RViz in ROS to be displayed and operated on the Web. In the multi-robot path planning and intelligent task scheduling part, the robot can transmit the environment and robot status to the server at any time while moving, and then integrate the multi-robot environmental data into the database through the server. , and then perform dynamic path analysis and planning. In addition to expanding the robot's map data, it can also perform dynamic and more complete path planning and prediction, as well as group task scheduling for multiple robots.

For the web-based ROS human-machine interface, although ROS provides the Rosbridge API, which also has a websocket tool, how can the outgoing messages be integrated with the web interface, and how can the operations on the web interface be transmitted to ROS via Rosbridge? , how to display and integrate graphical tools in ROS, these are not done by Rosbridge alone, so this study uses C++, HTML, JavaScript and CSS to develop a web-based human-computer interface, among which C++ is used to Rosbridge compiles the function packages needed by the robot in the ROS workspace, and is also used to process the messages sent back by Rosbridge from ROS into data structures that can be used by web pages. HTML is the basis for writing the entire web page, including basic web page elements such as text, pictures, buttons, etc. JavaScript is used to process ROS data and allow web pages to have richer element control. CSS optimizes the color, layout, font size, etc. of a web page to make it more beautiful and user-friendly.

Fig. 5 shows the screen of a handheld device. Since the screen of the handheld device is relatively small, only a portion of the screen is displayed. However, other functions can still be switched through the window. Fig. 6 shows the screen as seen on a computer. Because the computer screen is larger, the robot's operation and related windows can be seen simultaneously. These functions are completed through a web browser.

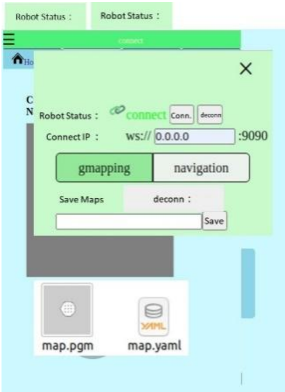


Fig.5 Block diagram of the robot control system.

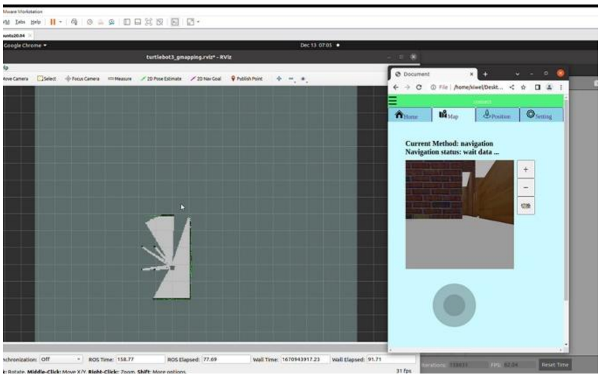


Fig.6 Block diagram of the robot control system.

3. Conclusion

We use ROS to convert lawn mowers and leaf sweepers that are already on the market into mobile robots that can operate autonomously. And also developed the Web base operation interface. After actual testing, the robots were able to operate normally and cooperate with each other to complete tasks. After this development experience, we also found that although ROS is very convenient, the related parameters and modifications are still too complicated. Therefore, we also developed a preliminary prototype so that different robot architectures and functions can be used in the future. This is done using the Web base method. In the future, we will bring this up again together with related tasks and map management for discussion.

Through this development, we have developed an agricultural robot that can really work on the farm. This result is also a beginning for us. We hope that in the future we can develop more agricultural robots to solve agricultural problems. The problem of insufficient manpower. We also try to simplify the development and use of robots. Finally, I hope that robots can truly serve people.

References

1. GILLER, Ken E., et al. The future of farming: Who will produce our food?. Food Security, 2021, 13.5: 1073-1099.
2. KASSIM, Mohamed Rawidean Mohd. Iot applications in smart agriculture: Issues and challenges. In: 2020 IEEE conference on open systems (ICOS). IEEE, 2020. p. 19-24.
3. CHUNJIANG, Z. H. A. O., et al. Agricultural robots: Technology progress, challenges and trends. *Smart Agriculture*, 2023, 5.4: 1.
4. OLIVEIRA, Luiz FP; MOREIRA, António P.; SILVA, Manuel F. Advances in agriculture robotics: A state-of-the-art review and challenges ahead. *Robotics*, 2021, 10.2: 52.
5. GUO, Jr-Hung. Applications of the Internet of Things technology in advanced planning systems. Sensors and Materials, 2018, 30.8: 1723-1734.

Authors Introduction

Dr. Jr. Hung Guo

He is an Assistant Professor at the Department of Electrical Engineering, National Yunlin University of Science & Technology, Douliou, Taiwan. He received his Ph.D. Degree from National Yunlin University of Science & Technology, Taiwan in 2012. His research interests include sensor network, intelligent systems, intelligent robot. and technical and vocational education

Dr. Kuo-Hsien Hsia

He received the Ph.D. degree in electrical engineering from the National Sun Yat-Sen University, Taiwan, in 1994. He is currently an Associate Pofessor of National Yunlin University of Science and Technology in Taiwan. His research interests are in the area of mobile robotics, fuzzy control and image processing.

The ROS-based web information center of small manipulator

Yue-Jie Wang
Graduate School of Intelligent Data Science, National Yunlin University of Science & Technology, 123 Section 3, University Road, Douliou, Yunlin 64002, Taiwan, R.O.C.

Jia-Ming Hsiao*
Bachelor Program in Intelligent Robotics, National Yunlin University of Science & Technology, 123 Section 3, University Road, Douliou, Yunlin 64002, Taiwan, R.O.C.

Shao-Yi Hsiao
Department of Intelligent Robotics, National Pingtung University ,
51 Min Sheng E. Road, Pingtung 900, Taiwan, R.O.C.
Email: M11156018@yuntech.edu.tw, hsiaojm@yuntech.edu.tw, shiao9212@gmail.com*

Abstract

A ROS-based web information center with respect to small manipulator is developed to show the feasibility of the one for factories. Small ROS-based 6-axis manipulators for teaching and research are utilized to simulate the applications of robot arms in the factories. The task of robot arm is designed to recognize the object color and then to grip the object to the desired storage area. Production operation data and status information from the robot arm are transmitted through the ROS rosbridge_suite to the web interface for display and saved in a database. A start-stop function is also implemented to remotely start and stop the operation of robot arm.

Keywords: Robot Operating System (ROS), rosbridge, manipulator

1. Introduction

With the development of Industry 4.0, smart manufacturing has become one of important issues of modern intelligent industry. Through smart manufacturing, companies from manufacturing industry can use advanced technologies such as data analysis, automation, network technology, and artificial intelligence to realize the intelligence and digitization of the production process. The three key technologies for building a smart manufacturing factory include data collection and analysis, automation and robotics, and digital manufacturing and virtual simulation [1]. There are variety of sensors and monitor facilities in the smart factory to collect manufacturing data, and this data needs to be processed and analyzed effectively. With the respect of monitoring production data, the graphical user interface (GUI) is an intuitive interface to show the real time process information in order to track and manage easily [2]. The GUI for monitoring production data can be easily implemented by dynamic web techniques.

Automation and robotics are the key technologies to reduce human errors and improve production efficiency and quality. However, the communications among different robot systems are difficult before the announcement of robot operating system (ROS). As the middleware of robot, the purpose of ROS is to solve issues of compatibility and interoperability. ROS provides system developers with modular development framework and more flexibility, and allows the combination of small modules to perform more complex tasks. Therefore, ROS simplifies the programming process, which also makes the ROS code more versatile [3]. As a package of ROS, rosbridge provides a JSON API to ROS functionality for non-ROS programs [4]. The communications between front webpages and robots can be made by rosbridge to implement GUI information center to show and monitor the process data, and even to remotely control the robot.

In the present study, we introduced a ROS-based web information center to monitor the process information of small ROS-based 6-axis manipulators. The ROS-based 6-axis manipulators for teaching and research are utilized to simulate the applications of robot arms in the factories. The task of robot arm is designed to recognize the object color and then to grip the object to the desired storage area. Production operation data and status information from the robot arm are transmitted through the ROS rosbridge package to the web interface for display and saved in a database. A start-stop function is also implemented to remotely start and stop the operation of robot arm.

2. Methodology

The proposed system consists of object identification, ROS-based manipulators, MySQL relational database management system and web station. The object identification function and ROS-based manipulators are combined to simulate a part of production process of packaging. The process data is stored in MySQL, and then can be shown in the front web page, as well as the working status of manipulators. The communications between web and manipulators, web and camera are implemented by ROS rosbridge package.

2.1. *Object color identification*

The production object is identified by the camera image color. A ROS node is designed to subscribe the topic published by the ROS node usb_cam from the package usb_cam in order to reduce the computation load of the manipulator. The color identification function is based on HSV (Hue, Saturation, Value) color model. First the BGR (Blue, Green, Red) color model of the camera image is transferred to HSV model. Then the calculations of HSV values from BGR values are performed. The calculation equations of HSV values from BGR values are omitted and can be referred to [5]. The (x, y) pixel threshold value of image color identification is obtained by Eq. (1), where L is the lower bound and U is the upper bound. Then AND bitwise operation is used to apply the mask generated by threshold filtering to the original image, retaining only the parts that match the color range. After the object color identification is done, the manipulator is activated to grip the target object to the designated area.

$$M(x,y) = \begin{cases} 1, if\ L \leq H(x,y), S(x,y), V(x,y) \leq U \\ 0, otherwise \end{cases} \quad (1)$$

2.2. *ROS-based 6-axis manipulator*

In this study mecharm270 pi which is a small ROS-based 6axis manipulator for teaching and research as shown in Fig. 1 is applied to simulate the production process in the factories. A python node callback.py is launched to subscribed the topic published from the web interface to perform the start and stop of the manipulator. The communication between web interface and callback.py node is established by rosbridge package as shown in Fig. 2. Two publisher nodes are implemented in the manipulator to publish the production data as shown in Fig. 3.

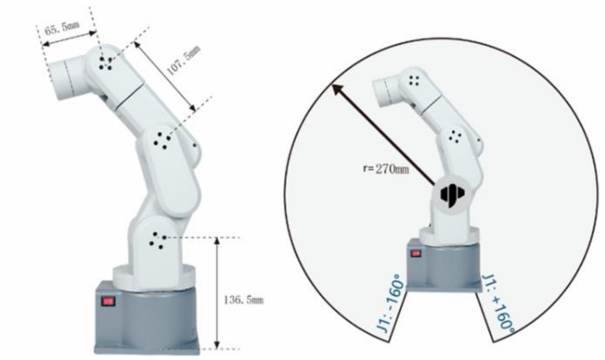


Fig. 1. mecharm270 pi

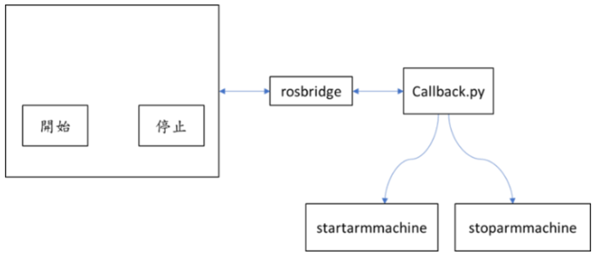


Fig. 2. Communication structure of web interface start-stop function and node callback.py

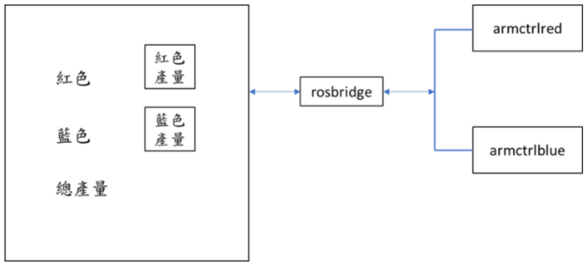


Fig. 3. Production data and publisher nodes

2.3. *Web interface and database*

The web interface is constructed by node.js and a MySQL database server runs for data storage. The layout for one status and data is arranged as Fig. 4. The left column is designed for tool icons and the right hand side is kept to show the manipulator information. The start and stop buttons function to remotely start and stop the manipulator movement. The process data is shown in the object amount numbers. The real time camera image is also monitored in the interface. It is shown in Fig. 5 that the communication between the web user interface which is a non-ROS client and the manipulator as ROS server is made by ROS rosbridge package.

The production process data is accessed in the backend of the web interface. The object amount data is stored in MySQL in order to track the process information in the future.

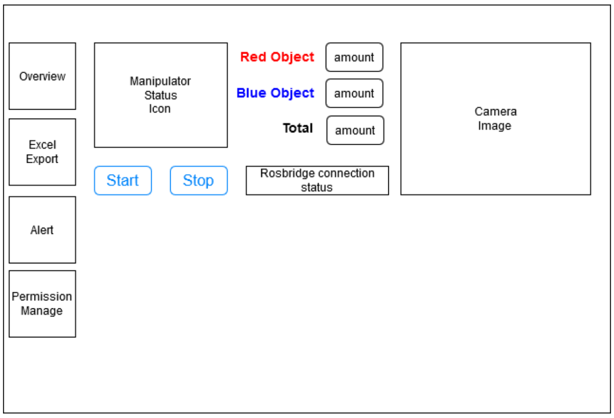


Fig. 4. Web interface layout.

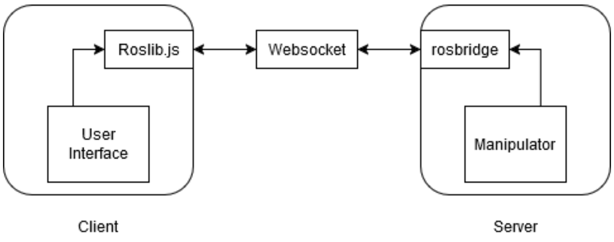


Fig. 5. Rosbridge communication.

3. Results and Discussion

The result web interface is shown as Fig. 6. There are 2 manipulators process data shown in the interface. The object image is obtained by the ROS usb_cam node. The topic /image_raw is published by usb_cam node and the node image_view and web interface subscribe it. The image_view output the image to the display window as shown in Fig. 7. The object color identification is performed by utilizing OpenCV python library with HSV model. The thresholding upper and lower bound of color red is selected as [10, 255, 255] and [0, 100, 100], respectively. The thresholding upper and lower bound of color blue is selected as [130, 255, 255] and [110, 50, 50], respectively. It is shown in Fig. 8 that identification results of red and blue objects. The initial image is transformed from ROS image format into OpenCV image format by our combined_color_recognition node and CvBridge package. After the target object is detected, the manipulator is activated to grip the target object.

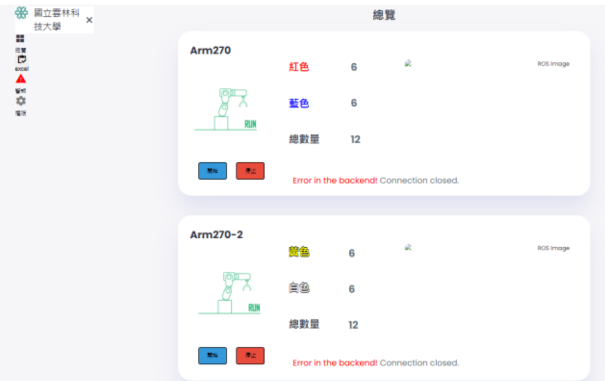


Fig. 6. Result web interface.

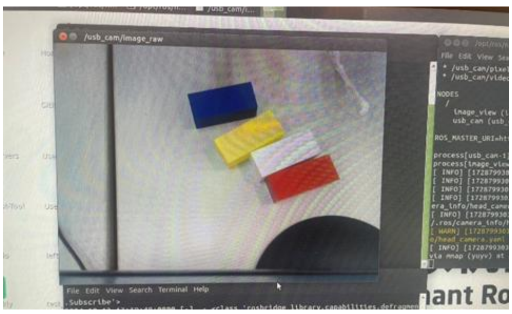


Fig. 7. The object image.

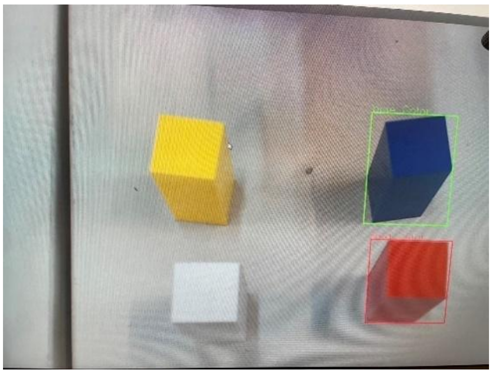


Fig. 8. identification results of red and blue objects.

After the gripping activation is done, the manipulator publishes a message of target color object amount. The web interface subscribes this message and calls self-defined function updateDatabase() to insert the process object data into database table with JSON format. Fig. 9 is a sample result for the database records.

id	red	blue	redtotal	bluetotal	total	timestamp_column
1	0	1	0	1	1	2024-09-30 16:23:25
2	1	0	1	1	2	2024-09-30 16:23:52
3	1	0	2	1	3	2024-09-30 16:24:26

Fig. 9. Process data records.

The process data is queried from database and shown in web interface as Fig. 6. This data can be exported to excel file by clicking the "excel" tool icon in the left column of web interface. The exportation result is shown in Fig. 10.

ID	red	blue	redtotal	bluetotal	total	Timestamp
1	0	1	0	1	1	2024-09-30 16:23:25
2	1	0	1	1	2	2024-09-30 16:23:52
3	1	0	2	1	3	2024-09-30 16:24:26
4	1	0	3	1	4	2024-09-30 16:32:50
5	0	1	3	2	5	2024-09-30 22:18:14
6	1	0	4	2	6	2024-09-30 22:20:14
7	1	0	5	2	7	2024-09-30 22:21:57
8	1	0	6	2	8	2024-09-30 22:22:28
9	1	0	7	2	9	2024-09-30 22:23:06
10	0	1	7	3	10	2024-09-30 22:23:37
11	0	1	7	4	11	2024-09-30 22:25:02
12	1	0	8	4	12	2024-09-30 22:25:34
13	1	0	9	4	13	2024-09-30 22:32:46
14	1	0	10	4	14	2024-10-01 13:25:11
15	1	0	10	4	14	2024-10-01 13:25:11
16	0	1	10	5	15	2024-10-13 16:17:08
17	0	1	10	6	16	2024-11-03 15:24:30
18	0	1	10	7	17	2024-11-03 15:25:59
19	0	1	10	8	18	2024-11-03 15:27:55
20	0	1	10	9	19	2024-11-03 15:29:29
21	0	1	10	10	8	2024-11-04 15:46:54
22	0	1	10	11	9	2024-11-04 15:48:30
23	0	1	10	9	19	2024-11-04 15:52:43
24	0	1	10	10	8	2024-11-04 15:54:10
25	0	1	10	11	8	2024-11-04 16:00:34
26	0	1	10	9	19	2024-11-04 16:02:50
27	0	1	0	1	1	2024-11-05 16:03:24

Production Data | Alert Data | Production2 Data | Alert2 Data

Fig. 10. Process data exported to excel file.

The manipulator status is also displayed in the interface with 3 icon images. The 3 status icons include Run, Idle and Alarm as shown in Fig. 11. The blue "Start" button is used to remotely activate the manipulator to work, and the red "stop" button is used to remotely stop the manipulator.

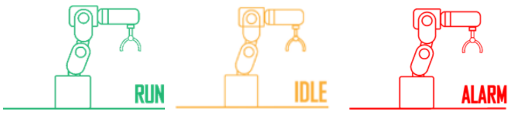


Fig. 11. Manipulator status icons.

4. Conclusion

Graphical user interfaces are an integral part of smart factories, but ROS is not friendly to novice developers. Developers need a deeper understanding of ROS, and there is a lack of specific suppliers to provide GUI options. This study proposes a customized web visual monitoring interface based on the ROS rosbridge package, which is applied to collect the equipment status, color classification and production data of ROS equipment operating in the factory. Rosbridge is used to connect ROS-based robot arms and non-ROS visual monitoring interfaces for bidirectional communication. In this way, non-ROS devices or systems can also be integrated into ROS systems. It will be convenient to develop the hardware and software integration under ROS framework.

Table. 1. Comparison of protocols

Protocol / Item	OPC	OPC UA	Rosbridge
Platform supported	MS Windows	MS Windows, Linux, iOS	MS Windows, Linux, iOS, ROS, Web
Security	High	High	Medium
Expansibility	Low	High	High
Development difficulty	Medium	High	Low
Typical application scenarios	Local Monitoring	Cross-system data integration, industrial Internet of Things	Robot control, Web monitoring, Data visualization

The comparison of open platform communications (OPC), OPC unified architecture (OPC UA) and Rosbridge is listed in Table. 1. Rosbridge has lower development difficulty and higher expansibility, but Rosbridge also has some problems that need to be optimized. For example, the security design of industrial applications is relatively simple, and for high standardization of industrial security, security is low. As the number of ROS users who develop robots grows year by year, the need to integrate ROS and non-ROS applications will be an issue worthy of attention.

References

1. Osterrieder, P., Budde, L., & Friedli, T., The smart factory as a key construct of Industry 4.0: A systematic literature review, International Journal of Production Economics, Vol. 221, 2020, 107476.
2. Kumar, N., & Lee, S. C., Human-machine interface in smart factory: A systematic literature review, Technological Forecasting and Social Change, Vol. 174, January 2022, 121284.
3. Hax, V. A., Duarte Filho, N. L., Da Costa Botelho, S. S., & Mendizabal, O. M., ROS as a middleware to Internet of Things, Journal of Applied Computing Research, Vol. 2, 2012, pp.91–97.
4. ROS wiki – rosbridge_suite, http://wiki.ros.org/rosbridge_suite
5. Loesdau, M., Chabrier, S., & Gabillon, A., Hue and saturation in the RGB color space, In Lecture Notes in Computer Science, 2014, pp.120–128.

Authors Introduction

Mr. Yue-Jia Wang

He received his Bachelor's degree in Management in 2022 from the Department of Industrial & Management, National Chin-Yi University of technology in Taiwan, R.O.C.. He is currently a master student in National Yunlin University of Science & Technology, Yunlin, Taiwan, R.O.C..

Prof. Jia-Ming Hsiao

He received the B.S. and Ph.D. degree in Electrical Engineering from National Sun Yat-Sen University, Kaohsiung, Taiwan in 1994 and 2002, respectively. He is an Assistant Professor in the Bachelor Program in Intelligent Robotics at National Yunlin University of Science & Technology, Yunlin, Taiwan, R.O.C.. He is also the 2022-2025 youth group vice presiding judge of mobile robotics at national skills competition in Taiwan. His current research interests include system and control theory, internet of things, mobile robotics and artificial intelligence.

Mr. Shao-Yi Hsiao

He is currently an undergraduate student in Department of Intelligent Robotics, National Pingtung University in Taiwan, R.O.C.. His current research interests include internet of things and mobile robotics.

Image-assisted Assembly and Disassembly Process Using TM Six-Axis Collaborative Robotic Arm

Kuo-Hsien Hsia
Bachelor's Program in Intelligent Robotics, National Yunlin University of Science and Technology
123, University Road, Sec. 3, Douliu, Yunlin, Taiwan

Yi-Yan Liao
AMPOC Far-East Co., Ltd.
17F., No.171, Sung-Teh Road, Taipei, Taiwan

Ching-Yuan Pan
MTmation Engineering of Automation Systems
28F-12, No. 99, Sec. 1, Xintai 5th Rd., Xizhi Dist., New Taipei City 221, Taiwan
Email: khhsia@yuntech.edu.tw, eane135759@gmail.com, yuan2295@gmail.com

Abstract

This paper explores the development of the TM collaborative robot arm in industrial applications. With its in-house developed TMflow software, TM robot streamlines the intricate human-machine interface of industrial robots and modularizes various tool functions, allowing operators to quickly familiarize themselves. The focus of this paper is on programming the TM collaborative robot arm using TMflow to achieve automatic image-assisted localization for assembly and disassembly. Collaborative arms improve the accuracy and efficiency of assembly and disassembly, reducing manual errors and wasted time. In terms of safety, compared with general robotic arms, collaborative arms are safer, allowing them to cooperate safely with human workers and reducing the incidence of workplace accidents.

Keywords: Collaborative robot, COBOT, TMflow, assembly and disassembly process

1. Introduction

Collaborative robot (Cobot) arms are an advanced form of robotic technology designed to work alongside human workers to complete a variety of tasks and jobs. These robotic arms have highly flexible movement capabilities and advanced sensing systems that can identify and adapt to different working environments. Traditional industrial robot arms are usually set up in fixed work areas, with operators controlling the robot arms, and strict safety protection measures are required to avoid collisions with people. However, as manufacturing continues to grow and advance, so does the need for more flexible, safer and more efficient ways of working.

There are many applications of robotic arms. Fu-Chun Liang proposed a new way of making Takoyaki with a robot arm on the premise that humans prepared the materials [1]. Yu-Ting Chen used a robotic arm to grind the surface of workpieces [2]. The grinding moving path of the robot arm was planned using the surface joining method and the hierarchical algorithm based on the information about the surface of the workpiece by a structured light scanner.

Robot arms are often combined with machine vision in applications. Bing-Ting Tsai used machine vision to identify the position of objects in [3], so that the robotic arm can grip the object from different directions. In [4], Cheng-Long Lee used machine vision to know the position and distance of the object to intelligently generate the moving path of the robotic arm, so that the robotic arm can move to the corresponding position and then use eye-in-hand photography to find defects on the object. In [5], Yi-Ru Wu used machine vision combined with a robotic arm to pick up items, and designed an adaptive clamp that can transform items of different shapes into parallel or open-angle clamps for elastic clamping, through deep learning training. The results estimate that the center and shape of the model can effectively complete clamping.

This paper is organized as follows: In section 1, there is a brief introduction of this paper and relative works in the past. In section 2, there is a brief introduction of the collaborative robot TM5-900 used in this paper. In section 3, there is a brief description of the experiment environment. In section 4, some discussions about the assembly and disassembly processes are shown. In section 5, there will be a short discussions and conclusions.

2. TM5-900 collaborative robot

Techman collaborative robot (TM Cobot) arm TM5-900 [6], as shown in Fig. 1, is used in this research. It is a 6-axis robotic arm that covers a reach of 946 mm and carries a load capacity of up to 4 kg. It is suitable for 3C industry, automobile industry, food industry and other fields. TM Cobot is the only robot arm equipped with an eye in hand, e.g. a camera at the end of the robot arm. Its hand-eye camera uses a global shutter. The positioning accuracy measured by TM laboratory with a working distance of 100mm is 0.1mm for 2D condition. The positioning accuracy of TM Landmark 3D is 0.24mm. In reality, accuracy will vary due to factors such as ambient light sources, object characteristics, and vision programming. In addition to the built-in hand-eye camera, the TM5-900 can also be equipped with an external camera to expand its functionality. TM5-900 uses TMflow as its programming environment.

TMflow is a fully graphical design that can significantly reduce the learning threshold. The programming process is based on a step-oriented approach, making the operation more intuitive and easily understood. There is a FREE button and a POINT button at the end of the arm, allowing users to easily carry out hand movement and positioning operations. This design allows users to quickly and easily program the TM5-900 robot arm and flexibly respond to different application requirements.

Fig.1. Techman TM5-900 Cobot [6].

3. Experiment Environment

The experiment uses a TM Cobot to assemble and disassemble a solid-state drive (SSD). The components of the SSD are shown in Fig. 2. They are assembled from three parts, namely the upper cover, the printed circuit board (PCB) and the lower cover. The Cobot arm uses the vacuum suction cup in Fig. 3(a) and the electric screwdriver in Fig. 3(b) for assembly and disassembly. We use the suction cup to assemble and disassemble the three parts of the SSD, and use the electric screwdriver to tighten or remove the M2 flat head screws to loosen the SSD components. Two trays are designed for the task, as shown in Fig. 4. The left one is for accommodating 24 M2 flat head screws, and the right one, the assembly tray, provides a fixed assembly position. TM Landmark is used for positioning for both trays. The removed screws will be placed in the business card box as shown in Fig. 5. The business card box is designed with grooves to allow the screws to escape the magnetic force of the electric screwdriver. After the business card box is full of screws, pour the screws in the box into the screw organizer so that the removed screws can be used again.

In addition, in order to enable the Cobot to install two end effectors at the same time, we designed an adapter block made by 3D printing so that the two end effectors

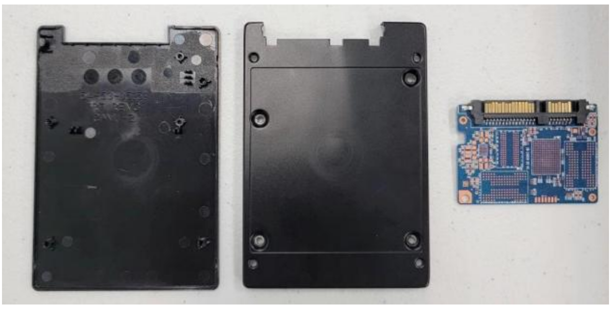


Fig. 2. Components of the SSD for this paper.

(a) vacuum suction cup (b) AE-SDL-1800 electrical screwdriver

Fig. 3. End-effector used in this paper.

Fig. 4. Trays for the task.

can be fixed on the end flange surface of the Cobot at the same time. As shown in Fig. 6(a), the vacuum suction cup and the electric screwdriver can be fixed on the black jig at the same time, and then installed on the end flange surface of the Cobot, as shown in Fig. 6(b). The layout of the overall environment is shown in Fig. 7. The PCB is transported from the previous process to this station of the assembly process via a conveyor belt. In Fig. 7, the square with red TM inside is the TM Landmark for positioning.

4. SSD Assembly and Disassembly

4.1. *Assembly and Disassembly Processes*

The assembly process can be divided into two parts: SSD assembly and screw locking. The Cobot pick up the lower cover by the suction cup and then places it on the assembly tray. Then pick up the PCB by the suction cup and put the board into the slot of the lower cover in an oblique manner. Finally, pick up the upper cover by the suction cup and move it at an oblique angle so that the tenon of the upper cover engages the slot of the lower cover. Then enter the process of screw tightening. First, move the electric screwdriver to a position above any

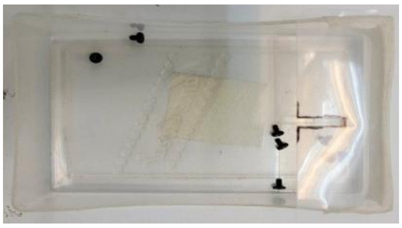
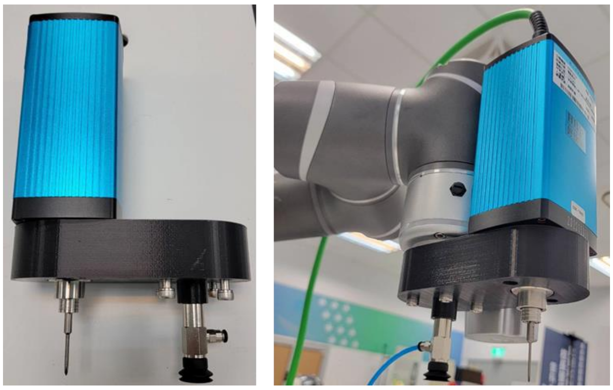


Fig. 5. Screw-recycling box.



(a) Adapter block with 2 end-effectors. (b) TM Cobot with dual end-effectors.

Fig. 6. Adapter block for dual end-effectors.

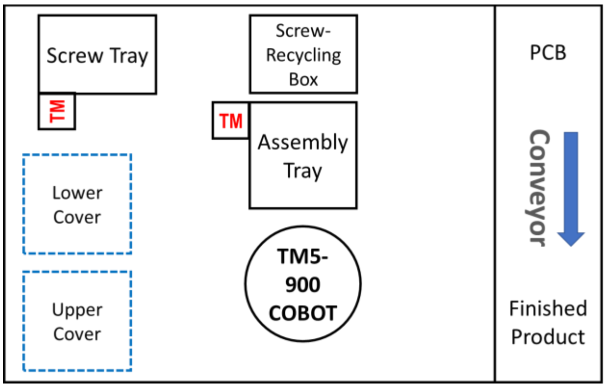


Fig. 7. Layout of the whole system.

screw in the Screw Tray, close to the screw with a low rotation speed to pick up the screw. Then move to the top of one screw hole. Similarly, first in a low rotation speed to make the screw engage the screw hole, and then rotate at a high speed to lock the screw. Repeat until all 4 screws are locked.

The disassembly process is also divided into two parts: screw removal and SSD disassembly. First, use the electric screwdriver on the Cobot to remove the four screws. When removing the screws, the screwdriver should first gradually approach the screws at a low speed to engage the screws, then loosen the screws at a high speed, and use the grooves on the screw recovery box to disengage the screws. Then use the suction cup to pick up the upper cover, PCB, and lower cover in sequence and place them in position.

4.2. *Troubles in Assembly Process*

There might be some troubles in assembly process. Three cases are considered in this paper. The first case is the problem of insufficient ambient light. This problem can be solved using the ring light around the hand-eye camera on the TM Cobot. The second case is the problem of poor image processing. At this time, one can adjust the shooting light source and shooting position to achieve the best image capture and simplify the image processing process. The third case is object stacking. This may lead to object detection errors and confusion in the order in which objects are retrieved. The object detection problem can be solved with the previously mentioned solutions to the light source and image processing problems, while the picking sequence problem can be solved with the parameters of the object detection function. In the object detection function, when multiple objects are detected, they can be sorted according to the detection score, left/right, top/bottom, or center of the image. As shown in Fig. 8, the lower covers or upper covers are randomly stacked, and through the sorting method of object detection, the Cobot can pick up the top object.

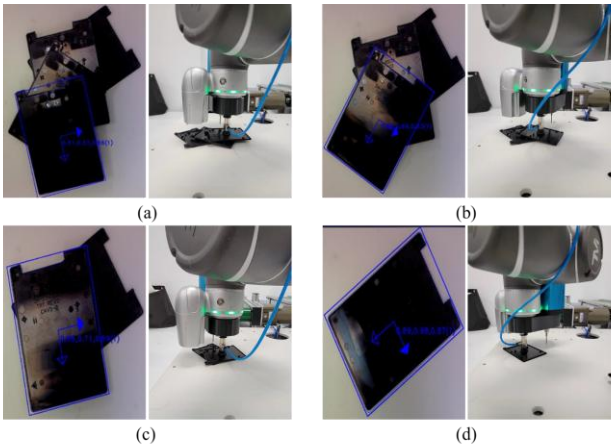


Fig. 8. Sequence of object-picking.

5. Conclusion

In this paper, we use TMflow to develop programs to control TM Cobot with vacuum suction cup and electric screwdriver to complete automated image positioning, SSD assembly and disassembly. By simulating the conditions of the real production line, the experimental environment is arranged to solve similar problems in the real case. These include acquisition and image processing issues, assembly and disassembly issues, and screw-related issues. This scenario will occur in many factories. It is very worthy of reference for manufacturers who want to automate production lines but have not yet done so.

References

1. F.C. Liang, Human-Machine Collaborative Cooking Robot - Using Takoyaki as an Example, Master Thesis, Department of Electronic Engineering, National Formosa Iniversity, 2020.
2. Y.T. Chen, Autonomous 6-Axis Robot Grinding System for Golf Club Head, Master Thesis, Department of Mechanical Engineering, National Taiwan University of Science and Technology, 2022.
3. B.T. Tsai, Six-axis Robot Manipulator Control by Integrating Image Processing Algorithm, Master Thesis, Department of Mechanical Engineering, Min Chi University of Technology, 2021.
4. C.L. Lee, Intelligent Six-Axis Robotic Arm in Industry Using Computer Vision, Master Thesis, Department of Electronic Engineering, National Yunlin University of Science and Technology, 2023.
5. Y.J. Wu, Intelligent Robotic Grasping Using Deep Learning and Stereo Images, Department of Power Mechanical Engineering, National Tsing Hua University, 2019.
6. Specification of TM5-900, Techman Robot Inc.

Authors Introduction

Prof. Kuo-Hsien Hsia

He received the Ph.D. degree in electrical engineering from the National Sun Yat-Sen University, Taiwan, in 1994. He is currently an Associate Professor of National Yunlin University of Science and Technology in Taiwan. His research interests are in the area of mobile robotics, fuzzy control and image processing.

Mr. Yi-Yan Liao

He received his master's degree in Electrical Engineering from National Yunlin University of Science and Technology in 2024. His research interests are collaborative robotic arms and image processing. He currently works as an engineer in automation.

Mr. Ching-Yuan Pan

He received his master's degree in Electrical Engineering from National Yunlin University of Science and Technology in 2024. His research interests are collaborative robotic arms and image processing. He currently works as an automatic control engineer in automotive automation production.

Quality Inspection of PVC Shoe Chopsticks: A Research Study

Shu-Li Pai
Institute of Engineering Science and Technology, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliu, Yunlin 64002, Taiwan

Kuo-Hsien Hsia
Bachelor's Program in Intelligent Robotics, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliu, Yunlin 64002, Taiwan

Chian-Cheng Ho
Department of Electrical Engineering, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliu, Yunlin 64002, Taiwan
Email: D11310205@yuntech.edu.tw, khssia@yuntech.edu.tw, futureho@yuntech.edu.tw

Abstract

In the past decade, the advancement of artificial intelligence technology has led to substantial growth in industrial automation. However, most manufacturing industries are still at the stage of physical labor automation, especially in the area of product inspection, where manual inspection is predominantly used. Taking a PVC shoe chopstick factory as an example, the production line operates 24 hours a day, but quality inspection requires employees to conduct checks on the next working day. This approach makes it difficult to promptly address any defects that occur during production. By incorporating computer vision technology into the PVC production line to track and measure products, it is possible to reduce the defect rate in production and decrease the personnel costs and time delays associated with manual inspection.

Keywords: Computer Vision, Product Inspection, Quality Control, Labor Cost.

1. Introduction

The Taiwanese plastics industry has a development history spanning over 50 years, establishing a comprehensive industrial chain from upstream naphtha cracking to downstream plastic product processing. Polyvinyl chloride (PVC) is a critical material within this industry, and Taiwan was once the second-largest producer globally. However, increasing competition from China and other Southeast Asian countries has led to a decline in Taiwan's competitiveness in terms of scale. Although some companies continue to operate by leveraging advanced molding and processing technologies, maintaining both product quality and price competitiveness remains a significant challenge. Furthermore, the majority of Taiwanese plastic manufacturers are small and medium-sized enterprises (SMEs), which face substantial barriers to adopting Industry 4.0 initiatives. While many stakeholders understand and express interest in these innovations, they often lack the necessary resources and initial investments. This challenge is particularly pronounced in traditional industries, where the concept of technological innovation remains relatively unfamiliar. Moving forward, the development of plastic products will prioritize environmental sustainability and product quality, necessitating innovative approaches to enhance production standards.

The PVC shoe chopsticks investigated in this study are produced using an extrusion machine [1]. The manufacturing process involves melting granular PVC raw materials combined with additives and extruding the molten material through a mold to create continuous tubular products, where the mold defines the final product shape. Quality inspection for PVC shoe chopsticks is primarily conducted during the first and second stages of processing on the production line, focusing on parameters such as product dimensions (length and width), bending angles, and quantity statistics. This study aims to develop an automated inspection system to replace manual inspections, enhance product quality, and meet the demands of industry standards and market competitiveness.

2. Research Framework and Methodology

After capturing images on the production line, the inspection process is divided into three parts: image grayscale binarization, contour extraction with the drawing of the target object's bounding rectangle, and quality inspection. The quality inspection consists of three components: measurement of the target object's length and width, detection of bending angles, and quantity statistics with inspection reporting. The inspection framework is illustrated in Fig. 1.

2.1. Measuring the Dimensions of the Target Object

The principle of scale ratio [2] is used to calculate the actual spatial distance as the ratio between the image distance and the specified reference distance. This method, also known as the Pixel Ratio per Metric, determines the target object's actual dimensions based on the pixel size of a reference object. The process involves selecting the target object within the image and applying the scale ratio principle to calculate its actual dimensions.

2.2. Pipe Bending Angle

After applying Canny edge detection [3] on the image, the bounding rectangle of the contour is calculated using OpenCV's bounding rectangle API. In this study, the method is based on selecting the largest contour by pixel size. First, the set of contour points is obtained, and the rectangle's parameters—x, y, w, and h—are returned. Using the positions of the four contour points, circles and lines are drawn to outline the target object's bounding rectangle.

After determining the position of the bounding rectangle, three points are selected to calculate the angle of the target object.[4] First, the vector calculation for points A, B, and C is performed using formulas (2a, 2b, and 2c). Then, the cosine rule is applied using formula (2d). Finally, the angles between vectors AB and AC are individually calculated.

$$\mathrm{AB} = (b.x - a.x, b,y - a,y) \tag{2a}$$

$$\mathrm{AC} = (c.x - a.x, c,y - a,y) \tag{2b}$$

$$\cos A = (\mathrm{AB}*\mathrm{AC})/(|\mathrm{AB}| * |\mathrm{AC}|) \tag{2c}$$

$$\cos A = (b*b + c*c - a*a)/(2*b*c) \tag{2d}$$

2.3. Quantity Count

Since the bounding rectangle of the target object is already drawn using FindContours [5] during the measurement of the object's length and width, the quantity statistics are directly obtained by traversing all the bounding rectangles to count the number of target objects in each image.

2.4. Inspection Reportt

After completing the three types of inspections, Power Query retrieves the inspection data, analyzes it, and displays the results on the report UI, including the quantities of defective and non-defective products, inspection time, and the status of the cutting machine.

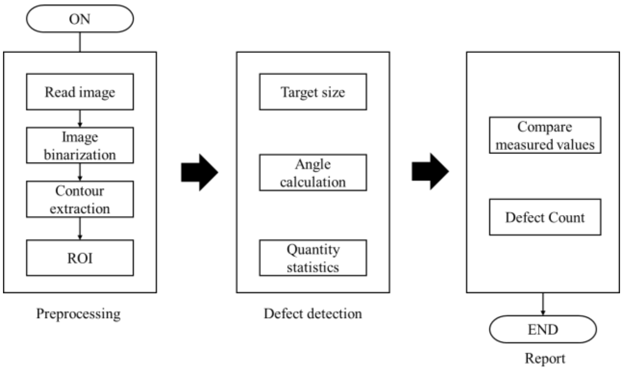


Fig. 1 Research Framework

3. Results and Discussion

In this study, we performed image recognition on production line images. The final results provide information on the target object's dimensions, processed bending angles, quantity statistics, and reports.

3.1. Target Object Dimensions

The primary inspection focuses on whether the cutting machine at the extrusion process's final stage can perform cutting according to the set angle. The results are shown in Fig. 2.

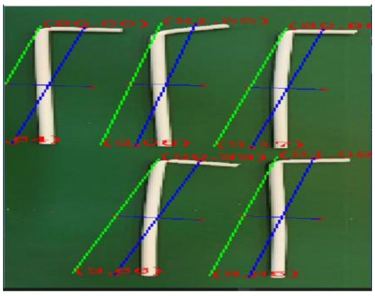


Fig 2 Dimensional Inspection Results

3.2. Processed Pipe Bending Angle

Since the shoe chopsticks are straight pipes after cutting, they require bending to be considered finished products. Therefore, measuring the bending angle, as shown in Fig, 3, serves as the final quality inspection step.

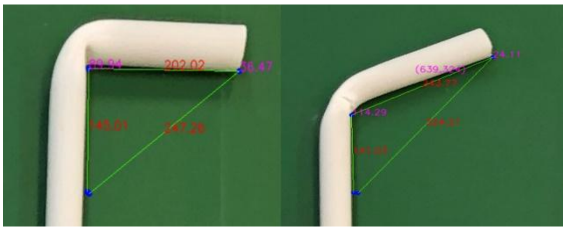


Fig 3 Bending Angle Inspection

3.3. Quantity Statistics

After the inspection is completed, the production quantities for each batch need to be statistically recorded, as shown in Fig. 4.

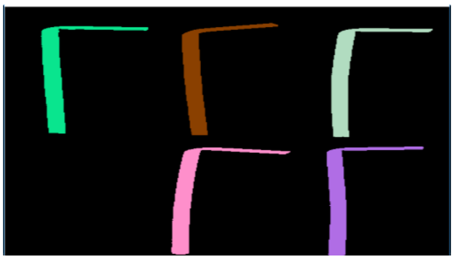


Fig 4 "Quantity Statistics"

3.4. Inspection Report

Since the first three tasks are all results of image recognition processing, a report is necessary to provide users with a clear understanding of the current production line status. This report enables monitoring of the shoe chopsticks' production quality and allows users to determine whether the production line machinery needs to be halted based on defect results. When defects in the shoe chopsticks occur, they are immediately displayed in the report, as shown in Fig. 5.

Inspection Report					
Target Quantity	Qualified Quantity	Unqualified Quantity	Yield Rate	Cutting Machine Status	Last Inspection Time
500	200 / 500	300 / 500	Low	Warning	Thu Jun 24 03:26:49 2024

Updata

a. Inspection Report

PVC-width	3.64	PVC-length	20.5
PVC-width	3.66	PVC-length	20.33
PVC-width	3.08	PVC-length	21.55
PVC-width	3.17	PVC-length	20.96
PVC-width	3.25	PVC-length	21.09
PVC-width	3.64	PVC-length	20.5
PVC-width	3.66	PVC-length	20.33
PVC-width	3.08	PVC-length	21.55
PVC-width	3.17	PVC-length	20.96
PVC-width	3.25	PVC-length	21.09
PVC-width	3.64	PVC-length	20.5
PVC-width	3.66	PVC-length	20.33
PVC-width	3.08	PVC-length	21.55
PVC-width	3.17	PVC-length	20.96
PVC-width	3.25	PVC-length	21.09

b. PVC Inspection Data

Fig 5 Inspection Report

4. Conclusion

This study successfully developed an automated quality inspection method for PVC shoe chopsticks by integrating computer vision technology into the production line. The proposed method addresses challenges in traditional manual inspections, such as high labor costs, inspection delays, and the inability to promptly identify production defects.

The method consists of three core steps: measurement of target object dimensions, bending angle detection, and quantity statistics. Through grayscale binarization, contour extraction, and advanced image processing techniques, the inspection method achieves high precision in defect detection and generates real-time inspection reports.

The findings demonstrate that the application of this method not only improves product quality and reduces defect rates but also provides actionable insights through detailed reports. These insights enable operators to monitor production line performance and make timely adjustments, ensuring compliance with industry standards and maintaining market competitiveness.

Future research directions may focus on expanding the method's applications, such as handling more complex product shapes, inspecting multiple production lines, and integrating with Industry 4.0 frameworks to further enhance production efficiency and quality control.

References

1. RONG SHAO Precision Plastic Injection Co., Ltd., Extrusion Machine Introduction Website.
2. Chapra, S. C., & Canale, R. P. (2021). Numerical Methods for Engineers (8th ed.). McGraw-Hill Education.
3. Bradski, G., & Kaehler, A., Learning OpenCV: Computer Vision with the OpenCV Library, O'Reilly Media, 2008.
4. Weisstein, E. W. (n.d.). Law of Cosines. MathWorld – A Wolfram Web Resource.
5. OpenCV Documentation, findContours Edge Detection.

Authors Introduction

Mr. Shu-Li Pai

He is currently a first-year Ph.D. student at the Graduate School of Engineering Science and Technology, National Yunlin University of Science and Technology. His research interests include big data, smart systems, and network applications.

Prof. Kuo-Hsien Hsia

He received the Ph.D. degree in electrical engineering from the National Sun Yat-Sen University, Taiwan, in 1994. He is currently an Associate Professor of National Yunlin University of Science and Technology in Taiwan. His research interests are in the area of mobile robotics, fuzzy control and image processing.

Dr. Chian-Cheng Ho

He received his Ph.D. in Electrical Engineering from National Chung Cheng University, Chiayi, Taiwan, in 2000. He is currently an Associate Professor at National Yunlin University of Science and Technology, Taiwan. His research interests include intelligent robotics.

Obstructing PLC Operations through Modbus Command Manipulation

Nai-Yu Chen
M.S. Degree Program on Cyber-Security Intelligence, National Cheng Kung University
No.1, University Rd., East Dist., Tainan City 701401, Taiwan

Cheng-Ying He
M.S. Degree Program on Cyber-Security Intelligence, National Cheng Kung University
No.1, University Rd., East Dist., Tainan City 701401, Taiwan

Jung-Shian Li
Department of Electrical Engineering / Institute of Computer and Communication Engineering,
National Cheng Kung University
No.1, University Rd., East Dist., Tainan City 701401, Taiwan

Chu-Sing Yang
Department of Electrical Engineering, National Cheng Kung University
No.1, University Rd., East Dist., Tainan City 701401, Taiwan

I-Hsien Liu
Department of Electrical Engineering, National Cheng Kung University
No.1, University Rd., East Dist., Tainan City 701401, Taiwan

E- mail: nychen@cans.ee.ncku.edu.tw, cyhe@cans.ee.ncku.edu.tw, jsli@cans.ee.ncku.edu.tw, csyang@ee.ncku.edu.tw, ihliu@cans.ee.ncku.edu.tw *
www.ncku.edu.tw

Abstract

Security vulnerabilities in Programmable Logic Controllers (PLCs) within Industrial Control Systems (ICS) using the Modbus/TCP protocol pose significant risks, particularly through stop-and-start command injection attacks that impact PLC operations and cause severe industrial consequences. Supported by Taiwan's National Science and Technology Council (NSTC) and the Water Resources Agency, this research establishes a cybersecurity testbed for water resource systems to investigate these threats. Unauthorized or forged commands are shown to manipulate PLC configurations and ladder logic diagrams, revealing critical weaknesses. Flowchart analyses and Modbus packet examinations highlight the risks and offer actionable insights into effective defense mechanisms for enhancing ICS security.

Keywords: Industrial Control Systems, Programmable Logic Controller, Modbus/TCP, Cybersecurity

1. Introduction

The integration of industrial control systems (ICS) with information technology (IT) has heightened the role of programmable logic controllers (PLCs) in automation [1]. However, this integration has also introduced significant cybersecurity risks. The Modbus protocol, designed for isolated industrial networks, lacks encryption and authentication, making it vulnerable in environments where Modbus TCP is widely used [2].

Recent cyberattacks, such as the 2022 Industroyer2 attack in Ukraine, have demonstrated these vulnerabilities [3]. Attackers exploited weaknesses in the Modbus TCP protocol to manipulate PLCs and disrupt critical infrastructure operations. Due to the protocol's lack of security features, such as encryption and authentication, it is highly susceptible to these types of attacks [4].

This study examines the CVE-2021-22779 vulnerability in PLC commands under the Modbus TCP protocol, involving an authentication bypass that allows unauthorized access via deceptive Modbus communication between engineering software and the controller. The vulnerability poses a significant security risk, enabling unauthorized command execution. Schneider Electric's TwidoSuite programming software was used to simulate various attack scenarios, demonstrating how system security can be compromised through manipulation of PLC configuration files and commands. Modbus TCP packets captured using Wireshark were analyzed to assess the impact of these attacks on ICS operations. Additionally, UMAS, a protocol based on Modbus TCP and commonly used in Schneider Electric devices, shares similar vulnerabilities, further emphasizing the importance of robust security measures to protect PLC systems from unauthorized access and manipulation [5].

This research aims to highlight the cybersecurity threats posed by such attacks and provide data to support future defense strategies, especially in environments limited by the inherent vulnerabilities of the Modbus protocol, where enhancing ICS security is critical.

2. Background

2.1. The risks of PLC's misconfigurations

In ICS, the configuration and programming of PLCs are essential for system security. Manufacturers provide development environments like Siemens' TIA Portal and TwidoSuite for PLC configuration. When integrated into ICS via Ethernet, default settings can expose vulnerabilities, allowing attackers to infiltrate the network [6] and execute unauthorized operations on PLCs [7].

Attackers can exploit TwidoSuite to control a PLC by configuring "Programming Mode" and entering the PLC's IP address to connect to the network. Once connected, TwidoSuite reveals configuration details, including I/O statuses and Ethernet statistics. Attackers can then download and modify the PLC program, adjusting I/O settings, creating new control commands, or altering parameters like temperature sensors or time controllers. They can upload the modified program back to the PLC, leading to malicious code execution and system failures.

TwidoSuite also allows attackers to monitor and capture operational data, including command executions and variable changes, facilitating continuous monitoring for more advanced attacks. These steps highlight the risks of using tools like TwidoSuite without proper security measures, emphasizing the need for enhanced PLC security in ICS environments [8].

2.2. Using Modbus Protocol to Control PLC

Modbus TCP is a widely used protocol in industrial control systems (ICS) for communication between PLCs and devices over TCP/IP networks [9]. It uses a TCP connection to transmit packets containing function codes and data, with responses indicating success. Modbus TCP packets consist of the Modbus Application Data Unit (ADU) and the Modbus Protocol Data Unit (PDU). The ADU contains the TCP header, address, function code, data, and checksum. The TCP header manages the connection, the address identifies the target device, the function code specifies the operation, the data field contains the relevant information, and the checksum ensures data integrity during transmission [10].

PLCs have two modes: Operation mode for normal tasks and Development mode for updates. Development mode, while useful, poses risks such as unauthorized memory access or control logic changes.

UMAS (Unified Messaging Application Services), Schneider Electric's proprietary protocol, extends Modbus TCP with custom function codes for advanced tasks like memory manipulation and program updates. However, both protocols lack encryption and strong authentication, leaving them vulnerable to interception, replay attacks, and manipulation.

3. Exploiting UMAS Protocol vulnerabilities to disrupt PLC control

This study examines how attackers can exploit vulnerabilities in the UMAS protocol to interfere with PLC control, sending malicious packets to achieve unauthorized operations. To simulate these attack scenarios, a PLC and the TwidoSuite programming software were utilized. The experimental setup included an unencrypted PLC connected to a network switch, with the switch configured to mirror all traffic to an analysis computer for monitoring. Wireshark was employed to filter and analyze the captured UMAS protocol packets [11] (Fig. 1).

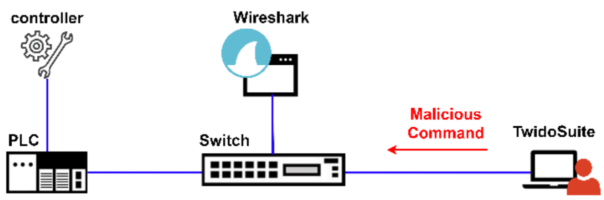


Fig. 1. PLC, Switch, Wireshark, and TwidoSuite Integration Architecture.

Using Wireshark, abnormal commands for stopping and starting the PLC, simulated through TwidoSuite, were intercepted to observe their impact on PLC operations. A packet with code 5a and function code 90 was detected. This function code, also known as "Unity Schneider," is a specialized command in Schneider Electric's systems used to execute stop and start instructions for the PLC [12].

Typical malicious commands include stopping and starting the PLC. The stop command, designed for emergencies, can be exploited to halt production or damage equipment. Conversely, the start command can lead to unintended operations or production accidents (Fig. 2).

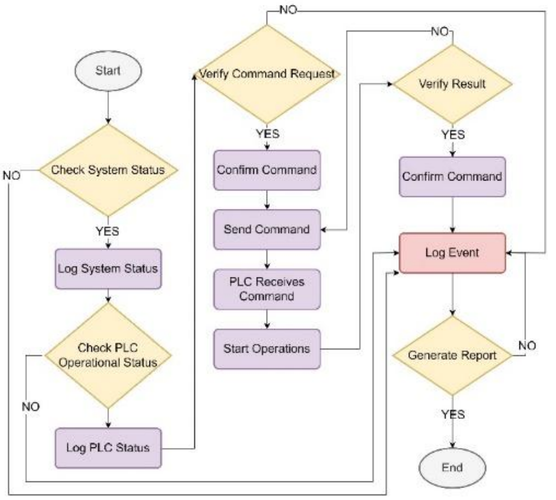


Fig. 2. The Flow Chart of PLC Stop Command Execution.

4. Experiments and Analysis

4.1. *Experimental Design*

The experimental design includes the following key steps:

1) PLC and Network Connection: Ensuring the PLC and switch are on the same network segment, with a mirror port routing traffic to the analysis computer, simulates an industrial control environment for monitoring PLC activities.

2) Malicious Command Simulation: Potential attack methods are simulated, including the modification or imitation of legitimate commands to send malicious instructions to the PLC. These actions are designed to disrupt or take control of system operations.

3) Packet Capture and Analysis: Wireshark is used to capture and analyze Modbus TCP packets from the PLC, focusing on data fields, function codes, and potential security risks.

4.2. *Experimental Results*

A detailed packet analysis was conducted on PLC communication, revealing that for the stopping PLC command, 213,229 packets were sent, with 212,840 successfully intercepted, resulting in a match ratio of 99.81%. Similarly, for the starting PLC command, 219,640 packets were sent, and 219,556 were intercepted, achieving a match ratio of 99.96%. These results indicate a high interception rate, demonstrating that the commands were effectively captured and potentially executed (Fig. 3). The analysis identified critical packet segments that require prioritization in the development of defense mechanisms and their associated potential attack methods (Table 1). To mitigate risks, recommended defense strategies include encryption to protect sensitive data, packet filtering to block unauthorized traffic, and anomaly detection systems to identify and respond to abnormal activities, thereby enhancing the overall security of the PLC system.

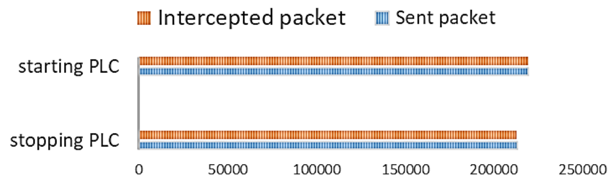


Fig. 3. PLC Command Packet Interception Comparison.

TABLE 1. Packet Header Segment Monitoring

Segment	Bytes	Possible attack methods
Ethernet Header Segment	0-13	MAC Spoofing, ARP Spoofing
IP Header Segment	14-33	IP Spoofing, IP Fragmentation Attack
TCP Header Segment	34-53	TCP Sequence Prediction, RST Attack, SYN Flood Attack
Data Segment	54-63	Command Injection, Data Manipulation

5. Conclusion

This study highlights the security risks of PLCs in ICS due to the lack of encryption and authentication protections. Simulated attack scenarios demonstrate that attackers can exploit PLC vulnerabilities to execute malicious actions, such as stopping or starting equipment, potentially disrupting processes or damaging machinery. Experimental results show that development mode enables attackers to manipulate configurations and ladder diagrams through unauthorized commands, while intercepting, simulating, or altering packets to send malicious instructions further compromises system integrity.

In addition to Modbus TCP, the study examines UMAS, a protocol based on Modbus TCP, revealing similar risks such as weak authentication and lack of encryption. Addressing these vulnerabilities requires measures like intrusion prevention systems, encryption protocols, and enhanced authentication. This research provides an empirical foundation for strengthening ICS security, with contributions to mitigating risks associated with both Modbus TCP and UMAS.

Acknowledgment

This work was supported by the National Science and Technology Council (NSTC) in Taiwan under contract numbers 113-2634-F-006-001-MBK, and by the Water Resources Agency (WRA) under the Ministry of Economic Affairs (MOEA) in Taiwan under contract number MOEAWRA1130243.

References

1. I-H. Liu, C.-C. Lai, J.-S. Li, C.-C.Wu, C.-F. Li, C.-G. Liu, "An Emulation Mechanism for PLC Communication

Features," Journal of Robotics, Networking and Artificial Life, Vol. 8, No. 3, pp. 175-179, 2021.

2. Wang, Yusheng et al. "Intrusion Detection of Industrial Control System Based on Modbus TCP Protocol," 2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS), 2017.
3. P. Kozak, I. Klaban, T. Šlajs, "Industroyer cyber-attacks on Ukraine's critical infrastructure," 2023 International Conference on Military Technologies (ICMT), Brno, Czech Republic, 2023.
4. K.-M. Sudar, P. Deepalakshmi, P. Nagaraj, V. Muneeswaran, " Analysis of Cyberattacks and its Detection Mechanisms," 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Bangalore, India, 2020.
5. L Rajesh, P. Satyanarayana. "Detection and Blocking of Replay, False Command, and False Access Injection Commands in SCADA Systems with Modbus Protocol," Security and Communication Networks, 2021.
6. I-H. Liu, K.-M. Su, J.-S. Li, "The Security Issue of ICS: The Use of IT Infrastructure," Journal of Robotics, Networking and Artificial Life, Vol. 8, No. 1, pp. 29-32, 2021.
7. K.-M. Su, I-H. Liu, J.-S. Li, "The Risk of Industrial Control System: Programmable Logic Controller Default Configurations", ICS 2020, Tainan, 17-19, December, 2020.
8. N.-Y. Chen, P.-W. Chou, J.-S. Li, I-H. Liu, "A Case Study of Network-Based Intrusion Detection System Deployment in Industrial Control Systems with Network Isolation", ICAROB 2024, Japan, 22-25, February,2024.
9. Ferst, Matheus K. et al. "Implementation of Secure Communication With Modbus and Transport Layer Security protocols," 2018 13th IEEE International Conference on Industry Applications (INDUSCON), 2018.
10. G. B. M. Guarese, F. G. Sieben, T. Webber, M. R. Dillenburg and C. Marcon, "Exploiting Modbus Protocol in Wired and Wireless Multilevel Communication Architecture," 2012 Brazilian Symposium on Computing System Engineering, Brazil, 2012.
11. Rahman, Ayesha et al. "Launch of denial of service attacks on the modbus/TCP protocol and development of its protection mechanisms," Int. J. Crit. Infrastructure Prot. 39, 2022.
12. S. Sandhya, S. Purkayastha, E. Joshua and A. Deep, "Assessment of website security by penetration testing using Wireshark," 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), India, 2017.

Authors Introduction

Ms. Nai-Yu Chen

She is a postgraduate of M.S. Degree Program on Cyber-Security Intelligence, National Cheng Kung University in Taiwan. She received her B.B.A. degree from the Bachelor of BioBusiness Management, National Chiayi University, Taiwan in 2021. Her interests are ICS Security and Network-Based Intrusion.

Ms. Cheng-Ying He

He is a postgraduate of M.S. Degree Program on Cyber-Security Intelligence, National Cheng Kung University in Taiwan. He received his B.B.A. degree from the Bachelor of Information Management, National Taiwan University of Science and Technology, Taiwan in 2022. His interests are ICS Security and Network-Based Intrusion.

Dr. Jung-Shian Li

He is a full Professor in the Department of Electrical Engineering, National Cheng Kung University, Taiwan. He graduated from the National Taiwan University, Taiwan, with B.S. in 1990 and M.S. degrees in 1992 in Electrical Engineering. He obtained his PhD in 1999 in Computer Science from the Technical University of Berlin, Germany. He teaches communication courses and his research interests include cybersecurity, cloud computing and network management. He is currently involved in funded research projects dealing with cybersecurity and critical infrastructure protection. He is the director of Taiwan Information Security Center @ National Cheng Kung University.

Dr. Chu-Sing Yang

He is a researcher in the Department of Electrical Engineering at National Cheng Kung University, Taiwan. He obtained his PhD in Electrical Engineering from National Cheng Kung University, Taiwan. His research interests include cybersecurity and network management.

Dr. I-Hsien Liu

He is an assistant professor in Department of Electrical Engineering, National Cheng Kung University, Taiwan. He obtained his Ph.D. in 2015 in Computer and Communication Engineering from the National Cheng Kung University. He teaches cybersecurity courses and his interests are Cyber-Security, OT Security, and Wired & Wireless Communication. He is the deputy director of Taiwan Information Security Center @ National Cheng Kung University (TWISC@NCKU).

Ultrasonic oscillator driver with digital frequency sweep function

Chung-Wen Hung, Yu-Hsing Su, Chun-Chieh Wang
National Yunlin University of Science and Technology, Taiwan
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C
Email: wenhung@yuntech.edu.tw, M11212048@yuntech.edu.tw, jasonccw@yuntech.edu.tw

Abstract

An Ultrasonic oscillator driver with digital frequency sweep function is proposed in this paper. Ultrasonic transducers are widely utilized in various applications, such as liquid atomization. The drive circuit causes the ultrasonic transducer to vibrate, while the attached atomization component converts the liquid into fine particles. In this paper, the LC resonant circuit is adopted to drive the ultrasonic transducer. Due to small variations in the resonant frequency of each transducer, the optimal operating frequency also varies and may change as physical conditions change. Then, microcontroller units (MCU) are used to control circuit switches to achieve frequency adjustment, scanning and tracking, so that the ultrasonic oscillator works in the best state.

Keywords: Ultrasonic Transducer, Frequency Scanning, Frequency Tracking

1. Introduction

Ultrasonic waves, which originated in the 19th century, are sound waves with frequencies exceeding 20 kHz. In recent years, the applications of ultrasonic waves have expanded significantly, encompassing medical tools such as ultrasonic scalpels [1] and industrial processes like ultrasonic cleaning and welding [2]. Common piezoelectric materials include Zinc Oxide (ZnO), Polyvinylidene Fluoride (PVDF), and Lead Zirconate Titanate (PZT). Among these, PZT is favored due to its higher piezoelectric coefficient, making it widely utilized in various applications. Ultrasonic transducers are constructed using PZT piezoelectric materials [3]. However, the resonance frequency of these materials can fluctuate due to changes in mechanical structure, component aging, and temperature increases resulting from prolonged operation [4], [5], [6]. To address this issue, this paper proposes an automatic frequency-tracking approach that periodically adjusts the transducer's frequency, ensuring it returns to the optimal oscillation point even after frequency drift, thereby maintaining adequate vibration output.

2. Load and System Architecture

2.1. *Ultrasonic Transducer Architecture*

The ultrasonic transducer selected for this study is a 50W, 28kHz bolt-clamped Langevin transducer (BLT) for testing purposes. This transducer is known for its high conversion efficiency and allows for selecting different operating frequencies based on specific applications [7].

This study analyzes the variation in impedance phase within the transducer by conducting a frequency sweep using an LCR meter. As shown in Fig. 1, the sweep results for a 28 kHz ultrasonic transducer reveal a resonance frequency of 28.06 kHz with an impedance of 29 Ω. To achieve sufficient vibration for atomizing water, an amplitude horn is added, as illustrated in Fig. **2**. The horn focuses the vibrations at a single point to produce a greater vibrational output. The sweep results after adding the horn are presented in Fig. **3**, where the overall impedance phase differs significantly, and the resonance frequency shifts to 27.22 kHz with an impedance of 49 Ω. The transducer equipped with the amplitude horn will be used as the test component in this study.

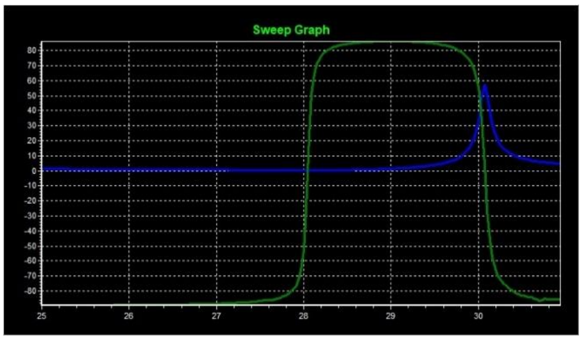


Fig. 1 Impedance Phase Plot of an Unloaded 28 kHz Transducer

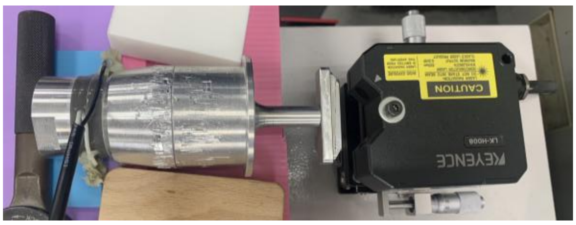


Fig. 2 Schematic diagram of transducer with amplitude horn

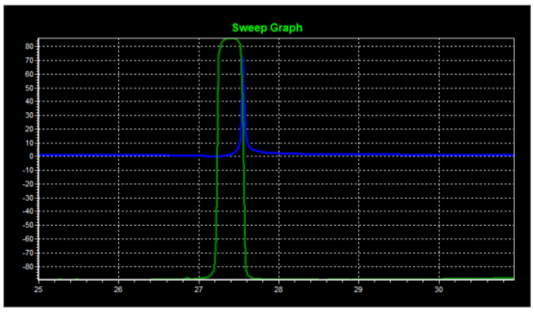


Fig. 3 Impedance phase plot of a 28kHz transducer with amplitude horn

2.2. *Hardware Architecture*

The hardware architecture of this study is illustrated in Fig. **4** The primary power source utilizes a power supply for voltage adjustment, enabling fine-tuning based on various oscillation effects. A half-bridge inverter with LC resonance functions as the main driving circuit, converting DC power into high-frequency AC at the desired frequency. The LC resonance effectively filters out higher harmonics, producing a sine wave for the transducer. The LC resonance frequency is calculated using Equation (**1**), where the LC components are optimized to achieve the best resonance. For the auxiliary power supply, a 25V input is converted to 18V and 3.3V using a buck converter to power the driver and the MCU, respectively. To monitor the transducer's current state, a current transformer (CT) is employed to detect current feedback to the MCU.

$$fr = \frac{1}{2\pi\sqrt{LC}} \quad (1)$$

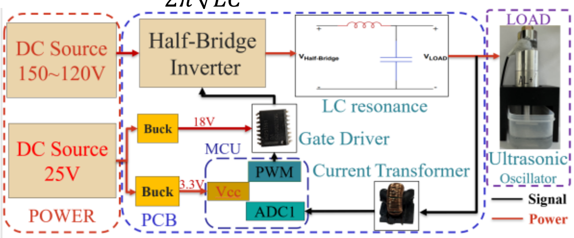


Fig. 4 Hardware Architecture Diagram

2.3. *Software Flowchart*

The resonant frequency range of the loaded transducer is approximately between 27 kHz and 27.5 kHz. As a capacitive component, the transducer exhibits slight frequency variations that may drift due to factors such as temperature fluctuations and long-term usage deterioration. In this study, a microcontroller unit (MCU) performs digital frequency sweeping and periodically monitors the frequency after each sweep. Fig. **5** illustrates the frequency-tracking flowchart: after identifying the operating frequency within the sweep range, a sweep of ±0.5 kHz around the current operating frequency is conducted every 3 minutes. The frequency with the highest current in this range is selected as the new operating frequency. The interval is set at 3 minutes to prevent excessive tracking, which could lead to rapid temperature increases in the transducer and result in adverse effects.

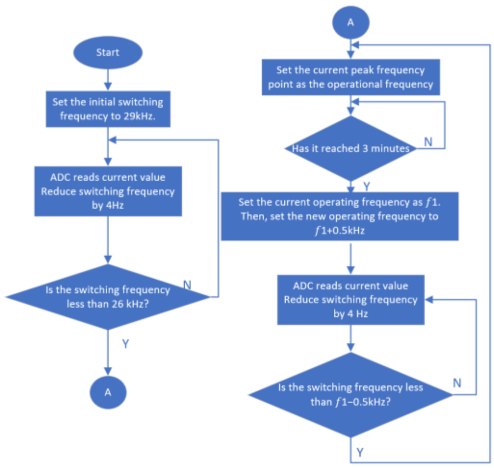


Fig. 5 Frequency Tracking Flowchart

3. Results and Discussion

In this study, an actual circuit is employed with a frequency-tracking method to drive the transducer, recording changes in temperature and transducer power over time to validate the feasibility of the tracking approach.

Since different transducers exhibit slight variations in resonant frequency, frequency sweeping is conducted to identify the optimal frequency for each transducer. Fig. 6 illustrates the actual sweep waveforms, with the red waveform representing transducer voltage and the green waveform representing transducer current. The frequency point with the highest current within a specific range is selected as the operating frequency for subsequent operations.

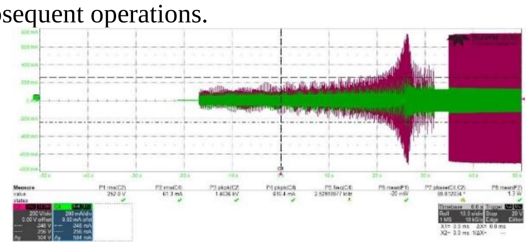


Fig. 6 Frequency Sweep Waveform

Table 1 presents the results of frequency tracking conducted every 3 minutes over a total duration of 9 minutes. The results indicate that the transducer's power decreases as operating time increases due to a shift in the optimal frequency. However, with frequency tracking, the transducer returns to its optimal operating point, nearly restoring its initial power level. The transducer's vibration amplitude, illustrated in Fig. 7, decreased by $9\mu m$ before tracking but returned to a higher state after tracking, as shown in Fig. **8**, demonstrating the effectiveness of the frequency-tracking method. To examine the overall impedance drift, an LCR meter was utilized to compare the transducer's impedance and phase before Fig. 9 and after testing Fig. 10 These figures demonstrate that, following a brief test, the zero-phase frequency shifted from 27.67 kHz to 27.11 kHz. The experimental results indicate that both frequency and vibration amplitude change over time, and frequency tracking effectively mitigates this issue.

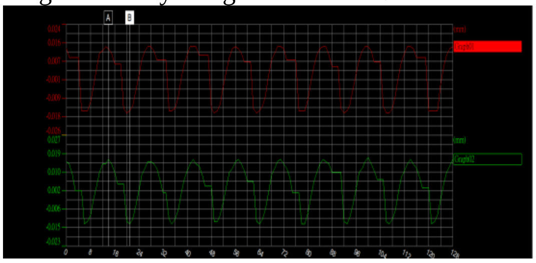


Fig. 7 Vibration Amplitude Before and After Damping

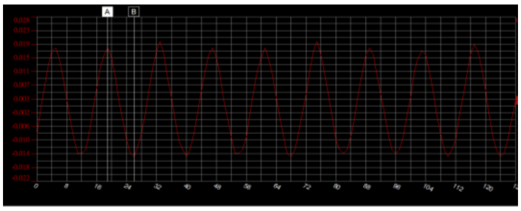


Fig. 8 Vibration Amplitude After Frequency Tracking

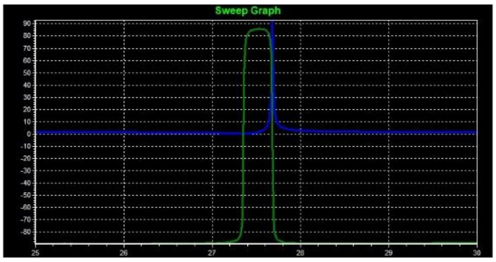


Fig. 9 Before Testing

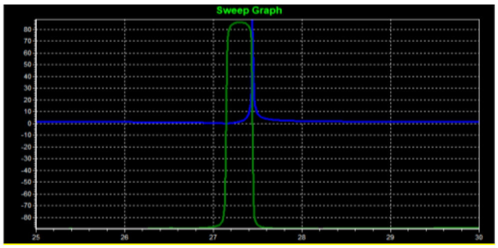


Fig. 10 After Testing

Table 1 Frequency Tracking Results

TIME (min)	V_{in} (DC)	L (mH)	*C* (*nF*)	f_r (kHz)	f_s (kHz)	V_{PZ} (V)	I_{PZ} (mA)	T_{PZ} (°C)	T_L (°C)	T_{water} (°C)	P_{in} (W)	P_{PZ} (W)
Start	100	3.3	9.4	30.2	27.43	605.0	107.0	27.6	25.3	28.6	44.7	40.4
1					27.42	645.5	115.8	34.1	28.6	41.3	47.0	42.8
3					27.39	588.0	125.9	35.0	28.3	43.7	43.6	39.4
9					27.30	379.0	109.5	39.8	28.7	45.2	29.0	27.6

4. Conclusion

This paper employs a digital frequency sweeping method to identify the optimal frequency point for various transducers under different loads and incorporates a frequency tracking feature. When frequency drift occurs, corrections are made at fixed intervals to maintain a consistent vibration amplitude. Experimental results demonstrate the effectiveness of this method. However, maintaining long-term frequency tracking for system stability is currently not feasible, likely due to factors such as temperature fluctuations, changes in mechanical structure, and the inability of the initially matched LC components to achieve optimal resonance during frequency drift. Future research will focus on addressing these potential issues.

5. References

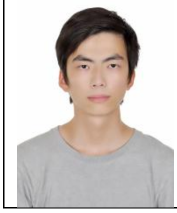
1. Tsujino, J., Ueoka, T., Hasegawa, K., Fujita, Y., Shiraki, T., Okada, T., & Tamura, T. (1996). New methods of ultrasonic welding of metal and plastic materials. *Ultrasonics, 34(2-5), 177-185.*
2. Li, X., Stritch, T., Manley, K., & Lucas, M. (2021). Limits and opportunities for miniaturizing ultrasonic surgical devices based on a Langevin transducer. *IEEE Transactions on Ultrasonics*, Ferroelectrics, and Frequency Control, 68(7), 2543-2553.
3. I. C. Lien, Y. C. Shu. (2008, Nov). Introduction to Vibration-Base Piezoelectric Energy Harvesting. *Industrial Materials Magazine*.
4. Yang, T., Zhu, Y., Li, S., An, D., Yang, M., & Cao, W. (2020). Dielectric loss and thermal effect in high power piezoelectric systems. *Sensors and Actuators A: Physical*, 303, 111724.
5. DeAngelis, D. A., & Schalcosky, D. C. (2006, October). P2O-10 The Effect of PZT8 Piezoelectric Crystal Aging on Mechanical and Electrical Resonances in Ultrasonic Transducers. In *2006 IEEE Ultrasonics Symposium* (pp. 1935-1938). IEEE.
6. Kazys, R., & Vaskeliene, V. (2021). High temperature ultrasonic transducers: A review. *Sensors*, 21(9), 3200.
7. Piezoelectric Components Business Unit. (2021, March 7). Bolt Clamped Langevin Transducer. https://www.unictron.com/piezoelectric-components/piezoelectric-technologies/bolt-clamped-langevin-transducer/?lang=zh-hant

Authors Introduction

Dr. Chung-Wen Hung

He received the Ph.D. degrees in Electrical Engineering from National Taiwan University in 2006. Currently he is a Professor in National Yunlin University of Science &Technology. His research interests include the IoT, IIoT, and AI application.

Mr. Yu-Hsing Su



He received the B.S. degrees and now he is studying for the M.S. degree in electrical engineering from National Yunlin University of Science and Technology.

Dr. Chun-Chieh Wang

He received the PhD degree in The graduate school of engineering science and technology from Nation Yunlin University of Science and Technology. Currently, he is a project professor at National Yunlin University of Science and Technology.

Design and development of foot pressure sensing massage stick

Yuting Hsiao, Dengchuan Cai, Chung-Wen Hung
Nation Yunlin University of Science and Technology, Taiwan
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C

Email: ergoresearch119@gmail.com, caidc@yuntech.edu.tw, wenhung@yuntech.edu.tw

Abstract

There are three key points in the operation of the Foot Massage Stick: sequence, direction and force. However, the force is not easy to be described and not easy to be learned. In order to provide the operator with visualization and information about the force during the operation, a pressure sensing massage stick was developed in this study. The features developed in this study are as follows. 1) The function of the assistive device is in line with the precision level of commercially available instruments. 2) The force and time duration of the operation can be displayed at any time during the execution process. 3) At the end of the execution, the maximum, average, standard deviation, and time duration of the force of the operation can be presented. This visualizes the force of the operation and makes it easy for the learner to check and meet the requirements for use.

Keywords: Foot Massage, assistive device, pressure sensing massage stick, force

1. Introduction

A foot massage is a type of massage that applies pressure to specific points on the feet and can be performed by a professional therapist or at home. There are two types of foot massage, one is Manual therapy (MT) and the other is using massage sticks. However, hand injuries are the second most common work-related musculoskeletal injuries among Manual therapy (MT) professionals [5]. This is because during massage, the hands must maintain repetitive movements and high-velocity forces to accomplish the task [6].

According to Cornwell et al., (2021) in the American Physical Therapy Association, 38.5% of Physical Therapists (PTs) suffered from Work-Related Musculoskeletal Disorders (WRMD). Among them, manual therapy (MT) caused the highest proportion of damage [3].

Albert, Currie-Jackson, & Duncan, (2008) surveyed 502 Canadian Registered Massage Therapists (RMTs) and found that musculoskeletal pain and discomfort was highest in the wrist and thumb, followed by the lower back, neck, and shoulders [1]. To effectively reduce the risk of musculoskeletal injury, preventive measures such as manual treatment or the development of new tools or equipment should be reduced [7].

The use of massage stick can replace the movements that cannot be operated by Manual Therapy (MT), including rolling, pushing, lifting, pressing, etc., and it can also press the acupoints more accurately. Therefore, the effect and operation satisfaction of massage stick is preferred by most of the massage practitioners. It can also reduce the incidence of hand pain and injury [2].

The technique of massage stick operation has three key points: sequence, direction and force [4]. Sequence and direction can be recorded visually, whereas force is less easily described and learned. The aim of this study was to develop a foot pressure sensing massage stick. It allows learners and operators to have visual information to refer to, and the power to record the operation process to assist the operator's learning and operation quality.

2. Methodology

The development of the foot pressure sensing massage stick in this study followed the product development procedure. At first, based on the analysis of market data, the study of existing technology principles and the investigation of usage requirements, the design objectives and specifications are formulated, and the functional design is carried out to achieve the design objectives. In this study, two technical applications such as piezoelectric elements and strain gauge elements are used as the basic components for the design. Secondly, the human factors design of the assistive device should be applied to the operation mode and human factors needs of the employees. Finally, by integrating the aforementioned functional design and ergonomic design, the development and pleasing appearance will meet the proposed design objectives and specifications.

3. Design Results

This study went through 5 designs and revisions, resulting in the development of an assistive device that could record the force exerted and met the design objectives. This model foot pressure sensing massage stick is the result of design 1-3 modifications. The pressure value is sensed by a single point pressure sensor (Fig.1). The internal structure has a pressure sensing mechanism in three directions: x, y, and z-axis. The topmost structure of a massage stick is utilized to sense pressure in three axes. In order to optimize the effect of the massage stick in transmitting pressure to the sensor, four different shapes of the top part of the massage stick were made for testing (Fig.2). The result of the test was that the top part of the fourth massage stick was the most effective, and could detect the pressure value of each foot massage technique (Fig.3).

Fig. 1 Single Point Pressure Sensor

Fig. 2 Four different shapes of the massage stick top part.

Fig. 3 foot pressure sensing massage stick testing

3.1. Final Design

This pressure sensing massage stick is the final modification result. This design is rechargeable and is charged using the USB connector. The applied data can be displayed on the small display of the massage stick, eliminating the need for an external large screen to display the measured data. During the foot massage, the massage aid itself will display the pressure value and the time of application at any time. When the operation program is completed, the maximum value, average value, standard deviation and total application time of the force applied during this operation will be displayed on the screen(Fig.4, Fig.5). Fig.6 shows the appearance of the final design entity. Fig.7 shows the internal electronics of the pressure sensing massage stick. Fig.8 shows the charging device of pressure sensing massage stick. Fig.9 shows the external dimensions of the final design of pressure sensing massage stick.

Fig. 4 Pressure value data display

Fig. 5 Pressure value data display

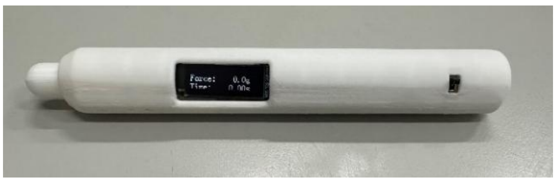


Fig. 6 pressure sensing massage stick final design of the physical appearance

Fig. 7 pressure sensing massage stick internal Electronic Components

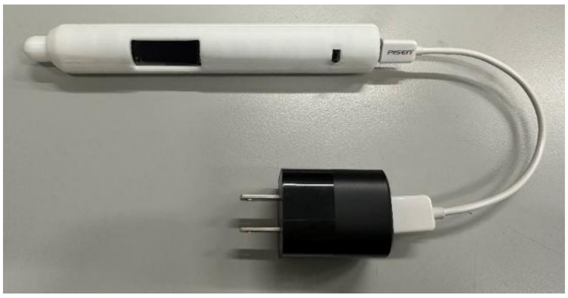


Fig. 8 pressure sensing massage stick charging device

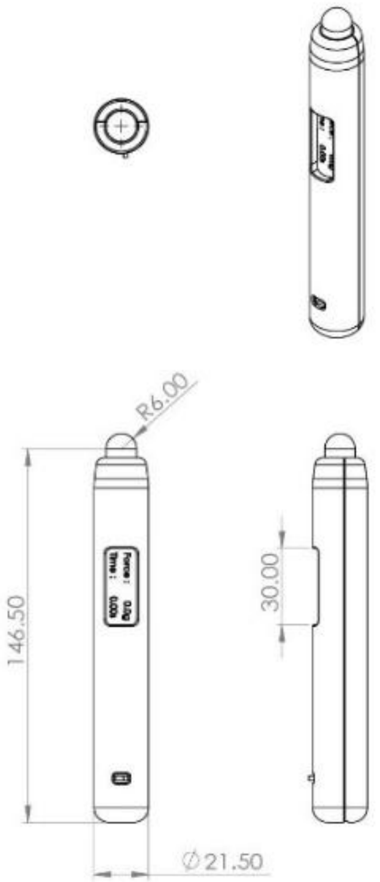


Fig. 9 External dimensions of the final design

3.2. Test and Verification

In order to test the functionality of the pressure sensing massage stick, a compression test was conducted in this study. Pressure tests were conducted using a commercially available pressure measurement device from G-CHEN Technology Corp. and a pressure sensing massage stick produced in this study. The test was conducted by pressing the top end of the pressure sensing massage stick developed in this study down on a G-CHEN Technology Corp. pressure pad and applying a force of 1kg. The results of the measurements were compared to the force values presented by the two devices. The results showed that the force value of 1kg was the same for both devices, confirming their validity. Fig.10 shows the test situation, the left side of the figure shows the pressure value of G-CHEN Technology Corp. equipment, and the right side is the final design of the pressure sensing massage stick applying force of 1kg. Fig.11 applies 1000g (1kg) of pressure to the pressure sensing massage stick. Fig.12 shows the result of pressure sensing massage stick pressure recording. 1kg of pressure is shown on the screen. Fig.13 shows the results of pressure sensing equipment testing by G-CHEN Technology Corp.

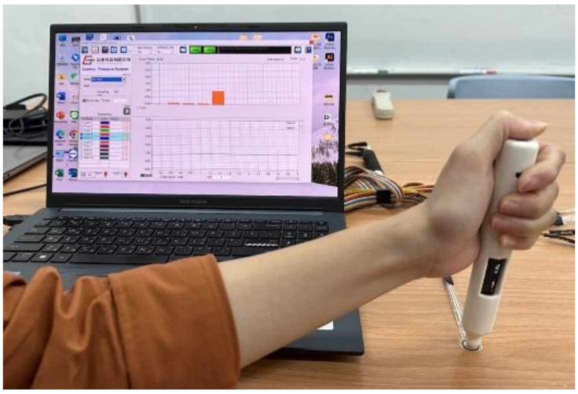


Fig.10 The left side shows the pressure value of the company's software, and the right side shows the 1kg force test of this design.

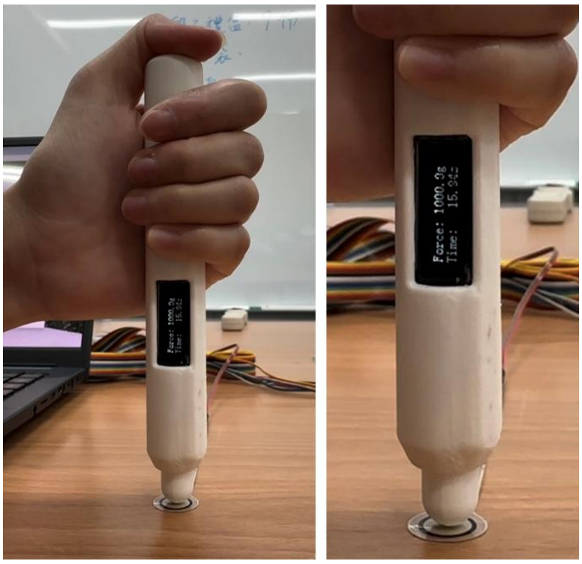


Fig.11 Pressure sensing massage stick applied 1000g of pressure.

Fig.12 Pressure sensing massage stick test results

Fig.13 G-CHEN Technology Corp. pressure Sensing Equipment Test Results

4. Conclusion

This study went through five designs and revisions to develop an assistive device that could record the force of application and meet the design objectives. The design features are as follows: 1) The function of the assistive device is in line with the precision level of commercially available instruments. 2) The force and time duration of the operation can be displayed at any time during the execution process. 3) At the end of the execution, the maximum, average, standard deviation, and time duration of the force of the operation can be presented. This visualizes the force of the operation and makes it easy for the learner to check and meet the requirements for use.

Acknowledgements

The authors thank the National Science and Technology Council of ROC (NSTC112-2410-H-224-019) for financial support.

References

1. Albert, W. J., Currie-Jackson, N., & Duncan, C. A. (2008). A survey of musculoskeletal injuries amongst Canadian massage therapists. Journal of bodywork and movement therapies, 12(1), 86-93.
2. Caragianis, S. (2002). The prevalence of occupational injuries among hand therapists in Australia and New Zealand. Journal of hand therapy, 15(3), 234-241.
3. Cornwell, L., Doyle, H., Stohner, M., & Hazle, C. (2021). Work-related musculoskeletal disorders in physical therapists attributable to manual therapy. Journal of Manual & Manipulative Therapy, 29(2), 92-98.
4. Eugster, J., & Hu, C.W. (2017). Self-healing Foot Care: Father Josef 's Method of Reflexology. Cosmax Publishing Co., Ltd. New Taipei City.
5. Gyer, G., Michael, J., & Inklebarger, J. (2018). Occupational hand injuries: a current review of the prevalence and proposed prevention strategies for physical therapists and similar healthcare professionals. Journal of integrative medicine, 16(2), 84-89.
6. Glover, W., McGregor, A., Sullivan, C., & Hague, J. (2005). Work-related musculoskeletal disorders affecting members of the Chartered Society of Physiotherapy. Physiotherapy, 91(3), 138-147.
7. Sharan, D., & Ajeesh, P. S. (2012). Injury prevention in physiotherapists-a scientific review. Work, 41(Supplement 1), 1855-1859.

Authors Introduction

Ms. Yuting Hsiao

She received her M.S. degree in design from National Yunlin University of Science and Technology. Currently she is a Ph.D. student in the School of Design of National Yunlin University of Science and Technology.

Prof. Dengchuan Cai

He is a professor in Department of Industrial design, National Yunlin University of Science and Technology, Taiwan. He received his Ph.D. degrees in management from Taiwan University of Science and Technology and M.S. degree in engineering from National Chiao Tung University, Taiwan.

Dr. Chung-Wen HUNG

He received the Ph.D. degrees in Electrical Engineering from National Taiwan University in 2006. Currently he is a Professor in National Yunlin University of Science &Technology. His research interests include the IoT, IIoT, and AI application.

Effects on physiological indicators of foot massage using a pressure sensing massage stick

Dengchuan Cai, Yuting Hsiao, Chung-Wen Hung
Nation Yunlin University of Science and Technology, Taiwan
123 University Road, Section 3, Douliu, Yunlin 64002, Taiwan, R.O.C

Email: caidc@yuntech.edu.tw, ergoresearch119@gmail.com, wenhung@yuntech.edu.tw

Abstract

Foot massage is often used as a complementary and alternative medicine. This study uses a special foot pressure-sensing massage stick to massage the soles of the feet, and tests the effects on the physiological indicators of the massaged person. After 3 sessions of massage, the results demonstrated that the assistive stick and the foot massage method used in this study were verified. The sleep quality score and body energy index were improved significantly with increasing massage frequency, whereas the respiration rate, heart rate, and blood oxygen were decreased significantly with increasing massage frequency. The psychological stress and physical ages were not significantly changed.

Keywords: Reflexology, Ergonomic design, complementary and alternative medicine, human factors, mechanical design, healthcare.

1. Introduction

Complementary and alternative therapies are increasingly recognized as a safe and effective way to reduce pain and illness [1]. It is a non-invasive and affordable form of health care that is beneficial to most people, including children, the elderly, cancer patients and pregnant women. Can be used [2]. The World Health Organization (World Health Organization, 1990) pointed out that combining traditional drug-free podiatry medicine with modern surgical medicine can help people understand their own health and increase medical security. In many countries, the reflexology and beauty industry is associated with unorthodox medicine.

Complementary and alternative therapies are actively used in healthcare settings in hospices, nursing homes and obstetrics. Attitudes towards complementary and alternative therapies are shifting around the world towards their use as secondary health care and their integration into mainstream medicine [1]. The British survey found that the average one-year use rate of auxiliary therapy is 41.1% [3]. The use rate of elderly people in the United States is 23%-62.9% [4]. The one-year use rate in Taiwan is 85.65%.

Cai et al., [5] searched the Web of Science for 31 literature published on foot reflexology from 1991 to 2021. Use "foot massage" or "reflexology" to search for all academic publications on foot reflexology, and a total of 801 papers were retrieved. The average annual publication volume in the first six years was 3.5 articles, and the average annual publication volume in the last six years was 72 articles, which was 20 times that of the first six years. It can be seen that this research topic has recently been receiving global attention and investing research resources.

Foot reflexology has been suggested to enhance blood flow, increase relaxation, and improve healing [6]. It is considered to be the most commonly used complementary and alternative therapy by the elderly [7]. It is also the complementary and alternative therapy that young people love to use [8]. Currently, foot reflexology in Taiwan is loved by a wide range of users and is used in daily life, becoming a part of complementary and alternative medicine.

The application areas of foot massage include: toes, inside of foot, instep, outside of foot, and sole of foot. Since the sole of the foot contains the most acupuncture points, has the best massage effect, and is easier to perform, this study first explores the massage of the sole of the foot and explores the impact of foot massage on physiological indicators.

2. Methodology

2.1. Experimental Design

The experiment is a 15-day period, and the detail plan is shown in Figure 1. During this period, the subjects were asked to wear Garmin Forerunner 945watches, record physiological indexes, and be responsible for keeping the Garmin watches.

During the study period, subjects are required to be foot massaged three times, on days 4, 8, and 12 respectively. Each application takes about 40 minutes and is performed by a professionally trained foot reflexology operator. During the experimental study, the subjects continued to work and rest as usual.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Start	====>	====>	Mas1	====>	====>	====>	Mas2	====>	====>	====>	Mas3	====>	====>	End
No action			Effects of Masssage 1				Effects of Masssage 2				Effects of Masssage 3			

Fig.1 the experimental plan

Fig.2 the recorder of physiological indexes.

2.2. Subjects

This study recruits 6 volunteer adults (3 males and 3 females), over 20 years old with no obvious trauma as subjects.

2.3. Experimental Instruments

There are two main instruments used in this experiment: the GARMIN Forerunner 945 wearable device (Figure 2) and the massage assistive device, a foot massage stick developed by this study (Figure 3). The massage parts on the sole of the feet was showed in Figure 4. The experimental location was the Ergonomic Research Laboratory of National Yunlin University of Science and Technology, Taiwan.


Fig.3 the massage stick.

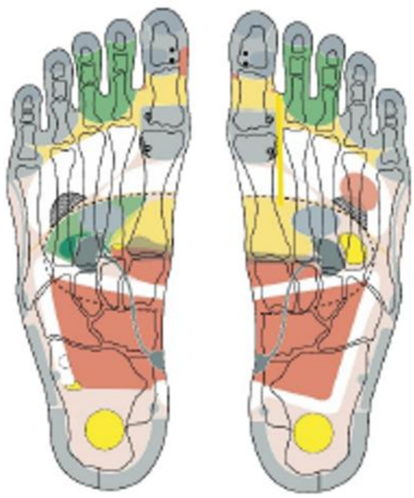


Fig. 4 The massage part on the sole of the feet.

2.4. Variables

Experimental variables: The independent variable of this study is the stages: stage 0 (no application; days 1, 2, 3), stage 1 (first massage on day 4; days 4, 5, 6, 7), stage 2 (second massage on day 8; days 8, 9, 10, 11 days), and stage 3 (third massage on day 12; days 12, 13, 14, 15). The plan of massage schedule was showed in Figure 1.

The dependent variables include: 1) sleep quality score, 2) psychological stress, 3) body energy index, 4) physical age, 5) respiration rate, 6) heart rate, and 7) blood oxygen.

3. Results and Discussion

3.1. Results

The age of the subjects was calculated by genders and showed in Table 1. The age ranged from 44-58 years with a mean age of 51.50 (SD= 5.92) years.

Table 1. the statistics of the age of the subjects

Genders	N	min	max	mean	SD
Male	3	50.0	58.0	53.67	4.04
Female	3	44.0	58.0	49.33	7.57
Total	6	44.0	58.0	51.50	5.92

The statistic results of the seven dependent variables are calculated based on the 4 stages: stage 0 (no action), stage 1 (first massage), stage 2 (second massage), and stage 3 (third massage) of the independent variable and showed in Table 2-8, respectively.

Table 2 showed the results of the sleep quality scores of the 4 stages. The sleep quality scores of stages 0, 1, 2, and 3 were 59.06, 71.21, 76.04, and 81.54, respectively. The sleep quality scores from 59.06 of stage 0 increased up to 81.54 of stage 3.

The table showed that the sleep quality of stage 2 and stage 3 were significantly greater than that of stage 0, and the sleep quality of stage 1 was equal to that of stage 0, 1, and 2. That means the sleep quality were improved when massaged twice and more.

Table 2. the statistics of sleep quality scores

Stages	N	min	max	mean	SD	t-tests
Stage 0	6	40	73.3	59.06	13.43	A
Stage 1	6	44	85.3	71.21	16.30	AB
Stage 2	6	54	87.5	76.04	11.82	B
Stage 3	6	73	91.8	81.54	6.49	B

The results of body energy indexes of the 4 stages were showed in Table 3. The body energy indexes of stages 0, 1, 2, and 3 were 68.28, 79.21, 83.50, and 84.17, respectively. The body energy indexes from 68.28 of stage 0 increased up to 84.17 of stage 3.

The table showed that the body energy indexes of stage 1, stage 2 and stage 3 were significantly greater than that of stage 0. That means the body energy indexes were improved when massaged for the first time, and improved better for massage twice and more.

However, the body energy indexes of stage 2 and stage 3 were equally. This signified that the body energy indexes for more than twice massage, the body energy indexes can maintain a constant and is not increasing continued.

Table 3. the statistics of body energy index

Stages	N	min	max	mean	SD	t-tests
Stage 0	6	52.7	74.7	68.28	8.56	A
Stage 1	6	68.3	91.8	79.21	7.75	B
Stage 2	6	67.3	94.3	83.50	9.01	C
Stage 3	6	76.0	88.8	84.17	5.00	C

The results of respiration rate of the 4 stages were showed in Table 4. The respiration rate of stage 0, 1, 2, and 3 were 14.00, 13.71, 13.54, and 13.33, respectively. The respiration rate decreased from 14.00 down to 13.33.

The table showed that the respiration rate of stage 3 was significantly greater than that of stage 0. That means the respiration rate was improved only when massaged for third time.

The respiration rate of stages 1 was equal to that of stage 2 and not greater than that of stage 0 significantly. This means that the massage effect was not significantly strong on respiration rate.

Table 4. the statistics of respiration rate

Stages	N	min	max	mean	SD	t-tests
Stage 0	6	14	14.0	14.00	0.00	A
Stage 1	6	13	14.0	13.71	0.29	AB
Stage 2	6	13	14.0	13.54	0.51	AB
Stage 3	6	13	14.0	13.33	0.52	B

The results of heart rate of the 4 stages were showed in Table 5. The heart rate of stage 0, 1, 2, and 3 were 76.39, 73.08, 71.42, and 67.92, respectively. The heart rate decreased from 76.39 down to 67.92.

The table showed that the heart rate of stage 1 was significantly greater than that of stage 0. That means the heart rate was not improved when massaged for the first time.

The heart rate of stages 2 was significantly greater than that of stage 0 and the heart rate of stages 3 was significantly greater than that of stage 0, stage 1, and stage 2. This means that the massage effect was significantly strong on heart rate when massage twice and more.

Table 5. the statistics of heart rate

Stages	N	min	max	mean	SD	t-tests
Stage 0	6	66	84.3	76.39	7.85	A
Stage 1	6	65	79.5	73.08	5.73	AB
Stage 2	6	65	76.8	71.42	4.54	B
Stage 3	6	59	71.8	67.92	5.01	C

The results of blood oxygen. of the 4 stages were showed in Table 6. The blood oxygen of stage 0, 1, 2, and 3 were 94.17, 93.42, 93.25, and 93.00, respectively. The blood oxygen decreased from 94.17 down to 93.00.

The table showed that the blood oxygen of stage 1, stage 2, and stage 3 were significantly greater than that of stage 0. That means the blood oxygen was improved when massaged for the first time.

However, the blood oxygen of stages 3, stage 2, and stage 1 were the same. That means that the blood oxygen was not significantly improved when massage twice and more.

Table 6. the statistics of blood oxygen.

Stages	N	min	max	mean	SD	t-tests
Stage 0	6	93	95.0	94.17	0.75	A
Stage 1	6	92	94.8	93.42	1.16	B
Stage 2	6	92	94.0	93.25	0.88	B
Stage 3	6	92	94.0	93.00	0.99	B

The results of psychological stress of the 4 stages were showed in Table 7. The psychological stress of stage 0, 1, 2, and 3 were 37.44, 35.29, 34.46, and 32.50, respectively. The psychological stress decreased from 37.44 down to 32.50.

The table showed that the psychological stress of stage 1, stage 2, and stage 3 were not significantly greater than that of stage 0 although the psychological stresses were slightly decreased. That means the psychological stress was not improved when foot massaged was applied for people regardless the massage frequency.

Table 7. the statistics of psychological stress

Stages	N	min	max	mean	SD	t-tests
Stage 0	6	30	49.3	37.44	6.80	A
Stage 1	6	25	46.3	35.29	8.20	A
Stage 2	6	24	44.3	34.46	7.99	A
Stage 3	6	27	37.5	32.50	3.80	A

The results of physical ages of the 4 stages were showed in Table 7. The physical ages of stage 0, 1, 2, and 3 were

50.75, 49.48, 49.13, and 48.83, respectively. The physical ages decreased from 50.75 down to 48,83.

The table showed that the physical ages of stage 1, stage 2, and stage 3 were not significantly greater than that of stage 0, although the physical ages had declined slightly. That means the physical ages was not improved when foot massaged was applied for people regardless the massage frequency.

Table 8. the statistics of physical ages

Stages	N	min	max	mean	SD	t-tests
Stage 0	6	42.2	57.3	50.75	6.01	A
Stage 1	6	42.0	57.0	49.48	5.84	A
Stage 2	6	42.0	57.0	49.13	5.52	A
Stage 3	6	42.0	57.0	48.83	5.27	A

3.2. Discussions

The results showed that the sleep quality score and body energy index were increased gradually according to the of massage times. On the contrary, the respiration rate, heart rate, and blood oxygen were decreased according to the of massage frequency. On the other hand, the psychological stress and physical age were not significantly changed although the data was slightly different. This illustrated that the effects of the physiological indicators were improved after foot massage using the assistive stick designed by this study.

4. Conclusion

This study conducted a three-stage foot massage using a pressure-sensing massage stick designed by the study. The results and procedure of this study demonstrated that the assistive stick designed by the study and the foot massage method operated in this study were verified. The results illustrated that the sleep quality score and body energy index were improved significantly with increasing massage frequency, whereas the respiration rate, heart rate, and blood oxygen were decreased significantly with increasing massage frequency. The psychological stress and physical ages were not decreased significantly according to the of massage frequency. The results could be a reference for health supplies design.

Acknowledgements

The authors thank the National Science Council of Taiwan, ROC for financial support under the project No. NSTC 112-2410-H-224-019 and all the researchers and participants for their participation.

References

1. McCullough, J., Liddle, S., Sinclair, M., Close, C., Hughes, C. J. E.-B. C., & Medicine, A. (2014). The physiological and biochemical outcomes associated with a reflexology treatment: a systematic review. 2014.
2. Tiran, D., & Mackereth, P. A. (2010). Clinical reflexology: a guide for integrated practice: Elsevier Health Sciences.
3. Posadzki, P., Watson, L. K., Alotaibi, A., & Ernst, E. J. C. m. (2013). Prevalence of use of complementary and alternative medicine (CAM) by patients/consumers in the UK: systematic review of surveys. 13(2), 126.
4. Groden, S. R., Woodward, A. T., Chatters, L. M., & Taylor, R. J. J. T. A. J. o. G. P. (2017). Use of complementary and alternative medicine among older adults: differences between baby boomers and pre-boomers. 25(12), 1393-1401.
5. Cai, DC., Chen, C.Y., Lo, T.Y., 2022, Foot Reflexology: Recent Research Trends and Prospects. Healthcare, 2022,11,9, 1-18.
6. Jirayingmongkol, P., Chantein, S., Phengchomjan, N., Bhanggananda, N. J. N., & Sciences, H. (2002). The effect of foot massage with biofeedback: A pilot study to enhance health promotion. 4(3), A4-A4.
7. Morrissey, A.-M., O'Neill, A., O'Sullivan, K., & Robinson, K. J. B. J. o. P. (2022). Complementary and alternative medicine use among older adults with musculoskeletal pain: findings from the European Social Survey (2014) special module on the social determinants of health. 16(1), 109-118.
8. Chang, M.-Y., Liu, C.-Y., & Chen, H.-Y. J. C. T. i. M. (2014). Changes in the use of complementary and alternative medicine in Taiwan: a comparison study of 2007 and 2011. 22(3), 489-499.

Authors Introduction

Prof. Dengchuan Cai

He is a professor in Department of Industrial design, National Yunlin University of Science and Technology, Taiwan. He received his Ph.D. degrees in management from Taiwan University of Science and Technology and M.S. degree in engineering from National Chiao Tung University, Taiwan.

Ms. Yuting Hsiao

She received her M.S. degree in design from National Yunlin University of Science and Technology. Currently she is a Ph.D. student in the School of Design of National Yunlin University of Science and Technology.

Dr. Chung-Wen HUNG

He received the Ph.D. degrees in Electrical Engineering from National Taiwan University in 2006. Currently he is a Professor in National Yunlin University of Science &Technology. His research interests include the IoT, IIoT, and AI application.

A Study on Maximum Power Point Tracking Technology for Solar Power Systems Using Power Variation to Adjust Step Response

Chun-Liang Liu, Chung-Wen Hung
Department of Electrical Engineering, National Yunlin University of Science and Technology
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.

Yi-Feng Luo
Graduate Institute of A.I. Cross-disciplinary Tech, National Taiwan University of Science and Technology

Guan-Jhu Chen
Department of Electrical Engineering, National Changhua University of Education

Cheng-Sin Hu
Department of Electrical Engineering, National Yunlin University of Science and Technology

E-mail: clliu@yuntech.edu.tw, wenhung@yuntech.edu.tw, YF.Luo@mail.ntust.edu.tw, gjchen@cc.ncue.edu.tw, M11212069@yuntech.edu.tw

Abstract

The Perturb and Observe (P&O) method is a widely used Maximum Power Point Tracking (MPPT) algorithm in photovoltaic power generation systems. However, it faces a trade-off between minimizing steady-state oscillation losses and accelerating the transient response. This study introduces an enhanced P&O method with an adjustable step size to address this issue. The method starts with a larger step size during the initial tracking phase, and when the operating point surpasses the Maximum Power Point (MPP), the step size is dynamically adjusted by multiplying it by a factor F, reducing the perturbation magnitude. The step size gradually decreases until it reaches the minimum value that the hardware or system can set up and execute. The proposed method retains the transient advantage of a larger perturbation step size while reducing power loss by minimizing steady-state oscillations. Compared to fixed perturbation step sizes and a fixed multiplier, the proposed method achieves faster perturbation convergence, maintaining a 9 step rapid tracking with a 4% duty cycle in transients and up to 99.98% in steady-state tracking.

Keywords: Maximum power point tracking, P&O method, variable step-size P&O method.

1. Introduction

In the next five years, the installed capacity of renewable energy will continue to increase, with solar photovoltaic systems and wind energy accounting for a record 96%. With continued policy support, it is expected that by 2028, the installed capacity of solar photovoltaic and wind energy will more than double compared to 2022. Throughout the forecast period, records will be continuously broken, with solar installed capacity reaching nearly 539.6 GW (main case). Therefore, how to effectively utilize solar energy has become an important issue. This study primarily explores the advantages and disadvantages of solar MPPT algorithms. Using the simulation software MATLAB, it simulates the P&O methods, including fixed step size[1], [2], [3] the variable step size [4], [5], [6], [7] and the algorithm proposed in this paper. The steady-state and transient responses are analyzed in detail. The P&O for adjusting the step response of power changes proposed in this paper alters the duty cycle disturbance of the controller based on the path of the operating point movement and the power change. The algorithm proposed in this paper is easy to implement and has high steady-state tracking accuracy and fast-tracking performance, achieving the goal of MPPT.

2. PV MODULE

In this study, The PV cell is a single-diode equivalent circuit model, which is described in Fig. 1, including the diode(D), parallel resistance (RP), and series resistance (Rs). The output current and voltage of a solar cell is expressed as equation (1).

$$I = I_{pv} - I_s\left\{\exp\left[\frac{q(R_s I + V)}{nKT_k}\right] - 1\right\} - \frac{R_s I + V}{R_p} \qquad (1)$$

When K is the Boltzmann’s constant; q is the charge of an electron (1.602×10-19 coulomb); n is the ideality factor; TK is the temperature in Kelvin; Ig, Is and I is photogenerated current, saturation current and panel current.

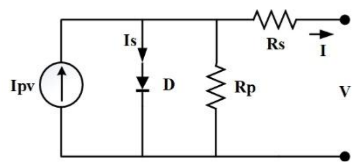


Fig 1. PV module equivalent circuit.

The simulation characteristic curves used in this study are based on the parameters listed in Table 1.

Table. 1 PV module parameters

Parameters	Symbol	Value
Maximum output power	P_{max}	234.24W
Open circuit voltage	V_{OC}	51.7 V
Maximum power voltage	V_{mp}	44.27 V
Short circuit current	I_{SC}	5.6 A
Maximum power current	I_{mp}	5.29 A

3. System Configuration

A DC-DC converter is a crucial element in PV systems. It provides an interface between the PV array and load. The PWM signal's duty cycle is adjusted based on the value determined by the MPPT algorithm for maximum power tracking. The relationship between the output voltage (V_o) and input voltage (V_{in}) of a boost converter is expressed as equation (2).

$$\frac{V_o}{V_{in}} = \frac{1}{1-\delta} \tag{2}$$

The duty cycle is denoted as δ. Consequently, the relationship between the output current and input current can be expressed as equation (3)

$$I_{in} = \frac{1}{(1-\delta)} I_o \tag{3}$$

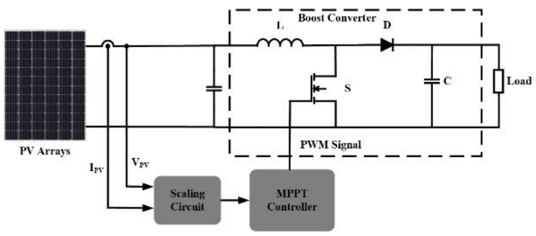


Fig. 2 System architecture diagram

To track the MPP of a solar cell, the output voltage and current signals of the solar cell are first captured. These feedback signals are then input into the controller for computation, which generates a new duty cycle to control the boost converter, thereby achieving MPPT. The system architecture is illustrated in Fig 2.

4. Variable Step Size and Proposed Algorithm

The algorithm introduced in this section is based on the principle of the P&O method. To address the limitations of the fixed-step P&O method, a variable-step P&O method is required for improvement. Many variable-step algorithms proposed in the literature involve numerous parameters, making them relatively complex. Based on the concept of variable steps and the P&O method, equation (4) represents the M-factor P&O method, and the proposed relationship is expressed as follows.

$$\Delta D(n) = M \times \Delta D(n\text{-}1) \tag{4}$$

Where $\Delta D(n)$ is the modified duty cycle.

$\Delta D(n\text{-}1)$ is the previous duty cycle.

M is a fixed proportional factor.

As seen from equation (4), the current duty cycle perturbation magnitude D is related to the previous duty cycle perturbation magnitude $D1$ and the M-factor. The relationship is straightforward.

To maintain the advantage of rapid response, the following conditions must be met before executing this relation (4) and (7), as shown in equations (5) and (6).

$$P(n-2) < P(n-1) \tag{5}$$

$$P(n-1) > P(n) \tag{6}$$

The parameters defined in the above two equations are as follows:

$P(n)$: Latest power value

$P(n-1)$: Previous power value.

$P(n-2)$: The power value from the two previous times.

The power variation adjustment step rule proposed in this paper is expressed in equation (7). $\Delta D(n)$ is the modified duty cycle. $\Delta D(n\text{-}1)$ is the previous duty cycle.

It utilizes a fixed duty cycle under steady-state variations to obtain the perturbation power change on the left side (ΔP_L) and the right side (ΔP_R). The maximum value of the power changes from both sides is taken as the denominator (ΔP_F) in equation (8). Here, ΔP represents the power difference between the latest power value and the previous power value.

$$\Delta D(n) = \frac{\Delta P(n)}{\Delta P_F} \times \Delta D(n\text{-}1) \tag{7}$$

$$\Delta P_F = \max(\Delta P_L, \Delta P_R) \tag{8}$$

equations (5) and (6) are the primary conditions for determining the use of the relationships in equations (4) and (7). In Fig 3, power tracking path 1 crosses the MPP, transitioning from the left half-plane of the P-V curve to the right half-plane, while path 2 follows the opposite direction. Both conditions involve crossing the MPP. If this phenomenon occurs, it indicates that the operating point is oscillating near the MPP. If the method used is the fixed step P&O rule to adjust the fixed duty cycle near the MPP, it is called the oscillation phenomenon, causing a loss in the steady state. Therefore, the duty cycle is multiplied by the M-factor from equation (4) to reduce its variation, thereby minimizing disturbances and improving steady-state tracking accuracy. This design can be applied to either the open-circuit voltage side or the short-circuit current side for tracking, achieving the same effect.

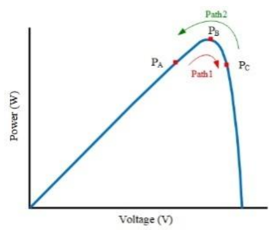


Fig. 3 Two paths passing through the MPP.

The flowchart is shown in Fig 4. First, the voltage and current values of the solar cell are measured, and the power value is calculated while continuously performing the P&O method. When the conditions in equations (5) and (6) are met, the fixed-proportion method uses equation (4) to update the duty cycle. Meanwhile, the proposed power variation adjustment step method utilizes equations (7) and (8) to calculate the latest duty cycle, continuing until the minimum preset perturbation magnitude is reached. From the above, it is known that the goal of achieving fast transient tracking to MPP and reducing oscillation at the MPP during steady state can be met. Therefore, this study conducts simulation experiments using a fixed duty cycle, the M fixed proportional factor, and the proposed power variation adjustment step response method.

5. Experimental Results

Using MATLAB to simulate the proposed method, variable step size with a fixed ratio M value of 0.9, and fixed duty cycles of 4% and 1% for P&O. Fig. 5(a) shows the waveform of these three methods for 1000 W/m^2 and 700 W/m^2 solar irradiances. Fig. 5(b) is the Duty cycle variation and stopping perturbation process. The performances of the tested methods are summarized in Table 2. The transient step is the number of steps required for the system to reach a steady state, and this value can be used to calculate its loss.

The proposed method requires 9 steps for transient rise at 1000 W/m², which is the same as the maximum fixed-step 4% P&O method in terms of time. Tracking accuracy is when the system reaches steady-state and the total recorded data contains 200 points for calculating the power average loss. Both the fixed-ratio perturbation and the proposed method ultimately stabilize at 0.1%, achieving the same accuracy of 99.98%. The average power losses are 8.55W, 12.87W, 3.32W, and 2.99W. Under solar irradiation of 700 W/m², the transient rise of the proposed method requires 7 steps, with a transient loss of 173.91 W. The tracking accuracy is 99.99%, and the average loss is 0.96 W. It can be seen that the proposed method controller can improve the transient and steady-state performance of the PV system simultaneously.

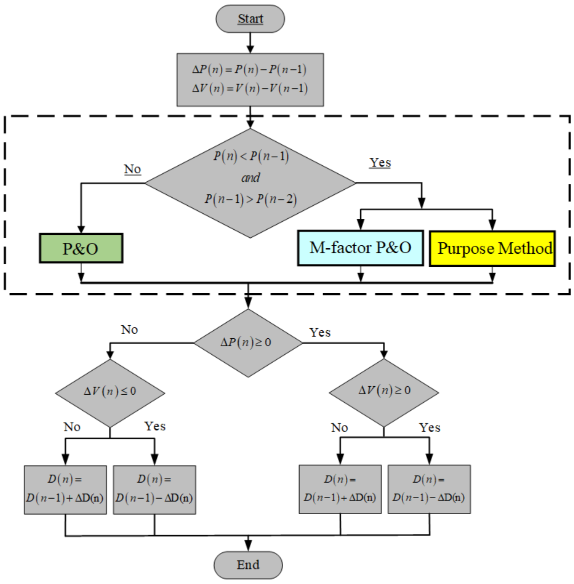


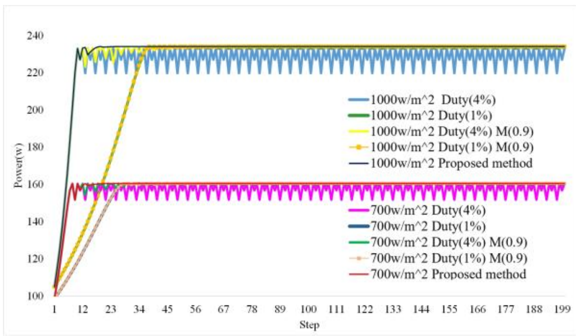
Fig 4 Flow chart of the three P&O methods

Table 2. Summarized performance of methods.

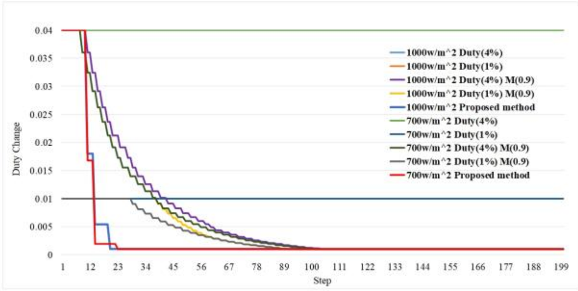
Methods	W/m2	P&O(4%)	P&O(1%)	M-factor (0.9)	Proposed
Transient Step	1000	9	36	9	9
	700	7	27	7	7
Transient Loss	1000	578.58W	2506W	578.58W	578.58W
	700	173.91W	785.48W	173.91W	173.91W
Tracking accuracy	1000	97.54%	99.82%	99.98%	99.98%
	700	98.07%	99.83%	99.99%	99.99%
Average loss	1000	8.55W	12.87W	3.32W	2.99W
	700	3.98W	4.16W	1.09W	0.96W

6. Conclusions

This study utilizes MATLAB software to simulate three methods of MPPT for solar systems. From the experimental results, it can be seen that the proposed method of adjusting the power change step P&O has high efficiency in steady-state response and fast transient response, effectively achieving MPPT for solar energy. The transient response maintains the rise time of the traditional 4% P&O, which takes 9 steps, while the steady-state tracking accuracy reaches as high as 99.98%. The average tracking power loss of 2.99W is the lowest in the test, thus the overall efficiency is high, which verifies its advantages and makes it easy to implement.



(a). PV source output performance.



(b). Duty cycle variation.

Fig. 5 Simulation results of three methods.

Acknowledgements

This work was supported by the National Science and Technology Council (NSTC), Taiwan [Project code NSTC 113-2222-E-224 -003 -].

References

1. R. I. Jabbar, S. Mekhilef, M. Mubin and K. K. Mohammed, "A Modified Perturb and Observe MPPT for a Fast and Accurate Tracking of MPP Under Varying Weather Conditions," IEEE Access, vol. 11, pp. 76166-76176, Jul. 2023.
2. V. K. P K and J. J J, "Implementation of Perturb and Observe MPPT Algorithm for the Performance Enhancement of Hybrid Power Generation Systems," International Conference on Computing Communication and Networking Technologies, pp. 1-5, 2023.
3. V. Kumar and M. Singh, "Derated Mode of Power Generation in PV System Using Modified Perturb and Observe MPPT Algorithm," Journal of Modern Power Systems and Clean Energy, vol. 9, no. 5, pp. 1183-1192, Sep. 2021.
4. S. R. Kiran, C. H. H. Basha, V. P. Singh, C. Dhanamjayulu, B. R. Prusty and B. Khan, "Reduced Simulative Performance Analysis of Variable Step Size ANN Based MPPT Techniques for Partially Shaded Solar PV Systems," IEEE Access, vol. 10, pp. 48875-48889, May. 2022.
5. A. Mohapatra, B. Nayak and C. Saiprakash, "Adaptive Perturb & Observe MPPT for PV System with Experimental Validation," IEEE International Conference on Sustainable Energy Technologies and Systems, pp. 257-261, 2019.
6. A. I. M. Ali, H. H. H. Mousa, H. R. A. Mohamed, S. Kamel, A. S. Hassan, Z. M. Alaas, E. E. M. Mohamed, A. R. Y. Abdallah, "An Enhanced P&O MPPT Algorithm With Concise Search Area for Grid-Tied PV Systems," IEEE Access, vol. 11, pp. 79408–79421, Jan. 2023.
7. E. M. Ahmed and M. Shoyama, "Scaling Factor Design Issues in Variable Step Size Incremental Resistance MPPT in PV Systems," IEEE Power Electronics and Drive Systems Conference, pp. 889-894, 2011.

Authors Introduction

Dr. Chun Liang Liu

He currently holds the position of Assistant Professor at the National Yunlin University of Science & Technology. His research interests include power electronics and battery management systems.

Prof. Chung-Wen Hung

He received the Ph.D. degrees in Electrical Engineering from National Taiwan University in 2006. Currently, he is a Professor in National Yunlin University of Science & Technology. His research interests include the IoT, IIoT, and AI applications.

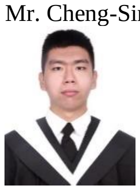
Dr. Yi Feng Luo

He currently holds the position of Assistant Professor at the National Taiwan University of Science and Technology, Taipei, Taiwan. His research interests include encompass charger algorithms, battery management systems, and renewable energy.

Dr .Guan Jhu Chen

He currently holds the position of Assistant Professor at the National Changhua University of Education, Changhua, Taiwan. His research interests include power electronics, digital power control and Li-ion battery charging algorithm.

Mr. Cheng-Sin Hu



He received her Bachelor's degree from the Department of Electrical Engineering, National United University, Taiwan in 2023. He is currently a master student in National Yunlin University of Science and Technology, Taiwan.

Bearing faulty prediction based on knowledge distillation

Chung-Wen HUNG, Zheng-Jie LIAO, Chun-Liang LIU
National Yunlin University of Science and Technology, Taiwan
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C

Email: wenhung@yuntech.edu.tw, M11212039@yuntech.edu.tw, clliu@yuntech.edu.tw

Abstract

This paper employs knowledge distillation to train teacher and student models using different motor bearing vibration datasets. The signal is transformed from the time domain to the frequency domain using Fast Fourier Transform (FFT), and a Convolutional Neural Network (CNN) model is used to recognize the bearing conditions. The teacher model is a deeper model trained with a larger dataset, while the student model is a shallower model trained with less data. The student model is guided by the soft labels provided by the teacher model. The results demonstrate that knowledge distillation improves the student model's recognition performance and enables knowledge transfer, allowing the student model to achieve good recognition accuracy even with limited training data.

Keywords: Bearing fault detection、Knowledge distillation、CNN

1. Introduction

Bearings are crucial to the operation of rotating motors, such as induction motors. Therefore, the condition of motor bearings significantly impacts the operational state of machinery 1. In recent years, advancements in artificial intelligence have facilitated the widespread application of deep learning, with techniques based on deep neural networks proven effective for diagnostics using vibration signals [2], [3]. However, existing fault diagnosis networks are large in scale and contain numerous parameters, making them difficult to deploy on embedded computing platforms. This study utilizes model compression techniques to reduce the model size while maintaining a certain level of accuracy, with the goal of enabling deployment on edge devices.

2. Materials and Methods

2.1. *Artificial Neural Networks*

Artificial Neural Networks (ANN) are the foundational structure of deep learning, comprising an Input Layer, Hidden Layers, and an Output Layer. Each hidden layer consists of a different number of neuron nodes, which connect with one another through each node's multiplication by various weights and addition of a bias term. The calculation is shown in Equation (1), where F represents the chosen activation function, y_n is the output of the current layer (and the input to the next layer), x_i represents the node of the current layer, M is the number of nodes in the previous layer, $w_{j,i}$ is the weight connecting the nodes, and $bias$ is the offset term. Before passing to the next layer, each neuron undergoes a linear or nonlinear transformation through an activation function. This transformation allows neural networks to learn and simulate data patterns and decision boundaries of varying complexity, enhancing the model's flexibility.

$$y_n(n) = \sum_{i=0}^{M} F * \left(x_i(n) \times w_{j,i}(n) + bias_n\right) \qquad \textbf{(1)}$$

2.2. *Convolution Neural Network*

Convolutional Neural Networks (CNNs) are the most widely used method for feature extraction. The structure of a one-dimensional CNN consists of a Convolution Layer, a Pooling Layer, and a Fully Connected Layer. The Fully Connected Layer, typically organized as an artificial neural network, performs classification after computing the weights. CNNs are a frequently used type of neural network in deep learning, particularly for tasks that require efficient feature extraction and classification.

2.3. *Knowledge Distillation*

Knowledge Distillation (KD) 4 is a model compression technique designed to train a simpler student model to replicate the behavior of a more complex teacher model, achieving comparable or even superior recognition performance. The principle of knowledge distillation is shown in Equation (1), where z_i represents the output logits for the i-th class, and a temperature variable T is introduced. A higher T results in a smoother probability distribution from the softmax output, increasing the entropy of the distribution, which helps the model focus more on negative labels. This can enhance the generalization ability of the student model. During high-temperature distillation, as shown in Equation (2), the objective function consists of a weighted combination of the distillation loss $\mathcal{L}_{soft}$ and the student loss $\mathcal{L}_{hard}$ is introduced to avoid fully trusting the teacher model, effectively reducing the propagation of errors to the student model. The weights α and β represent the respective proportions, with their sum equal to 1.

$$q_i = \frac{\exp\left(\frac{z_i}{T}\right)}{\sum_j \exp\left(\frac{z_j}{T}\right)} \quad (1)$$

$$\mathcal{L} = \alpha\mathcal{L}_{soft} + \beta\mathcal{L}_{hard} \quad (2)$$

3. Data and Results

3.1. *Case Western Reserve University dataset*

In this study, we utilized the Case Western Reserve University (CWRU) ball bearing vibration dataset [5], which is commonly employed for fault diagnosis in induction motor bearings. The vibration testing platform, shown in Fig. 1, consists of components arranged from left to right: motor, torque sensor, and dynamometer. Two accelerometers were installed on the motor's drive end and fan end, sampling at 48 kHz and 12 kHz, respectively. For this study, we selected the 48 kHz drive-end bearing data to evaluate the proposed classification method.

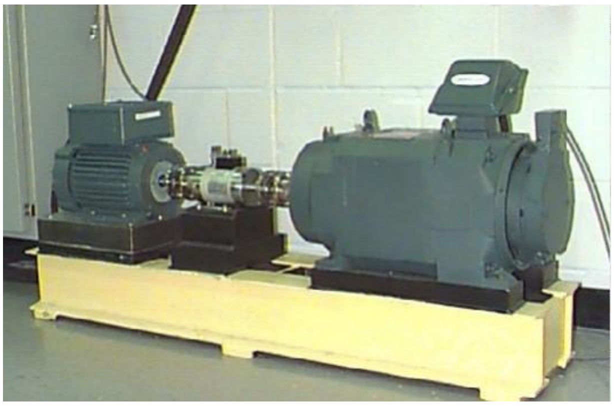


Fig. 1 CWRU platform

The drive-end bearing has four distinct conditions: inner race fault, outer race fault, ball fault, and normal condition. Fault diameters are categorized into three sizes: 0.007, 0.014, and 0.021 inches. The motor was tested under four different load levels: 0 hp, 1 hp, 2 hp, and 3 hp, which correspond to rotational speeds of 1797 rpm, 1772 rpm, 1750 rpm, and 1730 rpm, respectively. Table 1 presents the various fault conditions utilized in this study, along with the labels assigned to inner race faults, outer race faults, and normal conditions.

Table 1 CWRU fault conditions

Condition	Fault size(in.)	label
normal	0	0
Inner	0.007	1
Inner	0.014	2
Inner	0.021	3
Outer	0.007	4
Outer	0.014	5
Outer	0.021	6

3.2. *Triaxial bearing vibration dataset*

The triaxial bearing vibration dataset (TBVD) [6] contains three-dimensional (X, Y, and Z axis) vibration data for induction motor bearing faults. The testing platform, shown in Fig. 2, consists of a three-phase induction motor and an AC generator with a variable load. An accelerometer is mounted on the housing near the drive-end bearing of the three-phase induction motor, sampling at a rate of 10 kHz. The bearing conditions include inner race faults, outer race faults, and normal conditions, with fault severity levels of 0.7, 0.9, 1.1, 1.3, 1.5, and 1.7 mm. Vibration data were collected under load levels of 100 W, 200 W, and 300 W. Table 2 provides details of the various faults in the dataset and the labels assigned to the selected inner race faults, outer race faults, and normal conditions.

Fig. 2 TBVD platform

Table 2 TBVD fault conditions

Condition	Fault size (mm)	label
normal	0	0
Inner	0.7	1
Inner	1.3	2
Inner	1.7	3
Outer	0.7	4
Outer	1.3	5
Outer	1.7	6

In this study, the CWRU vibration dataset was used as training data for the teacher model. The input data for the model involved downsampling the 48 kHz vibration signal to 10 kHz using linear interpolation. The data was then segmented into segments of 2048 points each, followed by a Fast Fourier Transform (FFT). This process yielded a total of 1122 samples. The waveforms for each condition are shown in Fig. 3, demonstrating distinct frequency distributions under various fault conditions.

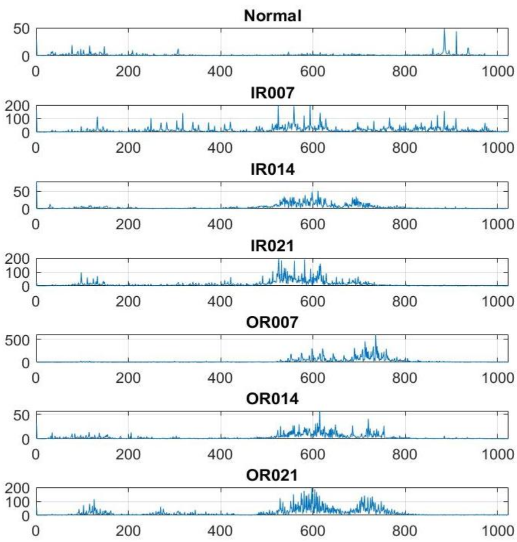


Fig. 3 CWRU data waveform

For output classification, faults of the same severity across three load levels were grouped into a single class, resulting in a total of seven conditions. Detailed model training parameters are listed in Table 3. The training and testing data were divided in a 7:3 ratio. Accuracy (ACC) was used as the evaluation metric, calculated as shown in Equation (1), where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. The confusion matrix for the teacher model on the test set is shown in Fig. 4, with an ACC of 0.997, which indicates excellent classification performance.

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \qquad (1)$$

Table 3 Teacher model parameter

CNN Filters	22
CNN Stride	8
CNN Kernel	10
CNN Activate Function	Swish
ANN Hidden layer	64-64-64-64-64-64
Hidden layer Activate Function	Sigmoid
Classes	7
Params	39,667
Learning Rate	0.0001
Epoch	200

True / Predicted	0	1	2	3	4	5	6
0	53	0	0	0	0	0	0
1	0	49	0	0	0	0	0
2	0	0	40	0	0	1	0
3	0	0	0	58	0	0	0
4	0	0	0	0	45	0	0
5	0	0	0	0	0	50	0
6	0	0	0	0	0	0	41

Fig. 4 confusion matrix

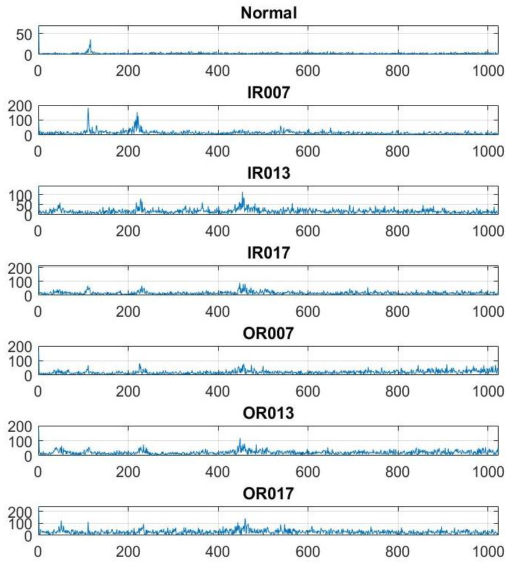


Fig. 5 TBVD data waveform

For the student model, the triaxial bearing vibration dataset was used. Since the teacher model only takes single-axis vibration data as input, it is essential to align the data dimensions between the teacher and student models to enable soft-label supervision from the teacher model. Therefore, the X-axis data from the triaxial dataset was selected as the input for the student model. The 10 kHz vibration data was segmented into 2048-point segments and transformed using FFT, resulting in 455 samples. The waveforms for each condition are shown in Fig5, revealing distinct frequency distributions under different fault states.

The data was divided into training and testing sets in an 8:2 ratio, and a 5-fold cross-validation method was applied to the dataset, with ACC averaged over the five folds. Detailed training parameters for the student model are shown in Table 4. The structure of the student model has fewer parameters compared to the teacher model.

Table 4 Student model parameter

CNN Filters	22
CNN Stride	8
CNN Kernel	10
CNN Activate Function	Swish
ANN Hidden layer	32-64-32-64-32-64
Hidden layer Activate Function	Sigmoid
Classes	7
Params	22,291
Learning Rate	0.0001
Epoch	1000

The training results are summarized in Table 5. When the student model was trained independently, it achieved an accuracy of 0.824. When the student model reused the CNN layers from the teacher model and trained only the subsequent ANN classification layers, accuracy increased to 0.839. By performing knowledge distillation with soft labels provided by the teacher model, the student model's recognition accuracy reached 0.892. This represents a 6.8% improvement over training of the student model independently and only a 2% difference from the performance of the teacher model. Moreover, the student model achieved comparable recognition performance while requiring only 56% of the parameters of the teacher model.

Table 5 model accuracy

Model	ACC
Teacher	0.997
Triaxial data with the teacher model architecture	0.912
Independent training of the student model	0.824
KD for the student model's ANN layers	0.839
KD student model	0.892

4. Conclusion

This study utilizes knowledge distillation to facilitate the student model's learning from the teacher model's expertise. Under similar operating conditions but different working states, experimental results show that knowledge distillation can effectively reduce the model size to 56% of the teacher model. Furthermore, it allows for knowledge transfer, resulting in only a 2% difference in accuracy between the student model and the teacher model.

5. References

1. Xie, J., et al.: A novel bearing fault classification method based on XGBoost: the fusion of Deep Learning-based features and empirical features. IEEE Trans. Instrum. Meas. 70, 1–9 (2021)
2. Yang, R., Zhang, Z., Chen, Y.: Analysis of vibration signals for a ball bearing-rotor system with raceway local defects and rotor eccentricity. Mech. Mach. Theor. 169, 104594 (2022).
3. Li, X., et al.: A vibration fault signal identification method via sest. Electronics 11(9), 1300 (2022).
4. Hinton, Geoffrey, Oriol Vinyals, and Jeff Dean. "Distilling the knowledge in a neural network." arXiv preprint arXiv:1503.02531 (2015).
5. Brjapon: 'CWRU Bearing Dataset'. https://www.kaggle.com/datasets/brjapon/cwru-bearing-datasets. (2023)
6. Kumar, D., et al.: Triaxial bearing vibration dataset of induction motor under varying load conditions. Data Brief 42, 108315 (2022).

Authors Introduction

Dr. Chung-Wen HUNG

He received the Ph.D. degrees in Electrical Engineering from National Taiwan University in 2006. Currently he is a Professor in National Yunlin University of Science &Technology. His research interests include the IoT, IIoT, and AI application.

Mr. Zheng-Jie LIAO

He received the B.S. degrees and now he is studying for the M.S. degree in electrical engineering from National Yunlin University of Science and Technology.

Dr. Chun-Liang LIU

He received the Ph.D. degrees in Electrical Engineering from National Taiwan University of Science and Technology in 2014. Currently he is a Professor in National Yunlin University of Science &Technology.

Accurate Brain Age Prediction Through Advanced Preprocessing and 3D ResNet-50 Modeling

Ting-An Chang
Department of Electrical Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan

Chiang-Ming Yeh
Department of Electrical Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan

Chun-Liang Liu
Department of Electrical Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan
E-mail: changta@yuntech.edu.tw, M11212035@yuntech.edu.tw, clliu@yuntech.edu.tw

Abstract

Accurate brain age prediction from structural magnetic resonance imaging (MRI) holds significant potential for advancing our understanding of the aging process and its effects on neural structures. In this paper, a robust preprocessing pipeline and two state-of-the-art 3D convolutional neural network architectures, 3D ResNet-50 and 3D DenseNet-121, were employed to develop and evaluate a brain age prediction model. The preprocessing steps included skull removal, spatial normalization to the Montreal Neurological Institute (MNI) template, and brain tissue segmentation into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). These steps ensured consistency and accuracy in the input data. The experimental results demonstrated that the 3D ResNet-50 architecture achieved superior performance, with a mean absolute error (MAE) of 3.9 for individuals over 50 years of age, surpassing the MAE of 4.1 achieved by the 3D DenseNet-121 model. These findings validate the efficacy of the proposed preprocessing pipeline and highlight the critical role of tailored deep learning architectures in brain age prediction. Future research could further enhance prediction accuracy by integrating multimodal imaging data and exploring hybrid model architectures.

Keywords: brain age prediction; magnetic resonance imaging; 3D ResNet; 3D DenseNet; preprocessing.

1. Introduction

The emergence of machine learning techniques has enabled the automatic prediction of diseases from medical imaging data [1], [2], [3]. With the recent advancements in deep learning, prediction accuracy has been elevated to levels that, in some scenarios, surpass human performance. These advancements have been recognized as valuable tools in assisting clinical diagnosis and treatment decision-making processes.

Structural magnetic resonance imaging (MRI) studies have demonstrated that profound neuroanatomical changes are experienced by the brain during normal development and aging. It has been observed that global gray matter volume decreases linearly with age, while regional gray matter is affected in a more heterogeneous manner. Significant reductions in gray matter volume have been identified in the frontal and parietal lobes, as well as in certain regions of the temporal lobe. However, non-linear patterns of age-related changes have been reported in specific subcortical regions, such as the caudate and hippocampus.

In addition, age-related alterations at the network level have been detected through structural MRI studies. More recently, efforts to estimate brain age using structural MRI data have been increasingly pursued. These efforts are regarded as essential for advancing the understanding of the relationship between brain morphometry and brain-predicted age, offering deeper insights into the aging process and its effects on neural structures.

2. Datasets and Feature Statistics

In this paper, T1-weighted magnetic resonance images (MRIs) were collected from individuals with normal cognitive function to construct a comprehensive and representative brain age prediction model. To ensure a diverse and inclusive sample of healthy brain structures spanning multiple age groups, T1-weighted images were meticulously selected from three widely recognized neuroimaging databases: the Alzheimer's Disease Neuroimaging Initiative (ADNI) [4], Information eXtraction from Images (IXI), and Cambridge Centre for Ageing and Neuroscience datasets (Cam-CAN). Each of these databases contributed data that were critical to the creation of a model that could account for the structural nuances of the brain across a broad spectrum of life stages.

A total of 769 T1-weighted images were obtained from the ADNI database, comprising cognitively normal elderly participants with a mean age of 72.46 years and a standard deviation of 6.89 years. These images were essential in providing insights into the normal aging process, allowing the study to explore structural changes in the brain that occur in older adults. Meanwhile, the IXI database contributed 275 T1-weighted images from healthy British adults, with an average age of 63.45 years and a standard deviation of 7.68 years. This set of data offered a crucial perspective on the structural characteristics of the middle-aged brain, bridging the gap between youthful development and the aging process. Additionally, the addition of the Cam-CAN dataset helps create a more continuous age range between these age groups. Cam-CAN provided T1-weighted images of 379 subjects with a mean age of 68.32 years and a standard deviation of 10.50 years. These scans proved invaluable for studying typical brain development during adolescence and early adulthood.

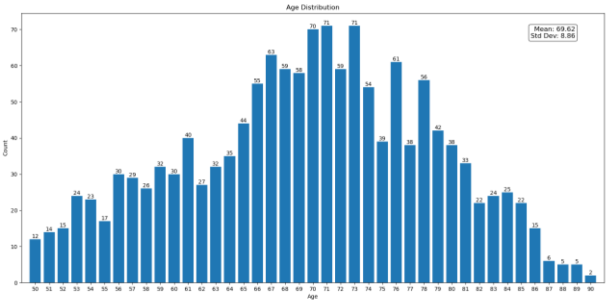


Figure 1. Age distribution of the collected datasets.

The combination of the ADNI, IXI, and Cam-CAN neuroimaging databases ensured that the sample focused exclusively on the older adult age range, capturing a comprehensive representation of late-life stages. This approach facilitated the inclusion of diverse brain structures across various life stages, significantly enhancing the robustness of the resulting brain age prediction model. In total, 1423 T1-weighted image datasets were sourced from these three databases as shown in Figure 1, providing a solid and diverse foundation for the modeling and analysis of brain development and aging trajectories.

3. Proposed System

3.1. *Data Preprocessing*

The initial step involved the removal of the skull from the original T1-weighted magnetic resonance images. This skull removal process was performed to eliminate non-brain tissue, which could interfere with the subsequent analysis, thereby increasing the accuracy and reliability of the study's results. After completing the skull removal, next use ANTs for rigid registration to align the images to the template mni_icbm152_t1_tal_nlin_sym_09a.nii. This template is a high-precision standard brain template designed for standardized processing in neuroimaging studies. This template originates from the Montreal Neurological Institute (MNI) and was developed by the International Consortium for Brain Mapping (ICBM). It was generated based on T1-weighted MRI data from 152 healthy adult volunteers aged 20 to 50 years. The template has a resolution of 1 mm and dimensions of 189 × 197 × 233, which helps ensure spatial consistency of brain images from different subjects. The use of rigid registration helps avoid unnecessary stretch deformation and aligns images more precisely, thereby improving the accuracy and stability of inter-individual comparisons.

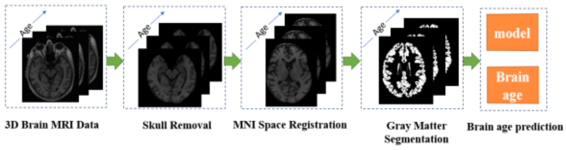


Figure 2. Proposed system flow chat.

Following the completion of spatial normalization, the Statistical Parametric Mapping version 12 (SPM12) software was employed to perform brain tissue segmentation. This step allowed the brain images to be precisely divided into three key tissue types: Gray Matter (GM), White Matter (WM), and Cerebrospinal Fluid (CSF). The segmentation process was carried out with high precision, enabling the extraction of distinct brain tissue components. This approach facilitated a more detailed and nuanced analysis, allowing for the independent investigation of different brain structures. By segmenting the brain into GM, WM, and CSF, more subtle neural structural information could be captured, providing deeper insights into the complexities of brain anatomy.

These Preprocessing techniques, including skull removal, spatial normalization, and tissue segmentation, laid the foundation for a more accurate and detailed analysis of brain structure. The input data were prepared in a way that maximized the reliability and precision of the brain age prediction model.

3.2. *Accurate Brain Age Prediction Algorithm Design*

Following the completion of Preprocessing and tissue segmentation of the brain images, we will focus on using gray matter (GM) images to train the 3D ResNet-50 model. The proposed system flow chat as shown in Figure 2. The reason for choosing the ResNet-50 architecture is that its effectiveness in processing high-dimensional medical imaging data has been widely proven. Its densely connected design not only facilitates the efficient transfer and reuse of features, but also effectively learns rich feature representations and captures complex spatial features inside the brain, thereby enhancing the performance capabilities of the model.

In the model training stage, the input gray matter image is extracted by 3D ResNet-50 [5], and then further processed through a fully connected layer, and finally a floating-point number is generated, representing the model's predicted value of the subject's brain age. The loss function of the model is the Mean Absolute Error (MAE). This method can intuitively measure the accuracy of the model prediction by calculating the mean absolute difference between the predicted age and the actual age.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (1)$$

Where N is the number of samples, y_i is the actual age, $\hat{y}_i$ is the model predicted age.

During the model training process, the Adam optimizer is used for training. The final learning rate is 0.001, the weight attenuation is 1e-6, and the epochs are 500. In terms of hardware configuration, the model is trained on the NVIDIA A100 GPU, ensuring high performance and efficiency of the operation.

4. Experimental results

To validate the accuracy of the proposed preprocessing method and model for predicting brain age, two 3D convolutional neural network architectures, 3D DenseNet-121 [6] and 3D ResNet-50, were selected for experimentation and comparative analysis. The selection of 3D ResNet-50 was informed by its residual learning capabilities, which have been demonstrated to effectively capture early age-related features of brain structure and enable the accurate learning of subtle, age-specific structural changes. The 3D DenseNet-121 architecture was chosen for its dense connection design, which facilitates the efficient transfer and reuse of features, providing notable advantages in detecting fine variations in brain structure.

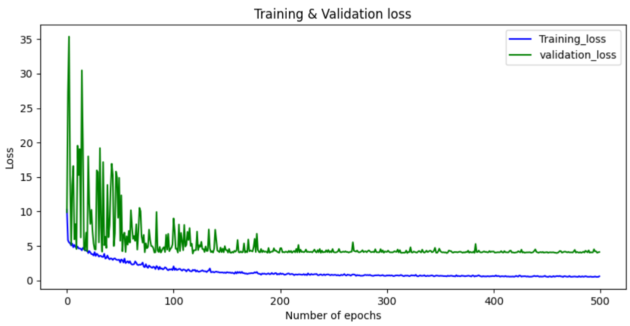


Figure 3. Mean absolute error performance of 3D ResNet-50

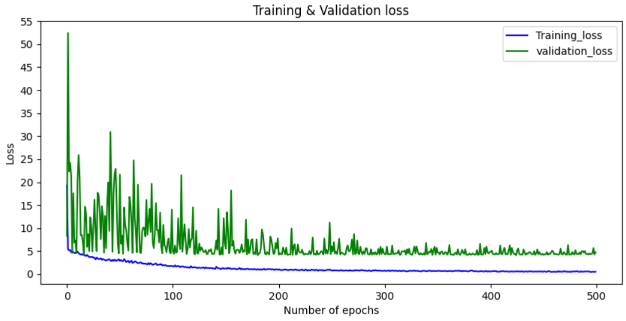


Figure 4. Mean absolute error performance of 3D DenseNet-121

The accuracy of the Preprocessing method and the 3D ResNet-50 architecture in predicting brain age was assessed through a series of experiments. For individuals over 50 years of age, it was observed that 3D ResNet-50 achieved a mean absolute error (MAE) of 3.9, indicating excellent predictive performance as shown in Figure 3.

To obtain a more comprehensive evaluation of the model's capabilities, a comparative analysis was conducted between 3D ResNet-50 and 3D DenseNet-121, another commonly used 3D convolutional neural network. For the same age group, 3D DenseNet-121 was found to achieve an MAE of 4.1 as shown in Figure 4. These experimental results demonstrated that 3D ResNet-50 significantly outperformed 3D DenseNet-121, exhibiting superior stability and accuracy in its predictions for individuals over 50 years of age.

5. Conclusions

In this paper, a comprehensive approach was developed and validated for brain age prediction using advanced preprocessing methods and two state-of-the-art 3D convolutional neural network architectures: 3D ResNet-50 and 3D DenseNet-121. The experimental results demonstrated the superior performance of 3D ResNet-50, particularly in predicting brain age for individuals over 50 years of age. The MAE of 3.9 was achieved by 3D ResNet-50, surpassing the MAE of 4.1 obtained by 3D DenseNet-121. These findings underscore the effectiveness of residual learning in capturing early and subtle age-related changes in brain structure, contributing to more stable and accurate predictions. This work provides a foundation for further research in brain age prediction, with potential applications in understanding the relationship between brain morphometry and biological aging. Future studies may explore integrating multimodal imaging data or employing hybrid deep learning models to enhance prediction accuracy and extend these findings to clinical and neurological applications.

Acknowledgment

The study was funded by National Science and Technology Council of Taiwan and National Yunlin University of Science and Technology, grant numbers: NSC 113-2221-E-224-042-.

References

1. Peng, H., Gong, W., Beckmann, C. F., Vedaldi, A., & Smith, S. M. (2021). Accurate brain age prediction with lightweight deep neural networks. *Medical image analysis*, *68*, 101871.
2. Baecker, L., Dafflon, J., Da Costa, P. F., Garcia-Dias, R., Vieira, S., Scarpazza, C., ... & Pinaya, W. H. (2021). Brain age prediction: A comparison between machine learning models using region-and voxel-based morphometric data. *Human brain mapping*, *42*(8), 2332-2346.
3. Sun, J., Tu, Z., Meng, D., Gong, Y., Zhang, M., & Xu, J. (2022). Interpretation for individual brain age prediction based on gray matter volume. *Brain Sciences*, *12*(11), 1517.
4. Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., ... & Weiner, M. W. (2010). Alzheimer's disease Neuroimaging Initiative (ADNI) clinical characterization. Neurology, 74(3), 201-209.
5. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
6. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

Authors Introduction

Dr. Ting-An Chang

He is an Assistant Professor of Electrical Engineering at the National Yunlin University of Science and Technology in Taiwan. He received his PhD degree in the Institute of Computer and Communication Engineering from National Cheng Kung University in 2018. His research interests smart healthcare and computer vision.

Mr. Chiang-Ming Yeh

He received his B.S. degree in Engineering in 2023 from Ming Chi University of Technology in Taiwan. He is acquiring an M.E. in National Yunlin University of Science and Technology.

Dr. Chun-Liang Liu

He currently holds the position of Assistant Professor at the National Yunlin University of Science & Technology. His research interests include power electronics and battery management systems.

Intelligent agricultural landscape identification system

Ching-Ju Chen
Department of Electrical Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan

Yu-Cheng Chen
Department of Electrical Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan

Jing-Yao Lin
Department of Electrical Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan

Rung-Tsung Chen
Taiwan Biodiversity Research Institute, Nantou, Taiwan

Candera Wijaya
Agricultural Engineering Research Center, Taoyuan, Taiwan

E-mail: cjchen@yuntech.edu.tw, M11212080@yuntech.edu.tw, M11212086@yuntech.edu.tw, tsung2358@gmail.com, rbhung@aerc.org.tw

Abstract

In recent years, Taiwan's economic development and land restructuring have decreased agricultural land area. The decrease in agricultural land area will affect the food supply for human beings and reduce the habitat of wild animals, destroying biodiversity. With the development of drone technology, the survey of farmland ecology no longer requires a large workforce to visit the site for inspection and analysis. This paper proposes a recognition system based on the Semantic segmentation method to classify agricultural landscapes, watersheds, and habitats automatically. We use two models for training: U-Net with the VGG16 model and U-Net with the Resnet50 model. The experiments in this paper show that the U-Net with VGG16 model and the U-Net with Resnet50 model applied to semantic segmentation of farmland landscape images have their classification categories. Still, some categories may be misclassified due to the similarity of the features, such as grassland, fallow field, and upland fields. This paper suggests that in the future, the number of data sets and the diversity of samples.

Keywords: Semantic Segmentation, Artificial Intelligence, Image Recognition, Biodiversity, Deep Learning

1. Introduction

Agricultural land serves not only as a source of human food but also as a habitat for wildlife and a provider of ecological services [1]. Reducing, overusing, or mismanaging agricultural land disrupts biodiversity. Agricultural biodiversity, a critical component of conservation, was integrated into the United Nations' Sustainable Development Goals (SDGs) in 2015.

This study proposes using UAVs to capture aerial images and establish an identification system for agricultural landscapes, watersheds, and habitats. This approach enables real-time biodiversity monitoring and supports adequate wildlife habitat protection and management, aligning with and advancing the SDGs for biodiversity conservation.

Recent advancements in artificial intelligence (AI) and deep learning have expanded the use of image recognition technologies. This study utilizes VGGNet, ResNet, and the U-Net semantic segmentation algorithm, as described below:

1. VGGNet

   The Visual Geometry Group [2] proposed in 2014 that VGGNet simplifies network design using 3x3 convolutional kernels and includes versions like VGG16 and VGG19, denoting the number of convolutional and fully connected layers. VGGNet demonstrated strong performance in the ImageNet competition.

2. ResNet

   Developed by Kaiming He et al. [3] in 2015, ResNet introduces Residual Blocks and Skip Connections to enable deeper networks while addressing model degradation. Popular versions include ResNet50 and ResNet101, indicating the number of layers.

3. Image Segmentation

   Image segmentation assigns pixels to specific categories, making it suitable for complex farmland landscapes like upland fields, freshwater fields, and grasslands. This study employs the U-Net model, introduced by Ronneberger et al. [4] in 2015, featuring an encoder-decoder architecture with skip connections to retain spatial details. Initially designed for medical imaging, U-Net has been adapted for various tasks, including aerial image segmentation, with variations like Attention U-Net. [5]

2. Research methods

In this study, different artificial intelligence algorithms based on semantic segmentation are collected, among which the model of U-Net architecture, whose structure is suitable for dealing with complex scenes and high-resolution aerial images, is also applied to aerial photos. The relevant data of the study also proves the applicability of the U-Net architecture to aerial photos. Therefore, this study proposes to build a U-Net architecture-based drone aerial image recognition system for farmland scenes.

2.1. *Input image*

In this study, we used 10cm/pixel high-resolution aerial photos provided by the Agricultural Engineering Research Center (AERC), and selected samples suitable for the identification of farmland landscapes from the provided images, as shown in Fig.1 and Fig.2 below, which should cover diverse farmland landscapes such as farmland and water.

Fig.1. Map Number 94202015_20200605

Fig.2. Map Number 94202004_20200712

2.2. *Classification category definition*

Because of the complexity and diversity of Taiwan's agricultural landscapes, a total of 21 categories were classified based on the actual context of the agrarian landscapes in the images, which were Freshwater field, Upland Field, Fallow fields, Land grading, Mixed forest, Grasslands, Bare land, Building, Orchard, (Ecological pond), Water pool, Fishery farm, Road, Farm road, Solar panel, Duck farm, Irrigation and drainage system, Unused land, Agricultural production facilities, No data.

2.3. *Image Labeling*

In this study, we use the open-source labeling software Labelme to mark 80 images with the size of 2048x2048, and the file (*.json) generated after labeling is processed to create the corresponding mask.

2.4. *Model training*

This study employs two backbone networks, U-Net+VGG16 and U-Net+ResNet50, using 800 images (560 for training, 120 for validation, and 120 for testing). The models were trained with a batch size of 1, 200 iterations, a learning rate of 0.00005, and Cosine Annealing for learning rate decay. Mixed precision training was applied due to GPU memory limitations, with Adam optimizer selected for better performance over SGD. The focal loss function was used to address class imbalance.

The training was conducted on a Z790 Pro CPU, a Gigabyte RTX4080 16GB GPU, and 128GB of RAM, using PyCharm (version 2023.1) and Python (version 3.11).

2.5. *Model Evaluation Metrics*

In this study, we use standard evaluation metrics of semantic segmentation, including mIoU, mPA, mRecall, and mPrecision, as an essential reference to measure the performance of the model, which are introduced as follows

(1) mIoU(mean Intersection over Union)

Calculate the intersection and concurrency ratio between the predicted and real regions and then average the IoU for all categories. The calculation formula is Eq.(1).

$$mIoU = \frac{1}{N}\sum_{i=1}^{N}\frac{TPi}{TPi + FPi + FNi} \tag{1}$$

(2) mPA(mean Pixel Accuracy)

Calculate the ratio of the number of pixels correctly predicted by the model to the total number of pixels, and then take the average value of Pixel Accuracy for all categories. The calculation formula is Eq.(2).

$$mPA = \frac{1}{N}\sum_{i=1}^{N}\frac{TPi}{TPi+FNi} \tag{2}$$

(3) mRecall

Calculate the proportion of True Positive pixels correctly recognized by the model, averaged over all classes of Recall, the calculation formula is as Eq.(3).

$$mRecall = \frac{1}{N}\sum_{i=1}^{N}\frac{TPi}{TPi + FNi} \tag{3}$$

(4) mPrecision

Calculate the proportion of pixels predicted by the model to be Positive that is Positive, then average the Precision values across all categories; the calculation formula is as Eq.(4).

$$mPrecision = \frac{1}{N}\sum_{i=1}^{N}\frac{TPi}{TPi+FPi} \tag{4}$$

2.6. *Model Prediction*

When the model is trained, the weights will be accessed and a new image will be imported. In the prediction, the image size needs to be the same as the training image size, which needs to be cut into several 2048x2048 images.

Then the prediction will be done for each 2048x2048 image, and the final image will be synthesized with the same size as the original imported image.

3. Experimental results

3.1. *Training results*

We calculate various evaluation metrics based on the test set's and training set's prediction results to assess the model's performance, serving as an important basis for subsequent parameter adjustments. As shown in Table. 1, U-Net+VGG16 achieves a mIoU of 70.94, mPA of 79.69%, mRecall of 79.69%, and mPrecision of 81.09%. In comparison, U-Net+ResNet50 achieves an mIoU of 57.71, mPA of 68.21%, mRecall of 68.21%, and mPrecision of 73.29%. This indicates that U-Net+VGG16 outperforms U-Net+ResNet50 in all evaluation metrics. However, the mPA, mRecall, and mPrecision metrics of U-Net+ResNet50 demonstrate its potential in agricultural land scene recognition.

Table. 1. Training results

	VGG16	Resnet50
mIoU	70.94	57.71
mPA	79.69	68.21
mRecall	79.69	68.21
mPrecision	81.09	73.29

3.2. *Prediction results*

In terms of prediction, we would like to compare the prediction results of U-Net+VGG16 and U-Net+Resnet50 to further verify the performance of the two models in the same scenario and to visualize the prediction effect of the two models as well as their ability to handle boundaries, etc. We would like to understand further the advantages and disadvantages of the two models and their applicability. To further understand the strengths and weaknesses of the two models and their applicability, the following is an example of a localized map of the predictions in image number 94202006c.

In Fig. 3, both the U-Net+VGG16 and U-Net+ResNet50 methods demonstrate strong performance in accurately identifying solar panels, indicating the robustness of these models in detecting this specific category. However, both methods need help classifying unused land, often mislabeling it as bare land, farm roads, fishery farms, or other similar categories.

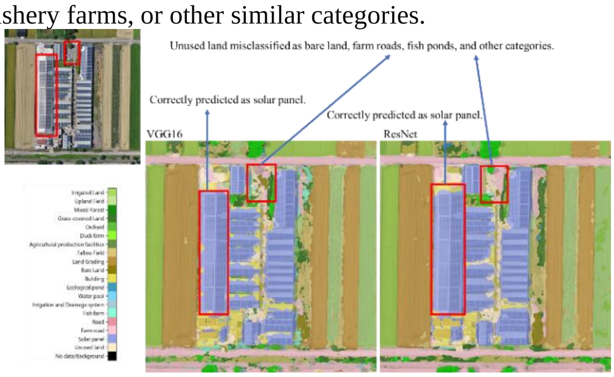


Fig. 3. Comparison of the partial map with the partial forecast map for map number 94202006c

3.3. *Calculate area*

In semantic segmentation, if the image resolution is known, then the number of pixels in each category can be used to calculate the size of the area covered by each category with the following formula: Eq. (5).

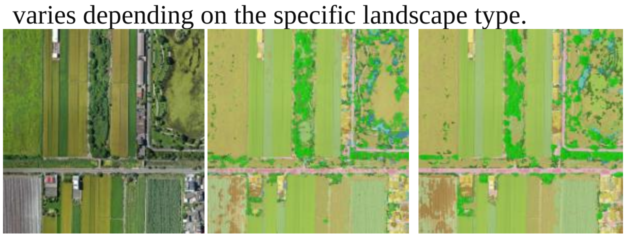
$$Area = \frac{Pixel\ count \times Area\ per\ pixel}{100{,}000{,}000} \quad (5)$$

Model prediction accuracy directly impacts area calculations, with incorrect predictions leading to inaccuracies. To evaluate this, we calculate the error between the predicted areas of U-Net+VGG16 and U-Net+ResNet50 and the actual area using Eq. (6), measuring the accuracy of both models.

$$\text{Error}(\%) = \frac{\text{Predicted area} - \text{Actual area}}{\text{Actual area}} \times 100\% \quad (6)$$

Fig. 4 illustrates the ground image and the prediction results of two models: U-Net+VGG16 (Fig. 4 a) and U-Net+ResNet50 (Fig. 4 b).

Table.2 compares the error values between the actual and predicted areas for various land types. The results show that U-Net+VGG16 provides more accurate predictions for freshwater fields, upland fields, land grading, mixed forests, bare lands, ecological ponds, roads, farm roads, irrigation systems, and unused lands. In contrast, U-Net+ResNet50 achieves greater accuracy for fallow fields, grasslands, buildings, orchards, fish farms, and agricultural production facilities. This indicates that the prediction accuracy of each model varies depending on the specific landscape type.



Actual image (a)U-Net+VGG16 (b)U-Net+Resnet50

Fig. 4. Actual image and Prediction result

Table.2.Comparison of actual area and predicted area

	Actual Area	U-Net +VGG16	Error	U-Net +Resnet50	Error
Freshwater field	3.58	3.68	2.79%	4.17	16.48%
Upland Field	2.58	2.53	1.94%	2.3	10.85%
Fallow field	3.65	5.17	41.64%	5.06	38.63%
Land grading	0	0.49		0.68	
Mixed forest	1.08	1.18	9.26%	1.47	36.11%
Grasslands	1.27	0.6	52.76%	0.64	49.61%
Bare land	0.04	0.03	25%	0.01	75%
Building	0.77	0.51	33.77%	0.57	25.97%
Orchard	0.06	0.76	1166.67%	0.4	566.67%
Ecological pond	1.70	0.25	85.29%	0.13	92.35%
Water pool	0	0	0%	0	0%
Fishery farm	0	0.21		0.12	
Road	0.49	0.13	73.47%	0.08	83.67%
Farm road	0.20	0.69	245%	0.73	265%
Solar panel	0	0	0%	0	0%

Duck farm	0	0	0%	0	0%
Irrigation and drainage system	0.66	0.13	80.3%	0.07	89.39%
Unused land	0.58	0.29	50%	0.02	96.55%
No data	0	0	0%	0	0%
Agricultural production facilities	0	0.01		0	0%

4. Conclusion

This study utilized U-Net+VGG16 and U-Net+ResNet50 models to develop an AI system for farmland landscape recognition. Five hundred sixty images were used for training, 120 for validation, and 120 for testing. U-Net+VGG16 achieved better performance with mIoU of 70.94%, mPA of 79.69%, mRecall of 79.69%, and mPrecision of 81.09%, compared to U-Net+ResNet50, which had mIoU of 57.71%, mPA of 68.21%, mRecall of 68.21%, and mPrecision of 73.29%.

In predictions, U-Net+VGG16 performed better in recognizing freshwater fields, upland fields, land grading, mixed forests, bare lands, ponds, roads, irrigation systems, and unused lands. Meanwhile, U-Net+ResNet50 identified fallow fields, grasslands, buildings, orchards, and agricultural facilities. Both models have strengths suitable for different segmentation tasks, but misjudgments due to similar landscape features remain challenging. Improvements in combining categories and increasing the diversity of image samples could improve the overall recognition accuracy of agricultural landscape segmentation.

Acknowledgment

This research was supported by the National Science and Technology Council, Taiwan (NSTC 113-2635-M-224-001-). Special thanks to the Agricultural Engineering Research Center (AERC) for providing aerial images and the Taiwan Biodiversity Research Institute for their expertise and resources, which greatly enriched the study.

References

1. Power Alison G., “Ecosystem services and agriculture: tradeoffs and synergies Phil.” Trans. R. Soc. B3652959–2971,2010, doi: http://doi.org/10.1098/rstb.2010.0143
2. Karen Simonyan, Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” 2014.,arXiv:1409.1556.
3. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition,” arXiv:1512.03385,2015.
4. Olaf Ronneberger, Philipp Fischer, Thomas Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” 2015, arXiv:1505.04597.
5. N. Gonthina, S. Adunuri, R. Mateti, S. S. Allampalli and L. V. Narasimha Prasad, “Accurate Semantic Segmentation of Aerial Imagery Using Attention Res U-Net Architecture,” 2024 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2024, pp. 1-5, doi: 10.1109/ESCI59607.2024.10497287.

Authors Introduction

Ching-Ju Chen

She received a Ph.D. in engineering science from National Cheng Kung University, Taiwan, in 2011. She is an Assistant Professor at the Department of Electrical Engineering, National Yunlin University, Taiwan. Her expertise includes digital image processing, AI, AIoT, UAV system integration, GIS, GPS, and RS.

Yu-Cheng Chen

He is pursuing a master's in Electrical Engineering at Yunlin National University of Science and Technology, Yunlin, Taiwan. His research interests include deep learning applications in aerial imaging and image recognition.

Jing-Yao Lin

He is pursuing a master's in Electrical Engineering at Yunlin National University of Science and Technology, Yunlin, Taiwan. His research interests include deep learning applications, GIS, and image processing.

Rung-Tsung Chen

He received a Ph.D. in fisheries science from National Taiwan University, Taiwan, in 2009. He is an Assistant Research Fellow with the Wild Animals Division at Taiwan Biodiversity Research Institute. He majored in the studies of organic agriculture's aquatic ecology and biodiversity.

Candera Wijaya

He earned a master's in civil engineering from National Central University. He focused on spatial information at the Agricultural Engineering Research Center, assisting the Ministry of Agriculture in promoting the use of drones in agriculture, forestry, fisheries, and animal husbandry.

Leveraging AIoT Visual Analytics for Optimizing Agricultural Sustainability and Efficiency

Hsueh-Yen Shih
Agricultural Engineering Research Center, Taoyuan, Taiwan

Xi-Wei Lin
Department of Electrical Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan

Zhao-Sheng Chen
Department of Electrical Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan

Ying-Cheng Chen
Tainan District Agricultural Research and Extension Station, Tainan, Taiwan

Ching-Ju Chen
Department of Electrical Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan

E-mail: isacyen@aerc.org.tw, M11212073@yuntech.edu.tw, M11312076@yuntech.edu.tw, larvaephage@gmail.com, cjchen@yuntech.edu.tw

Abstract

Bemisia tabaci is an essential pest in global agriculture, and its reproduction and distribution are affected by meteorological conditions, causing severe damage to economic crops. This study constructed a micro-meteorological station and a visualized weather alert system combining narrow-band Internet of Things (NB-IoT) and message queuing telemetry transmission (MQTT) technologies in an orchard to achieve precise pest monitoring and early warning. The micro-weather station can record high-resolution meteorological data (e.g., temperature, humidity, soil moisture, sunlight intensity, etc.) and transmit them stably and efficiently via NB-IoT and MQTT technologies to ensure real-time data updates over long distances. The early warning system combines artificial intelligence (AI) and geographic information system (GIS) to analyze the impact of meteorological conditions on *Bemisia tabaci*, make risk predictions, and provide prevention and control recommendations, such as adjusting pesticide application strategies to reduce losses and pollution. This study demonstrates an innovative agriculture solution based on NB-IoT and MQTT, which realizes data-driven and intelligent pest management in orchards, improves monitoring accuracy and control efficiency, and is not only applicable to whitefly control but can also be extended to other pests and diseases, providing an efficient and sustainable technology example for agricultural pest management.

Keywords: Bemisia tabaci formosana, Micro Weather Stations, Narrowband Internet of Things (NBIoT), Ecological Protection, Internet of Things (IoT)

1. Introduction

Bemisia tabaci (*B. tabaci*) secretes honeydew that coats the surface of the crop, providing a substrate for the growth of *B. tabaci*. This results in a black mold covering the surface of the crop, which affects photosynthesis and reduces the value of the commodity. *B. tabaci* is incredibly reproducible, forming large populations quickly under favorable conditions, and is resistant to various chemical pesticides, making prevention and control more difficult, increasing the use of pesticides and further aggravating the burden on the environment and agroecosystems. Global climate change, the increase in greenhouse environments, and the intensification of farming have all contributed to the reproduction and spread of *B. tabaci*, exacerbating the threat it poses to crops.

B. tabaci infestations devastate the yield and quality of economic crops, resulting in reduced incomes for farmers, lower market value of crops, and destabilization of agricultural ecosystems. Therefore, prevention of *B. tabaci* has become a substantial focus of agricultural research, including biological control (e.g., the use of natural enemies), improvement of agro-ecological environments, optimization of crop structures, and the development of *B. tabaci*-resistant varieties to minimize the damage caused by *B. tabaci* to crops, to safeguard food security, and to promote the sustainable development of the agricultural industry.

The reproduction and distribution of *B. tabaci* are influenced by meteorological factors such as temperature, humidity, soil moisture, and sunlight, which are crucial in regulating its population dynamics and impact. A micro-meteorological station was established in an orchard to collect high-resolution data on temperature, humidity, soil moisture, and sunlight to monitor these factors. This allows for modeling the relationship between *B. tabaci* populations and weather conditions and understanding how climate change affects population fluctuations.

The Visual Weather Alert website utilizes this data for real-time analysis and risk prediction, providing orchard managers with early warnings and recommendations. These include optimizing pesticide application timing, enhancing pest prevention measures, and improving crop management strategies to minimize damage caused by *B. tabaci*. The system integrates historical data and current meteorological conditions to offer targeted and actionable insights.

By combining precise meteorological data with ecological models, this approach fills gaps in traditional pest control strategies, quantifying the impact of climate conditions on *B. tabaci*. The system also supports smart agriculture by improving management efficiency, reducing pesticide use, and promoting sustainable agricultural practices. Beyond managing *B. tabaci*, this study serves as a model for digital and intelligent pest control applicable to other pests and diseases.

2. Research Methods

The system adopts solar power technology to realize a self-sufficient power supply and reduce the dependence on electricity for outdoor work. In addition, to effectively collect and transmit data, we compared the advantages and disadvantages of two transmission modes, the LoRa communication module and NB-IoT (Narrow Band Internet of Things), to ensure data transmission's stability and low power consumption. The advantage of this method is that it cannot only collect environmental data on *B. tabaci* growing environment stably for an extended period but also minimize the disturbance to the environment caused by manual observation.

2.1 Hardware Architecture Design

In the early stage of the study, we designed the first version of the device for environmental monitoring (shown in Fig. 1), which uses a LoRa communication module to transmit the monitoring data to the server. Due to the lack of stable network connection and power support in some monitoring sites, the operation of the first version of the device was limited by the network and power supply, which could not satisfy the long-term monitoring needs. To solve this problem, we further optimized and developed the second version of the device (shown in Fig. 2). This version introduces NB-IoT (Narrow-Band Internet of Things) technology, enabling the device to operate independently on a low-power cellular network without relying on external infrastructure. The device is also equipped with a solar power module to realize a self-sufficient power supply, which prolongs the device's operating life and reduces manual maintenance. We have upgraded the environmental sensors to improve the equipment's environmental adaptability and durability. The original light and temperature/humidity sensors have been replaced with a waterproof design to ensure stable operation in harsh environments such as high humidity and rain. This improvement improves the device's reliability and expands its range of applications, providing higher-quality monitoring data for subsequent data analysis.

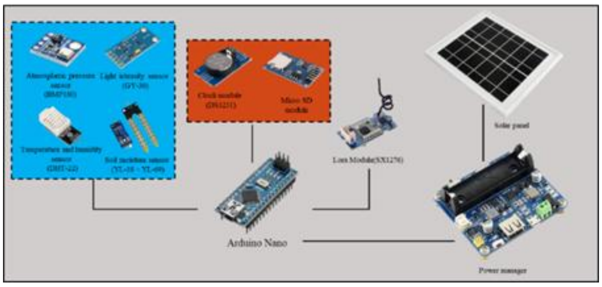


Fig. 1 First version of the device(LORA)

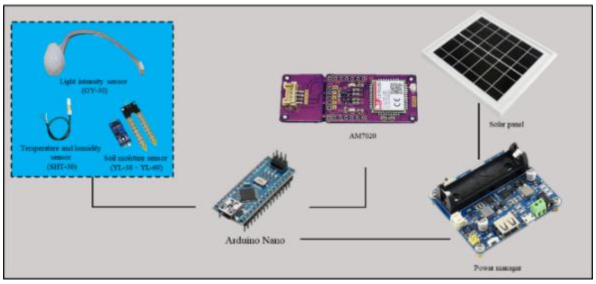


Fig. 2 The second version of the device (NB-IOT)

2.2 NB-IOT

Narrowband Internet of Things (NB-IoT) is a low-power wide-area (LPWA) technology designed for large-scale IoT deployments, offering efficient and environmentally friendly connectivity. As part of the LTE network [1], NB-IoT is a leading solution for machine-to-machine (M2M) communication due to its low power consumption, wide coverage, and cost-effectiveness, making it ideal for IoT applications.

NB-IoT's low energy consumption also makes it an environmentally friendly choice for IoT, contributing to innovative city development. It enables efficient device connectivity for energy management, environmental monitoring, and intelligent transportation, supporting the ITU-T's vision of sustainable smart cities [2]. These cities leverage ICT and innovation to improve quality of life, urban efficiency, and competitiveness while addressing current and future generations' economic, social, environmental, and cultural needs. NB-IoT serves as a vital foundation for achieving these goals.

2.3Communication Protocol

MQTT is a lightweight publish/subscribe communication protocol widely used in IoT and messaging applications, with implementations like Mosquitto (lightweight open-source broker), Hive-MQ (enterprise-grade broker), and Paho MQTT (open-source client library) [3]. Its core feature is the MQTT agent, which manages client communication and distributes messages based on subscribed topics, enhancing data transmission efficiency and reliability [4].

Experimental results (Table 1) show that MQTT significantly reduces power consumption and bandwidth requirements compared to HTTP and HTTPS while ensuring stable and efficient data transmission. This makes it ideal for large-scale IoT applications, especially in resource-constrained environments like cellular networks.

Table 1. Communication Costs.

	MQTT	HTTP GET	HTTP POST
data traffic	0.244K	1.12k	1.337K
30MB transmissions	125,523 times	27,422 times	22,971 times
Transmitted every minute	Can use 87 days	Can use 19 days	Can use 16 days
Transmitted every five minutes	Can use 435 days	Can use 95 days	Can use 80 days
Cost per 100 times	0.079 NT	0.364 NT	0.435 NT

2.4Automation Tools

Node-RED is an open-source Internet of Things (IoT) development tool based on Node.js, which is widely used for real-time processing and automated sensor data management [5]. With its flow-based programming visualization interface Fig. 3), developers can quickly build data collection, processing, and transfer applications without writing large amounts of code.

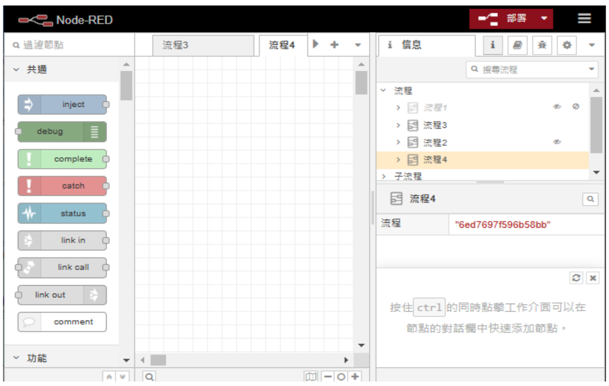


Fig. 3. Programming Visualization Interface

In this study, an automated data processing system is integrated using Node-RED (Fig. 4), which subscribes to the sensor data via the MQTT protocol and forwards the processed information to an SQL database for storage via HTTP. The system integrates LINE communication to push critical alert messages to users in real-time, realizing fully automated data management and notification [6], which enables users to track farm conditions in real-time from any location, improving management efficiency and decision-making accuracy [7].

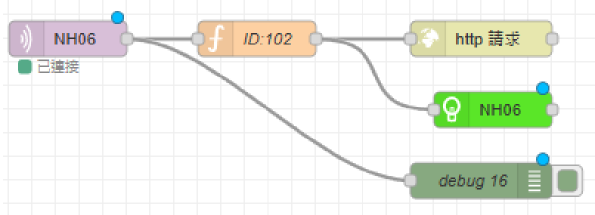
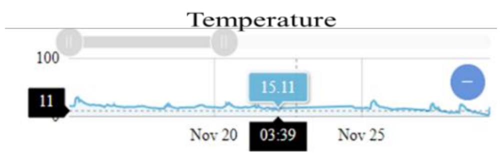
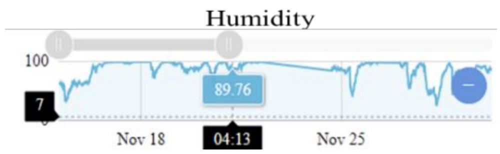


Fig. 4. Automated Data Processing System

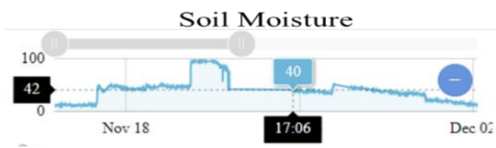
This study adopts HTML5 as the primary development framework, combining CSS and jQuery technologies to realize visual design and animation effects. CSS is used for page beautification to ensure a flexible and attractive interface, while jQuery handles dynamic effects and interactive functions to enhance user experience. The collected data is processed by a visualization tool and presented in graphs and charts, including temperature, humidity, soil moisture, and light intensity (Fig. 5). In addition, the system supports users to adjust the viewing time and select specific data points to facilitate flexible viewing of historical data and detailed analysis.



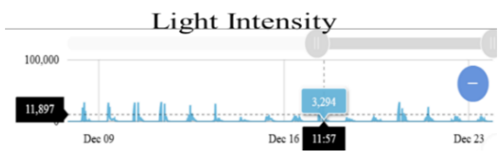
a. Temperature data curve



b. Humidity data curve



c. Soil moisture data curve



d. Light Intensity data curve

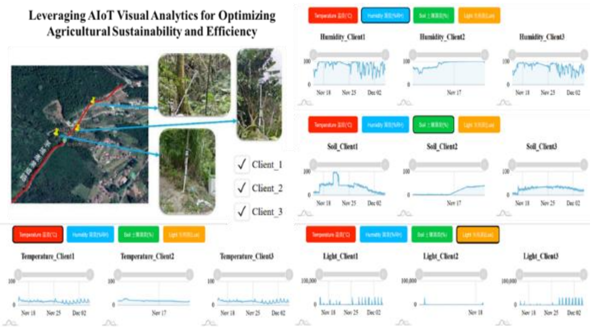


Fig. 5. Web Visualization Interface

3. Conclusions

This study successfully implemented a micrometeorological station and a visual weather alert system integrating NB-IoT and MQTT technologies to monitor whitefly infestations in orchards with high accuracy and risk prediction. The micrometeorological station captures precise weather changes, while NB-IoT ensures stable long-distance data transmission even in areas with limited infrastructure. MQTT enhances data transmission efficiency and supports simultaneous multi-device operations, enabling large-scale applications.

By combining AI models and GIS, the system analyzes weather impacts on whitefly populations, predicts risks, and provides targeted prevention strategies, such as optimizing pesticide use and crop management. This reduces crop damage, pesticide use, and environmental pollution while improving pest control efficiency and accuracy. The findings confirm that NB-IoT and MQTT-based solutions enable intelligent, data-driven pest management and are scalable for other pests and diseases. Future improvements aim to optimize system performance and broaden applications, supporting sustainable agricultural development with innovative solutions.

Acknowledgment

This research was supported by the National Science and Technology Council, Taiwan (NSTC 113-2635-M-224-001-), with support from the Tainan District Agricultural Research and Extension Station, and the Agricultural Engineering Research Center (AERC), which provided the experimental site, laboratory assistance, and data collection. Their expert guidance was crucial to the study's success.

References

1. S. K. Routray, "Narrowband Internet of Things", Encylopedia of Organizational Knowledge, no. 61, pp. 1-11, Jun. 2020.
2. S. K. Routray, D. Gopal, A. Javali and A. Sahoo, "Narrowband IoT (NBIoT) Assisted Smart Grids," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 1454-1458
3. International Business Machines Corporation (IBM), & Eurotech. (2010). MQTT V3.1 protocol specification. IBM.
4. Durkop Lars, Bjorn Czybik and Jurgen Jasperneite, "Performance evaluation of M2M nrotocols over cellular networks in a lab environment", Intelligence in Next Generation
5. Diego Clerissi, Maurizio Leotta, Gianna Reggio, and Filippo Ricca. 2018. Towards an approach for developing and testing Node-RED IoT systems. In Proceedings of the 1st ACM SIGSOFT International Workshop on Ensemble-Based Software Engineering (EnSEmble 2018). Association for Computing Machinery, New York, NY, USA, 1–8.
6. S. Chanthakit and C. Rattanapoka, "MQTT Based Air Quality Monitoring System using Node MCU and Node-RED," 2018 Seventh ICT International Student Project Conference (ICT-ISPC), Nakhonpathom, Thailand, 2018, pp. 1-5
7. V. David, H. Ragu, R. K. Duraiswamy and S. P, "IoT based Automated Indoor Agriculture System Using Node-RED and IBM Bluemix," 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2021, pp. 157-162

Authors Introduction

Hsueh-Yen Shih

He received a master's degree in civil engineering from the National Central University, Taiwan 2009. He is an assistant researcher at the Agricultural Engineering Research Center, Taoyuan, Taiwan, his expertise includes digital photogrammetry, geometrical processing for remotely sensed images, GIS, RS, and AI of crop interpretation.

Xi-Wei Lin

He received his B.S. in Engineering 2023 from the National Yunlin University of Science and Technology in Taiwan. He is acquiring an M.E. in National Yunlin University of Science and Technology.

Zhao-Sheng Chen

He received his B.S. in Engineering 2024 from Southern Taiwan University of Science and Technology in Taiwan. He is acquiring an M.E. in National Yunlin University of Science and Technology.

Ying-Cheng Chen

He earned a master's in entomology from National Taiwan University in 2009 and is an Assistant Researcher at the Tainan District Agricultural Research and Extension Station. His research focuses on pest control strategies for fruit trees, including pesticide screening, pest population dynamics, and behavior analysis.

Ching-Ju Chen

She received a Ph.D. in engineering science from National Cheng Kung University, Taiwan, in 2011. She is an Assistant Professor at the Department of Electrical Engineering, National Yunlin University of Science and Technology, Taiwan. Her expertise includes digital image processing, AI, AIoT, UAV system integration, UAV telemetry image processing, GIS, GPS, and RS.

A Study on Surface Defect Detection Algorithm of Strip Steel Based on YOLOv8n

Haozhe Sun [1*], Fengzhi Dai [1], Junjin Chen [2]

[1] *Tianjin University of Science and Technology, Tianjin, China*

[2] *SMC (Beijing) Manufacturing Co., LTD., Beijing, China*

*E-mail: * 2921407938@qq.com*

www.tust.edu.cn

Abstract

Hot rolled steel strip has been extensively applied in industrial production and processing due to its outstanding properties. Nevertheless, during the production procedure, as a result of technological constraints, defects will inevitably occur on the surface of the steel strip, significantly influencing the performance and safety of the steel strip. Hence, how to detect the surface defects of steel strips has turned into the key point. In this paper, an enhanced YOLOv8n network model is proposed to make it applicable for the surface defect detection tasks of hot rolled steel strips. The improved model introduces Dynamic Snake Convolution and Efficient Multi-Scale Attention Module. The average precision of the improved model is 6.8 percentage points higher than that of the original model.

Keywords: Deep learning, Surface defect detection, YOLOv8n, Dynamic Snake Convolution, Efficient Multi-Scale Attention Module

1. Introduction

Steel production technology serves as a crucial indicator for gauging the development level of a country's iron and steel industry. Among various types of steel, strip steel is extensively utilized in the automotive, construction, and aviation sectors because of its high dimensional accuracy, excellent surface quality, ease of processing, material-saving properties, and other advantages, and the demands for its product quality are also higher. Nevertheless, due to the influence of multiple factors such as raw materials, the rolling process, and system control, the surface of hot rolled steel strips frequently presents pitting, scratches, inclusions, rolling oxide scale, cracks, and other defects. These defects will, to varying degrees, impact the main performance indicators of the steel plate, such as wear resistance, fatigue resistance, corrosion resistance, and electromagnetic characteristics. Therefore, how to detect the surface defects of steel strips and analyze their causes to enhance product quality has become the key. The commonly used detection methods are mainly divided into three categories: manual detection, theoretical mechanism detection and machine vision detection [1].

Before 1970, strip surface defect detection mainly relied on manual visual observation. The advantage of this approach lies in its lower cost. However, due to the large scale of metal products production, surface defects are often small, especially prone to missed detection, false detection, and the speed is slow. The physical mechanism detection methods commonly used are: infrared detection, eddy current detection and magnetic leakage detection methods. The accuracy of such methods is generally greatly affected by the environment. With the development of science and technology, machine vision technology has been proposed and applied to defect detection tasks. Machine vision detection can be divided into traditional image processing detection methods and deep learning detection methods. The main principle of traditional image processing detection method is to identify defects by feature extraction of the preprocessed image. These methods rely on human-designed features. Moreover, the types of defects are complex, the differences between samples of different categories are large, and the production environment is complex, which leads to the poor adaptability and generalization ability of the method in the actual production link, and the stability of the detection results is not high. Visual inspection methods based on deep learning can adaptively learn defect features, thereby accurately and rapidly locating defects. The performance of the model tends to enhance as the number of data samples grows. Even when the amount of data is insufficient, image enhancement can also be employed to expand the sample. The trained model possesses strong universality, high robustness and fast detection speed, and is applicable to defect detection issues in industrial production. In this paper, the original model of YOLOv8n is used to improve it to enhance the ability to capture the feature information of small targets. After improving the model, the experiment is carried out, and the results are analyzed.

The rest of this paper is organized as follows. The second section introduces the structure of YOLOv8n network model. In the third part, the structure and principle of Dynamic Snake Convolution and Efficient Multi-Scale Attention Module are introduced. The fourth section gives the process of the experiment to verify the effectiveness of the improved deep learning model. The fifth part summarizes the main contents of this paper.

2. Structure and Working Principle of YOLOv8n

YOLO (You Only Look Once) [2] series object detection algorithm is a single-stage object detection algorithm. With the continuous change of YOLO model version, its speed and accuracy have been significantly improved. YOLOv8 is a widely used and newer version. In order to adapt to different application scenarios and hardware resources, it is divided into YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l and YOLOv8x versions. These versions mainly improve the detection performance by increasing the depth and width of the network. The main body of YOLOv8 network is divided into three modules: Backbone network, Neck network and Head network. Among them, the backbone network was responsible for feature extraction and passed to the neck network. The neck network fused and enhanced the extracted features. The head network uses the feature information derived from the previous network to predict the image. The YOLOv8 network structure is shown in Fig. 1.

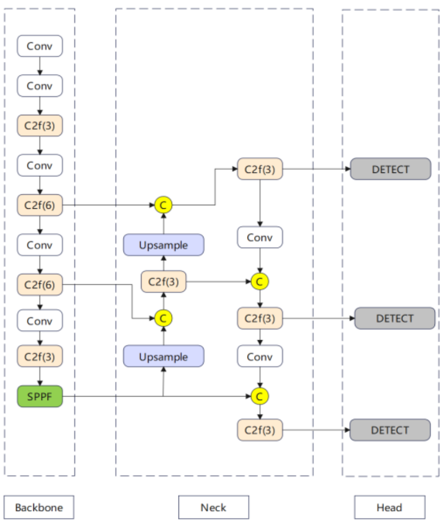


Fig. 1 Schematic diagram of YOLOv8 network structure

The backbone network of YOLOv8 adopts the CSPDarkNet [3] structure, which is constituted by convolutional modules CBS and C2f (CSPLayer With 2Conv), as well as Spatial Pyramid Pooling Fast (SPPF). CBS is a composite convolution module and serves as the most fundamental part of the entire backbone network. It is primarily composed of a two-dimensional convolution layer Conv, a two-dimensional batch normalization layer BatchNorm, and an activation function SiLU. In the C2f module, the input feature map initially traverses a convolution module on the trunk for feature extraction, and subsequently through n bottleneck modules to extract higher-level abstract feature information. These bottleneck modules employ the Split method to achieve cross-layer connection, while adding more parallel gradient flow branches. Thus, the expressiveness and performance of the entire network model are enhanced, and the ability of the model to capture complex features is improved without sacrificing computational efficiency, which contributes to enhancing the accuracy of the model in target detection tasks. The SPPF fast spatial pyramid pool module is situated at the last layer of the backbone network, enabling the feature map to acquire a more abundant receptive field, thereby strengthening the expression capacity of the network, as well as accelerating the operation speed and enhancing the efficiency of target detection. The neck network of YOLOv8 uses the structure of PAFPN, which is improved by the fusion of FPN and PAN. Its main purpose is to solve the problem of insufficient feature pyramid network in multi-scale detection tasks in object detection. The head network of YOLOv8 uses the Decoupled-Head structure, which uses two parallel branches to extract category features and location features respectively, and then uses a layer of CBS convolution and two-dimensional convolution to complete the classification and localization tasks.

3. Improve YOLOv8n Steel Surface Defect Detection Algorithm

3.1 Dynamic snake convolution

Due to the presence of minor flaws like fine cracks and rolled oxide scale in the NEU-DET [4] dataset, along with the insignificant difference between them and the background gray value, the detection accuracy is relatively low. To enhance the detection accuracy of the model for such minute defects, this paper adopts dynamic snake convolution to replace the original C2f module in YOLOv8n. Dynamic Snake Convolution (DSC) [5] is a technique originated from clinical medicine. The advent of this technology significantly contributes to the feature extraction work of cardiovascular targets such as slender, tortuous, and irregular ones. In the surface defects of steel strips investigated in this paper, the defects like fine cracks with poor detection results share similar structures, thereby this module is introduced to improve the detection accuracy.

Given the standard 2D convolution coordinates designated as K, the central coordinate is $Ki = (xi, yi)$. A 3×3 kernel K with a dilation of 1 is expressed as:

$$K = \{(x-1, y-1), (x-1, y), \cdots, (x+1, y+1)\} \quad (1)$$

To enable the convolution kernel to better focus on defect targets with complex geometric features, an offset Δ is introduced. Generally, the fixed convolution layer applies the same convolution operation to the feature maps of different levels, that is, the position of the sampled pixels is fixed. For instance, the convolution kernels of all CBS modules in the original YOLOv8n model are 3×3, and the feature information sampled by this approach will incorporate many background features. By adding a direction parameter to each convolution element, the deformable convolution kernel with a base of 3×3 can adaptively modify its shape during the training process to better match the characteristics of the input data. The characteristics of the deformable convolution kernel are presented in Fig. 2. The convolution kernel will extend from the green point area to the surrounding to change its shape, forming a new convolution kernel dominated by blue points. The middle arrow part is the introduced offset Δ.

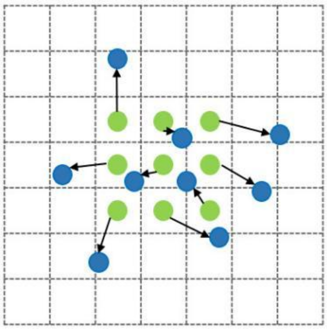


Fig. 2 Schematic diagram of deformable convolution

In DSC, the standard convolution kernel is elongated along the X-axis and Y-axis respectively. As presented in Fig. 3, the size of the convolution kernel subsequent to elongation is set at 9×9. Taking the X-axis direction as an exemplar, the specific location of each grid in K is expressed as $K_{i\pm c}=(x_{i\pm c},y_{i\pm c})$, where $c=\{0,1,2,3,4\}$ indicates the horizontal distance from the central grid. The determination of each grid position $K_{i\pm c}$ within the convolution kernel K is a cumulative procedure. Commencing from the central position K_i, the position deviating from the central grid is contingent upon the position of the preceding grid: K_{i+1} has an offset Δ relative to K_i, and$\Delta=\{\delta|\delta\epsilon[-1,1]\}$.

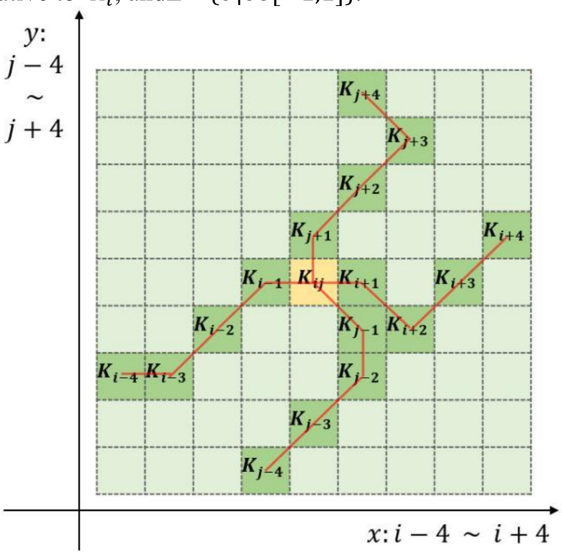


Fig. 3 Stretched convolution kernel

To guarantee that the receptive field does not deviate excessively from the target feature while the model adaptively acquires the defect feature offset, an iterative tactic is implemented to successively select the position that can be chosen at the next offset for each target to be processed, thereby ensuring the continuity of the convolution kernel's attention and preventing the receptive field from shifting too far due to the deformation of the convolution kernel. As depicted in Fig. 4, taking the X-axis direction as an example, the next offset position K_{i-1} for K_i is identified first, and its location is determined before proceeding to find the position of K_{i-2}.

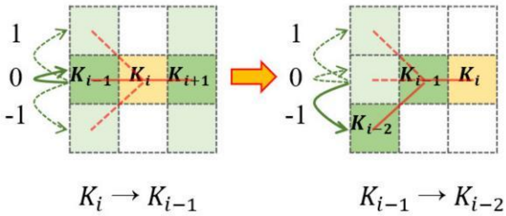


Fig. 4 Schematic diagram of K_i offset sequentially in the x-axis direction

The convolution kernel becomes Eq. (2) in the X-axis direction:

$$K_{i\pm c}=\begin{cases}(x_{i+c},y_{i+c})=(x_i+c,y_i+\sum_i^{i+c}\Delta y)\\(x_{i-c},y_{i-c})=(x_i-c,y_i+\sum_{i-c}^{i}\Delta y)\end{cases}\qquad(2)$$

The convolution kernel becomes Eq. (3) in the Y-axis direction:

$$K_{j\pm c}=\begin{cases}\left(x_{j+c},y_{j+c}\right)=(x_j+\sum_j^{j+c}\Delta x,y_j+c)\\\left(x_{j-c},y_{j-c}\right)=(x_j+\sum_{j-c}^{j}\Delta x,y_j-c)\end{cases}\qquad(3)$$

DSConv is designed to better adapt to the slender tubular structure based on the dynamic structures so as to better perceive the key features.

3.2 Efficient multi-scale attention module

Efficient Multi-Scale Attention(EMA) [6] is a novel attention mechanism specifically designed for computer vision tasks that aims to reduce computational overhead while preserving key information in each channel. The EMA module reconstructs part of the channels into batch dimensions and groups the channel dimensions into multiple sub-features so that the spatial semantic features are evenly distributed within each feature group. In addition to encoding the global information to recalibe the channel weights in each parallel branch, the output features of the two parallel branches are further aggregated by cross-dimensional interaction to capture pixel-level pairwise relationships.

A schematic of the Efficient Multi-Scale Attention (EMA) module is shown in Fig. 5. "g" denotes the number of groups into which the input channels are divided. X Avg Pool" and "Y Avg Pool" represent the one-dimensional horizontal and vertical global pooling operations, respectively. In the EMA module, the input is first grouped and then processed through different branches: one branch for one-dimensional global pooling and the other for feature extraction via a 3x3 convolution. The output features of the two branches are then modulated by a sigmoid function and a normalization operation, and finally merged by a cross-dimensional interaction module to capture the pixel-wise pairwise relationship. After the final sigmoid conditioning, the feature maps are output to enhance or weaken the original input features to obtain the final output.

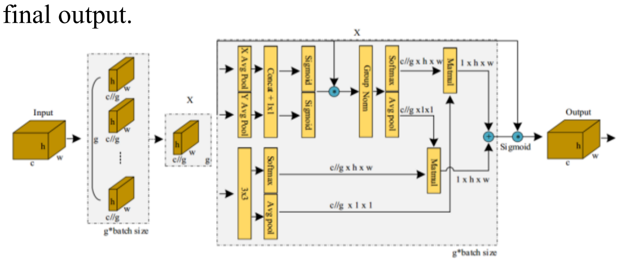


Fig. 5 Schematic diagram of EMA network structure

4. Experimental Setups

4.1 Processing of the dataset

In order to detect the common defects on the surface of steel strips, this paper uses the public data set for training. The data set is NEU-DET, the hot rolled steel strip surface defect detection data set organized by Professor Song Kecheng's team in Northeastern University. The dataset includes six defect types: crazing, inclusion, patches, pitted_surface, rolled-in scale, and scratches. There are 300 images of each defect, with a total of 1800 grayscale images. The original resolution of each image is 200×200 pixels.

The original data set of steel strip surface defects is in VOC data format, namely, the label file is in xml format. To facilitate the training of YOLO, a one-stage object detection model with high detection accuracy and fast detection speed for small objects, this paper converts the label file of the original data set into the txt label file used by YOLO through writing Python script code.

The original dataset only contains 1800 grayscale images, and there are few samples of each type of defect. In the actual training of deep learning network, it may cause overfitting phenomenon, resulting in that the model cannot fully learn the features of each type of defect, resulting in the reduction of defect detection accuracy. In order to solve the problem of sparse data sets, this paper expands the original steel strip defect data set, and uses image flipping, image rotation, adjusting image brightness and contrast to enhance the image, and expands the steel strip surface defect data set to 5400 images according to the ratio of 1:3.

After the expansion of the dataset, the corresponding label of each image also needs to be changed. Since the label content of the dataset is the defect category number and the normalized rectangular box information, the original label information is set as Eq. (4) here:

$$f = (i, x, y, w, h) \tag{4}$$

Where i is the defect category number, (x, y) is the center point coordinate of the defect rectangular box, and (w, h) is the width and height of the normalized defect rectangular box. Then the label information after horizontal flipping is shown in Eq. (5):

$$f' = (i, 1 - x, y, w, h) \tag{5}$$

The label information after vertical flipping is shown in Eq (6):

$$f'' = (i, x, 1 - y, w, h) \tag{6}$$

The 5400 images in the augmented dataset were divided into training set, validation set and test set according to the ratio of 8:1:1. That is, 4320 images in the training set, 540 images in the validation set, and 540 images in the test set.

4.2 Experimental environment and parameter setting

The network was built using the PyTorch deep learning framework on Windows 10 system. The processor used was the 10th Gen Intel(R) Core(TM) i9-10900K running at 3.70GHz. The graphics card employed was the NVIDIA GeForce RTX 2080Ti. The Python version utilized was 3.11, the PyTorch version utilized was 2.0.1, with a CUDA version of 11.8.

The training process consisted of 300 epochs, with the optimizer set to SGD. The initial learning rate was 0.01, workers was 4, and the batch size was 32. Set the model accuracy no longer rising within 50 rounds to trigger the early stop mechanism.

4.3 Evaluation index

mAP, Params, FLOPs were used as evaluation indicators.

Precision, The accuracy of the reaction model can be calculated by:

$$Precision = TP/TP + FP \tag{7}$$

where TP indicates that the object to be detected is correctly detected, and FP indicates that the non-object to be detected is correctly detected.

Recall, Represents the proportion of samples that are correctly predicted to belong to a class by the model among all samples that actually belong to a class, and is calculated as follows:

$$Recall = TP/TP + FN \tag{8}$$

where FN indicates that the object to be detected is incorrectly detected.

mAP@0.5, Represents the average precision value under different Intersection over Union (IoU) thresholds, calculated as:

$$mAP = \frac{1}{n}\sum_{i=0}^{n} AP(i) \tag{9}$$

where c is the total number of categories of images, i is the number of detections, and AP is the average recognition precision of a single category. This metric provides a comprehensive evaluation of the algorithm's performance. mAP@0.5 represents the average precision case with an IoU threshold of 0.5, and mAP@0.5:0.95 represents the average precision with an IoU threshold ranging from 0.5 to 0.95.

Params, is the sum of trainable parameters in a model, usually including the weights of convolutional layers, the weights of fully connected layers, and the bias term, and is calculated as follows:

$$Params = C_o \times (k_w \times k_h \times C_i + 1) \quad (10)$$

FLOPs is the number of floating-point operations required for the model to perform one forward propagation, which represents the computational burden of the model in a single inference process, and can be used to measure the complexity of the algorithm, calculated as:

$$FLOPs = 2 \times (C_i \times k_w \times k_h - 1) \times H \times W \times C_o \quad (11)$$

In the formula, H represents the height of the feature map, W represents the width of the feature map, and if the convolution kernel has a bias term, the $+1$ term in the parentheses is removed.

In this paper, mAP@0.5, Params and FLOPs are used as the main evaluation metrics of model detection accuracy.

4.4 Experimental result

Through the analysis of DSC and EMA modules mentioned above, all C2f modules in the original structure of YOLOv8 are replaced with DSC modules, and EMA attention mechanism is added to the 16th, 20th and 24th layers of the original model. The improved model structure diagram is shown in . The experimental results of the original YOLOv8n model and the experimental results of the improved YOLOv8_DSC_EMA model are compared to verify the influence of the improved network using DSConv and EMA modules on the surface defect detection results of steel strips, as shown in Table 1.

It can be seen from Table 1 that the average precision of YOLOv8_DSC_EMA model reaches 80.3%, which is 6.8 percentage points higher than that of YOLOv8n model. This shows that by replacing the deep convolution module of the original YOLOv8n with dynamic snake-like convolution, the extraction ability of feature information in the deep network is improved, especially for the detection accuracy of small target defects with tubular structure such as small cracks. In terms of the number of parameters, since the calculation process of snake convolution is more complex than that of C2f module, the number of parameters will increase and the amount of calculation will also increase slightly. Therefore, EMA is added to lightweight the model while retaining the key information of each channel. It is used for recalibration of channel weights, which enhances the ability of feature representation. Through comparison, it is found that the time used for image preprocessing and inference is little different from the original YOLOv8n network, but mAP@0.5 is greatly improved.

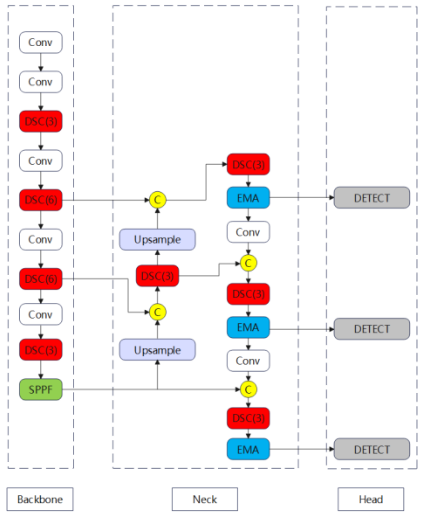


Fig. 6 Schematic diagram of YOLOv8_DSC_EMA network structure

Table 1 Comparison of model experiments for improving convolutional modules

Model	mAP @0.5	Params /M	FLOPs /G
YOLOv8n	0.735	3.01	8.1
YOLOv8_ DSC_EMA	0.803	3.91	9.9

The Precision-Recall curves obtained by training the original YOLOv8n model and YOLOv8_DSC_EMA model are shown in Fig. 7 and Fig. 8, respectively. By comparing the two groups of curves, it is found that the accuracy of the improved model has increased significantly.

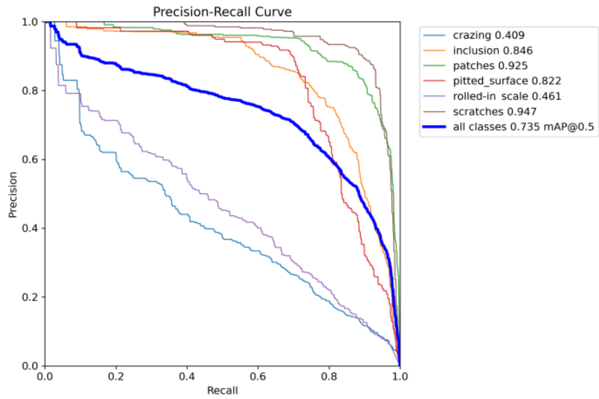


Fig. 7 Precision-Recall curve of the YOLOv8n model

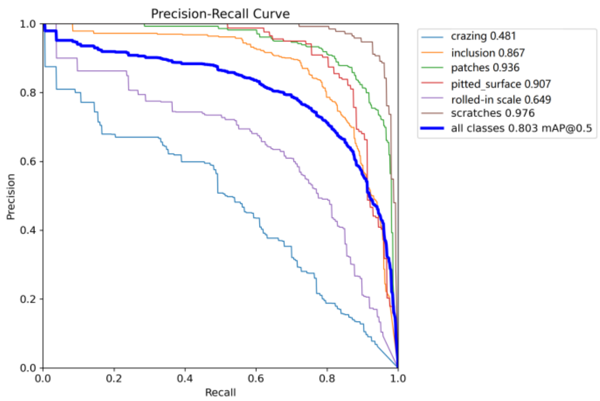


Fig. 8 Precision-Recall curve of the YOLOv8_DSC_EMA model

The comparison between the detection effect of the improved model and the detection effect of the pre-improved model is shown in Fig. 9.

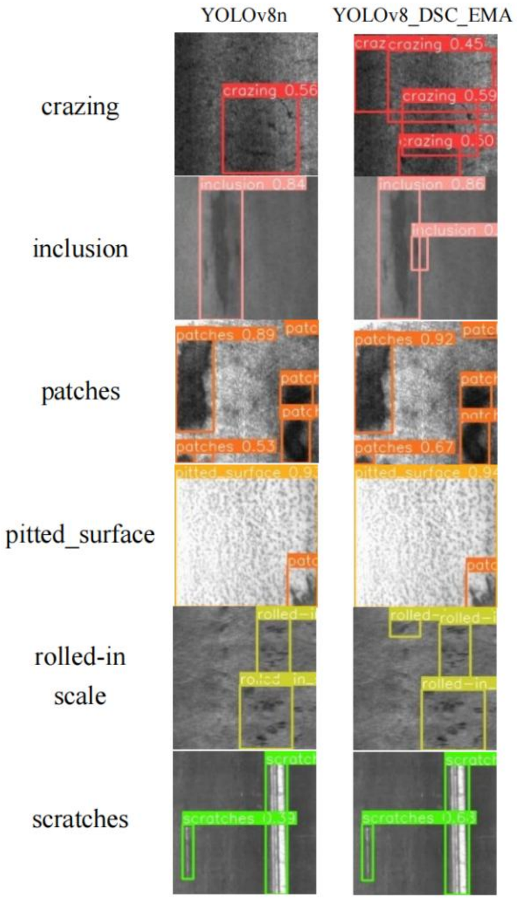


Fig. 9 Comparison chart of various defect detection and testing effects

Through the comparison figure, it can be seen that after the improvement, the average precision of the model is improved obviously, the confidence of the defect recognition in the detection result figure is improved to a certain extent, and the positioning of the recognition box is more accurate, which verifies the effectiveness of the improvement.

5. Conclusion

With the continuous innovation of deep learning theories and methods, a large number of deep learning algorithms have been applied to industrial production in the field of object detection. Based on the data set of steel strip surface defects collected by Professor Song Kechen's team at Northeastern University, this paper designs an improved surface defect detection algorithm based on YOLOv8n model, which has achieved good results in the detection accuracy and real-time performance of steel strip surface defects. Firstly, from the perspective of steel strip production process, combined with the defect types in the data set, the formation cause and visual characteristics of each defect are analyzed, and the YOLO series target detection method is selected. In order to facilitate model training, the labels of the dataset are converted from VOC format to YOLO format, and then the dataset is expanded, the 1800 images of the original dataset are expanded to 5400 images, and the image coordinate mapping formula is calculated and the corresponding labels are modified. Complete the preprocessing of the dataset. YOLOv8n was used as the baseline model for experiments, and the experimental results were analyzed. In order to solve the problem of low accuracy of defect detection such as small target and small edge gray change, DSC is used to replace the deep convolution calculation method of the backbone network and the EMA attention mechanism is introduced to lightweight the model, which can realize that the calculation amount is not greatly improved under the premise of ensuring the accuracy of the model, saving calculation cost and improving calculation speed. The experimental results indicate that the enhanced model achieves an average accuracy of 80.3%, representing a 6.8% improvement over the original model.

Although the improved method proposed in this paper has brought about a considerable improvement in accuracy, there is a definite increase in computational effort. In future research endeavors, attempts at lightweight enhancement can be made to reduce its computational load and enhance the detection speed of the network. Additionally, its accuracy still holds some potential for improvement, and a new detection head could be incorporated in the shallow layer to further augment the detection capability of small target defects.

References

1. Hong-Tao Z, Fa-Jie D, Ke-Qin D. Study on On-Line Surface Defect Detection Vision System for Steel Strip. *Chinese Journal of Sensors and Actuators*, 2007.
2. Redmon J , Divvala S , Girshick R ,et al. You Only Look Once: Unified, Real-Time Object Detection. *Computer Vision & Pattern Recognition.IEEE*, 2016.
3. Bochkovskiy A , Wang C Y , Liao H Y M .YOLOv4: Optimal Speed and Accuracy of Object Detection. *ArXiv*, 2020.
4. Song K C, Yan Y H. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Applied Surface Science*, 2013, Volume 285, Part B, 858-864.

5. Qi Y, He Y, Qi X, et al. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023: 6070-6079.
6. Ouyang D , He S , Zhan J ,et al. Efficient Multi-Scale Attention Module with Cross-Spatial Learning. *ArXiv*, 2023

Authors Introduction

Mr. Haozhe Sun

He received his B.S. degree from College of Electronic Information and Automation, Tianjin University of Science and Technoogy, China in 2023. He is currently a Master course student in Tianjin University of Science and Technology. His research area is about deep learning and image processing.

Dr. Fengzhi Dai

He received M.E. and Doctor of Engineering (PhD) from the Beijing Institute of Technology, China in 1998 and Oita University, Japan in 2004 respectively. His main research interests are artificial intelligence, pattern recognition and robotics. He worked in National Institute of Technology, Matsue College, Japan from 2003 to 2009. Since October 2009, he has been the staff in College of Electronic Information and Automation, Tianjin University of Science and Technology, China.

Mr. Junjin Chen

He received his bachelor's degree in Mechanical Design, Manufacturing and Automation from Beijing Union University in 2014. His research interests are industrial automation and robotics. From 2005 to 2007, he worked at Beijing Aeronautical Manufacturing Engineering Research Institute. Since September 2007, he has been working at SMC (Beijing) Manufacturing Co., LTD., China.

Prediction of Winter Wheat Growth Trends Based on NDVI Vegetation Index

Lu Kang *, Jiahao Xie, Chunli Li, Haoran Gong, Fengzhi Dai
Tianjin University of Science and Technology, Tianjin, China

*E-mail: *2788344006@qq.com*

Abstract

As modern technology emerged, the level of agricultural remote sensing has been further improved. This paper takes winter wheat as the research object, studying on Liangshan County in Shandong Province, where the planting coverage of winter wheat is high. Image preprocessing is carried out using ArcGIS, combined with ENVI to invoke satellite data in the near-infrared and infrared bands to calculate the NDVI index from the regreening stage to the maturity stage of winter wheat in this area. During the maturity stage of winter wheat, NDRE is used instead of NDVI to solve the problem of inaccurate NDVI measurement in high-density vegetation coverage. The simulation results show that the test data matches the actual winter wheat output value.

Keywords: NDVI, winter wheat, agricultural remote sensing, ENVI, ArcGIS

1. Introduction

The most important thing for the people is agriculture. Agriculture is the foundation of the national economy. With the continuous development of science and technology and agricultural digitization, agricultural remote sensing stands out with its advantages of wide observation range, large amount of information and long monitoring duration, and has become one of the important means of current agricultural management [1,2]. Winter wheat is one of the three major food crops in the world. How to accurately monitor crop growth and establish agricultural disaster intervention system has become an important task for agricultural workers. In this paper, by combining remote sensing and GIS technology, the typical vegetation index NDVI was selected to analyze the growth of winter wheat. This paper provided new ideas for decision-making management of winter wheat production.

2. Research areas and methods

2.1. Overview of the study area

Shandong Province is located in the middle and lower reaches of the Yellow River, with a warm temperate monsoon climate type, mild climate and sufficient sunlight. Shandong Province is suitable for the growth of various crops. Winter wheat is widely planted in the province, accounting for about 14% of the national wheat planting area. As one of the typical representatives of winter wheat planting areas, Liangshan County has the characteristics of suitable climate and flat terrain. Its planting area remains high, and the planting area will reach 938,000 mu in 2024. At the same time, most other crops are in the sowing period during the winter wheat growth period, with less interference.

2.2. Data source and preprocessing

The Landsat8 satellite is loaded with multi-spectral and thermal infrared sensors. Its panchromatic band can clearly capture the subtle features of the earth's surface. The data spatial resolution reaches 30 meters, and some resolutions are as high as 15 meters. At the same time, due to its long-term series and multi-band coverage, it can provide rich surface information. And it has significant advantages in analyte growth and changes.

On the geospatial data cloud platform, locate Liangshan County and select Landsat 8 OLI satellite products to download satellite data of Liangshan County. Then use the downloaded shapefile data of China's administrative regions. Query Liangshan County according to the attribute table of the surface data, export the graphical data of Liangshan County and combine it with the downloaded spatial data in ENVI. Image cropping was performed to obtain Liangshan County spatial remote sensing data. The spatial remote sensing data of Liangshan Count is shown in Fig1.

Fig.1. Atmospheric correction remote sensing image

2.3. NDVI Index Calculation

Vegetation normalization index (NDVI) is an important parameter used to reflect vegetation growth status and coverage. It can transform multispectral data into a single image band to characterize vegetation distribution characteristics [3]. At the same time, NDVI can reflect the background influence of plant canopy, such as soil, wet ground, snow, dead leaves, roughness, etc. Its value is between-1 and 1. The closer its value is to 1, the more vegetation is distributed and the greater the coverage rate. When it is 0, it represents the corresponding place for rock, bare soil, etc. While when its value is negative, it means that the ground cover is water, snow and other non-vegetation substances. The reflectivity of chlorophyll to the spectrum is obviously different. It mainly absorbs red light and blue-violet light for photosynthesis and reflects green light, making plants appear green. The near-infrared band is the main band reflected by the internal structure of plant leaves (such as cell wall, mesophyll, etc.). The NDVI is calculated by using near-infrared (Band5) and infrared band (Band4). The calculation formula is:

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}} \tag{1}$$

(where, NIR Is the reflectivity in the near infrared band,REDIs the reflectivity in the infrared band)

Although NDVI can quantitatively determine the growth of winter wheat in the region, NIR is difficult to reach the lower part of the plant canopy. Some of the lower canopy NIR light was ignored, and high levels of chlorophyll had already accumulated in the leaves during the winter wheat growth spurt. Measurements using NDRE are more accurate than NDVI. Using NDRE in its maturity stage, it can not only reflect its maturity, but also evaluate the health status of crops. In the mature stage of vegetation, NDRE begins to decrease due to the change of wave band, but it decreases slowly. If it decreases rapidly, it means that the vegetation may be in danger of pests and diseases, so it can be found and treated in time. At the point where NDVI falls to a certain threshold, it means that the winter wheat can start to be harvested. Harvesting winter wheat at this time of year prevents damage to yields by harvesting too early or too late. Utilization helps predict the timing of harvesting winter wheat and corresponding field management. The calculation formula is:

$$\mathrm{NDRE} = \frac{\mathrm{NIR} - \mathrm{RE}}{\mathrm{NIR} + \mathrm{RE}} \tag{2}$$

(HIRis the reflectivity in the near infrared band,RE is the reflectivity of the red edge band)

In order to visualize the growth of winter wheat, this paper establishes a geographic information display platform for crop growth. The Brouser-Server (BS) architecture is utilized so that the browser acts as a client in conjunction with the server to form the architecture, as shown in Fig.2. That is, the customer only needs to have a browser to carry out the appropriate services, and log in to the system by means of an account password, etc., in order to obtain the relevant privileges and services of the system. This paper utilizes the vs code integrated development environment to write a geographic information management system front-end framework, call leaflet open source JavaScrip library. The paper realizes the online map development of Liangshan County, as shown in Fig.3. The paper calls the Baidu map API interface to use Baidu map information as the core content of the visualized geographic screen and import the processed NDVI data into the Mysql database. It predicts the future growth condition of winter wheat in Liangshan County as well as the yield situation. And it sets up the early warning function through the threshold method discovers the abnormal state of winter wheat growth in time. It also discovers the abnormal state of winter wheat growth in time. The above measures enable rationalization suggestions and initiatives, and finally the platform is piggybacked to run on AliCloud servers.

Fig.2. Platform login interface

Fig.3. Liangshan County Visual Large Screen

3. Results and analysis

3.1. Analysis of horizontal differences in NDVI

The phenological stage of winter wheat shows temporal and spatial characteristics [4]. The main phenological stages are divided into regreening stage, jointing stage, heading stage and maturity stage. In different growth intervals, the chlorophyll content of winter wheat is different, and the corresponding NDVI changes. The four-month remote sensing data of winter wheat phenology in 2018 were processed as Fig.4, and two NDVI trend curves were fitted and drawn three times as Fig.5 Comparison of the results of the analysis shows that, in time, the two trend curve changes converge to show faster growth. Winter wheat thrived during the phenological period and chlorophyll area began to increase significantly, leading to a decrease in NIR. RED began to increase, and the NDVI index changed significantly from 0.373348 at the regreening stage to 0.735383 at the heading stage. In June, the regional average NDVI reached the highest value of around 0.44, and then entered a period of slow growth or even declined. At this time, winter wheat reached the maturity stage, chlorophyll began to decrease, RED began to decrease, and NDVI began to decrease in some areas. Spatially, the NDVI in the eastern and southern parts of Liangshan County took the lead in increasing to a high value, while the growth in the western parts was slower. As shown in Fig.4, the regional difference results were also highly correlated with the winter wheat yield in this region in 2018. So far, studies have confirmed this idea [5], and realize the application [6,7]. There is also a constant low-value area in the middle. According to the characteristics of remote sensing images, it can be further found that this area has been widely developed as a commercial area. And this area has been widely developed as a concentrated population residence. This conclusion can also be used to judge the prosperity of the city.

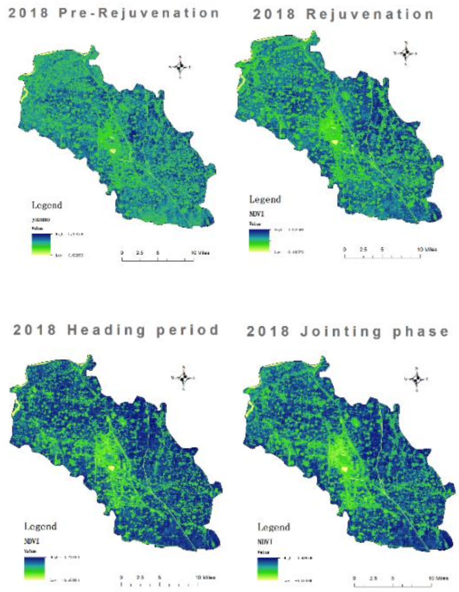


Fig.4. Distribution of NDVI in the winter small phenological period of 2018

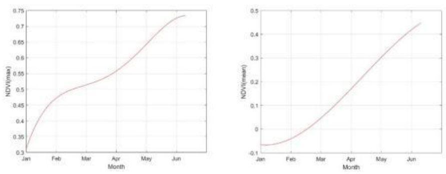


Fig.5. NDVI trend curve of winter phenology in 2018

According to the fitting data of two trend curves and using the method of threshold setting, we can evaluate and analyze the growth of any area in any period and draw the evaluation grade chart. The value of NDVI near the regional average is regarded as normal growth, the value far lower than the average is invalid, and the value significantly higher than the average and close to the maximum value is growth preference. The growth evaluation of the regreening period in 2018 is shown in Fig.5. The image intuitively shows the crop growth status of each village and town, and the information content is large. On the whole, the plants grow well in the north and south of Liangshan County. But not in the east, it has a certain guiding effect on the growth of winter wheat.

3.2. Comparison of longitudinal interannual changes of NDVI

Taking the regreening period as an example. This paper processed and analyzed the 2016-2021 five-year winter wheat regreening period tele-imagery (2017 data quality is poor). It was found that there were large inter-annual differences in NDVI. The overall trend shows an increase, which can be considered as an increase in winter wheat production. The results showed that there was a large interannual difference in NDVI, and the overall trend showed an increase. It can be considered that the yield of winter wheat increased. In order to eliminate the influence of irrelevant factors and judge the interannual difference of NDVI from the whole, the NDVI distribution image is binarized. 0 represents the part where NDVI is less than 0.3 and 1 represents the part where NDVI is greater than 0.3. As shown in Fig.7, the result is more intuitive. From the change curve in Fig.8, it can be intuitively seen that 2020 and 2021 have a certain fluctuation trend compared with previous years, from the average NDVI of regional winter wheat regreening period of 0.047593 in 2016 to the average NDVI of 0.127220 in 2020 and 0.058684 in 2021. The NDVI of winter wheat in the annual phenological period is treated by regional average value, and the results can be made as the evaluation grade map shown in Fig.6, and the growth of winter wheat in Liangshan County can be observed from the interannual scale. By consulting the relevant local policies and regional technology introduction strategies, it can be inferred that the winter wheat harvest in Liangshan County has been considerable in recent years. It shows that the introduction of new technologies has improved the agricultural yield and quality.

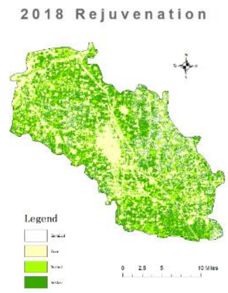


Fig.6. Grade Chart of Winter Wheat Growth Evaluation in Phenological Period in 2018

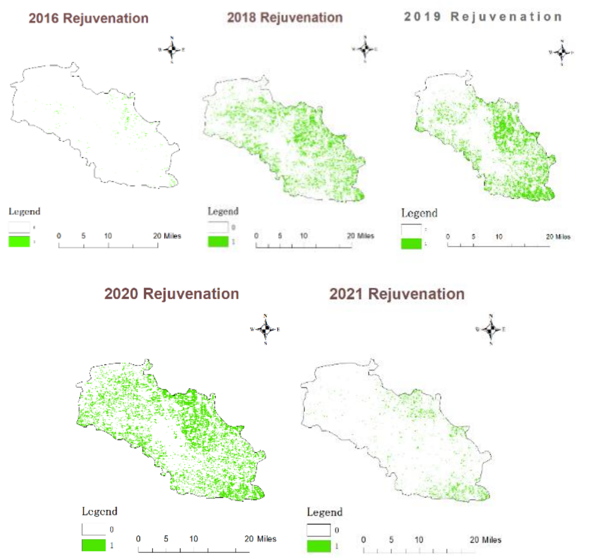


Fig.7. NDVI binarization image of Liangshan County from 2016 to 2021

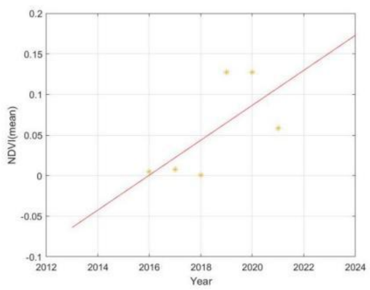


Fig.8. Interannual trend curve of NDVI

4. Conclusion and Prospect

In this paper, winter wheat was used as the research object and positioned in Liangshan County. In this paper, we use remote sensing technology tools to analyze the growth of winter wheat by calculating NDVI. The results showed that NDVI index could better reflect the growth status of winter wheat, and had a positive correlation with the yield. At the same time, the index had interannual and short-term temporal changes, which had important research significance for judging the growth and overall maturity of winter wheat in a region. Using NDRE instead of NDVI for follow-up analysis in the mature stage of winter wheat can better reflect the growth status of high-density vegetation and solve the problem of crown shading. In this paper, NDVI prediction of winter wheat growth also has some areas to be improved, and the influence of human activities and extreme weather may lead to the occurrence of outliers [8]. Some non-wheat crops may affect NDVI values to a small extent.

In future research, the application of other remote sensing indices in crop growth analysis can be further explored. Information from multiple sources can be combined for comprehensive analysis to improve the accuracy and reliability of the analysis. The latest research results generated are applied in the growth analysis of other crops to provide more comprehensive support and services for agricultural production.

Acknowledgments

This article partially supports the 2024 Student Innovation and Entrepreneurship Training Program project （202410057039）

References

1. Zhang Yeping, Zhang Mingwei, Sun Ruijing, et al. Application of Fengyun Meteorological Satellite in Agricultural Remote Sensing. Science and Technology Herald, 2021, 39 (15): 39-45.
2. Zhuang Jiayu, Bao Weijia, Su Wuzheng. Present situation and prospect of agricultural remote sensing application. Agricultural Outlook, 2024, 20 (04): 68-74.
3. Liu Qinghua, Ren Jintong, Ren Fang. Analysis of NDVI Variation Characteristics in Liuchong River Basin Based on PIE-Engine. Science and Technology and Innovation, 2023, (12): 26-28 +34.
4. Zhao Yijing, Wang Xiaoli, Hou Xiyong, et al. Spatial and temporal characteristics of key phenological stages of winter wheat in Shandong Province from 2003 to 2019. Acta Ecologica Sinica, 2021, 41 (19): 7785-7795.
5. Pan Yangyang. Remote sensing monitoring and yield evaluation of leaf area index of jujube garden. Tarim University, 2024. DOI: 10.27708/d.cnki.gtlmd.2024. 00 0157.
6. Wu Yulun. Early rice yield estimation in Tianmen Ci ty, Hubei Province based on NDVI Modern Agricultu re Research, 2023, 29 (09): 18-23. DOI: 10.19704/J.c nki.xdnyyj.2023. 09. 026.
7. Zhang Lei, Yao Mengyao, Liu Zhigang, et al. Estimation of corn yield based on UAV multispectral NDVI value. Xinjiang Agricultural Science, 2024, 61 (04): 845-851.
8. Liu Haihong, Yin Shuyan, Xu Liting, et al. The imp act of extreme climate and human activities on NDVI of different vegetation types in Shandong Province. Journal of Ecology, 2023, 43 (21): 8780-8792. DOI: 10.20103/j.stxb.202209092577.

Authors Introduction

Ms. Lu Kang

She enrolled in Tianjin University of Science and Technology in 2022, majoring in Marine Technology and specializing in the research fields of Physical Oceanography and Remote Sensing Analysis.

Mr. Jiahao Xie

He studied Oceanographic Sciences at Tianjin University of Science and Technology in 2022 and is currently a bachelor student of science majoring in Marine Technology. He expects to receive a bachelor's degree in 2026.

Ms. Chunli Li

She is a third-year undergraduate student pursuing her studies in Translation at Tianjin University of Science and Technology. She is focusing on Scientific and Technical English Translation, Sight Translation, and Comparative Translation between English and Chinese.

Mr. Haoran Gong

He is currently studying in the School of Electronic Information and Automation of Tianjin University of Science and Technology, and is proficient in embedded architecture with strong single-player working ability.

Dr. Fengzhi Dai

He received M.E. and Doctor of Engineering (PhD) from the Beijing Institute of Technology, China in 1998 and Oita University, Japan in 2004 respectively. His main research interests are artificial intelligence, pattern recognition and robotics. He worked in National Institute of Technology, Matsue College, Japan from 2003 to 2009. Since October 2009, he has been the staff in College of Electronic Information and Automation, Tianjin University of Science and Technology, China.

A Study on Artemia Culture System and Its Application

Wanying Zhang *, Yicheng Wu, Ziting Zhang,Yumei Huang
Tianjin University of Science and Technology, Tianjin, China
*E-mail: *3123559440@qq.com*

Abstract

Aimed at the shortcomings of the low efficiency and high cost of Artemia culture, this paper proposes a high-density Artemia culture system based on the Internet of Things technology. The system detects and controls the breeding environment through sensors and actuators, and uses a cloud platform to analyze and process the collected data. Automation control and remote monitoring of the system reduce the cost of breeding and human resource. The system's Internet of Things technology provides scientific basis and decision support for Artemia culture.

Keywords: the Internet of things, intelligent control, artemia, high density culture

1. Introduction

The "14th Five-Year Plan" emphasizes the importance of the fishery sector in implementing new development concepts, expanding food sources, and increasing farmers' income. It also proposes the use of science and technology to promote the high-quality development of the fishery industry. Currently, China’s aquaculture and seedling industries are world leaders (as shown in Fig1).

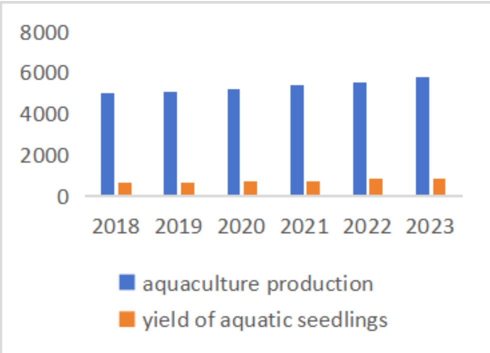


Fig. 1 Production of aquaculture and seed industry

The quality of feed for aquatic seedlings is a key factor in determining the quality of aquatic products. Artemia, with its high protein content and short growth cycle, is widely used as feed in marine aquaculture breeding. Studies indicate that about 90% of artemia eggs come from inland salt lakes. However, these salt lakes are at risk of drying up due to climate change. As the global aquaculture industry continues to expand, the demand for artemia is rising, yet its natural resources are becoming increasingly scarce. Currently, indoor artemia breeding requires frequent manual monitoring and feeding, which is time-consuming and labor-intensive, making large-scale breeding challenging. Therefore, industrialized high-density artemia culture holds significant economic potential. By integrating intelligent artemia breeding systems with Internet of Things technology, breeding efficiency can be greatly improved, costs reduced, and the quality of the final product enhanced.

The structure of this article is as follows: Section 2 presents related research on the growth conditions of artemia. Section 3 describes the overall design, as well as the detection and control modules of the system. Section 4 highlights the advantages of this design compared to traditional artemia breeding methods. Finally, Section 5 provides a summary of the main findings of the paper.

2. Study On Artemia Growth Conditions

Artemia is widely used in aquaculture as a feed for fish, shrimp, and crab larvae due to its rich nutritional value in early developmental stages. To efficiently culture artemia, achieve high hatching rates, and ensure optimal growth during the guarantee period, it is crucial to study the factors influencing its growth.

The growth and survival of artemia result from the combined effects of temperature and salinity. This study focuses on the American Great Salt Lake artemia. Through extensive experiments, it was found that when temperature is constant, artemia's growth length is positively correlated with salinity—higher salinity leads to greater growth. Conversely, when salinity is constant, survival time is negatively correlated with temperature; higher temperatures shorten the survival time of artemia. The results indicate that the optimal growth and survival conditions for artemia occur at 28°C and 20 ‰ salinity. Notably, at a temperature of 24°C, salinity does not significantly affect artemia hatching.[1]

Further studies have revealed that, in addition to temperature and salinity, other environmental factors also play a crucial role in the growth and development of artemia. For example, light intensity influences artemia's photosynthesis and biological rhythms. While appropriate light supports growth and reproduction, excessive or insufficient light inhibits development.[2] Dissolved oxygen is another key factor, as Artemia requires adequate oxygen for respiration and metabolism. Low oxygen levels can result in stunted growth or even mortality. Additionally, the pH level of the water affects artemia's nutrient absorption and enzyme activity.

Maintaining an optimal pH range ensures proper enzyme function and supports normal growth.

3. System Design

3.1. Overall system architecture

The high-density aquaculture system for artemia uses Internet of Things technology to detect and control the growth environment of brine worms. It can be divided into four parts: perception layer, network layer, platform layer, and application layer. Among them, the perception layer detects the environment through sensors and execution pumps; The network layer is implemented by IoT gateways and communication controllers for centralized data collection and transmission; The platform layer is the OneNET cloud platform, which conducts data mining and computation to optimize the incubation plan; The application layer consists of computer, mobile APP, and PAD (both computer and PAD are web versions). Users can monitor the breeding situation in real time and remotely control and manage the breeding environment through these three media.(as shown in Fig2)

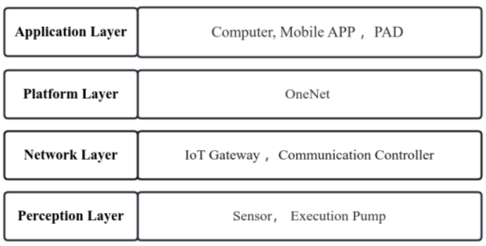


Fig. 2 Overall architecture diagram

3.2. System hardware design

The hardware equipment used in the high-density aquaculture system for brine worms based on Internet of Things technology mainly includes two parts: sensors and actuators. In terms of sensors, it mainly includes temperature sensors, salinity sensors, pH sensors, turbidity sensors, light sensors, dissolved oxygen sensors, etc., to ensure comprehensive monitoring of the aquaculture environment. The actuators include peristaltic pumps, circulation pumps, heaters, quantitative pumps, air pumps, and lighting control equipment. (as shown in Fig3)

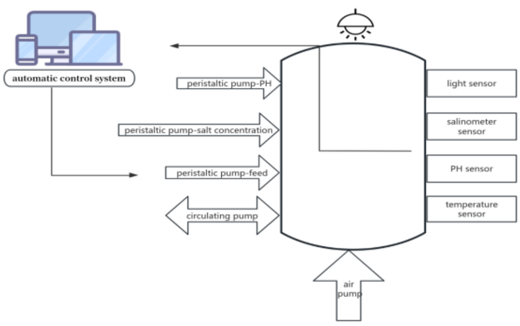


Fig. 3 System Hardware Design Diagram

(1) Water environment monitoring

In the process of brine worm cultivation, environmental conditions are the most important influencing factor on its growth. Different environmental conditions can cause changes in the growth rate of brine worms, and some species of brine worms require specific environmental conditions for growth and reproduction. The following provides a detailed explanation of the core technical principles of water environment monitoring.

Water salinity detection module: The working principle of the salinity sensor is to indirectly measure the salinity of saltwater by measuring the ion concentration in the saltwater using conductivity. That is, when the concentration of salt ions in the water increases, the conductivity also increases.

Water display module: In order to facilitate users to quickly understand the relevant information of the aquaculture environment, the system adopts the USART serial port screen as the water display module (as shown in Fig4). This display screen can display real-time water information such as temperature, light intensity, salinity, pH, etc., helping users accurately analyze and process data.

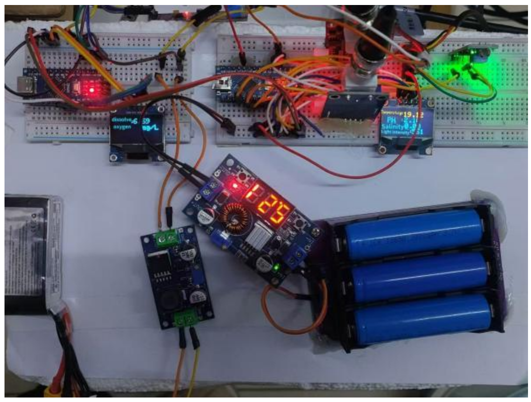


Fig. 4 Water display module

Water dissolved oxygen module [3]: The core technical principle of water dissolved oxygen sensors is fluorescence method, which involves the interaction between dissolved oxygen and fluorescent substances, resulting in a change in fluorescence intensity. When the concentration of dissolved oxygen increases, the fluorescence intensity decreases, and vice versa, it increases. Based on this characteristic, dissolved oxygen sensors can accurately determine the concentration of dissolved oxygen in water by measuring changes in fluorescence intensity.

Water pH detection module: The working principle of the pH sensor is based on the relative voltage between the signal electrode and the reference electrode to measure the pH value of the solution. When the pH value of the solution is 7 (neutral), the theoretical output voltage of the reference electrode is 0. As the acidity or alkalinity of the solution changes, the output voltage will correspondingly undergo positive or negative changes..(as shown in Fig5)

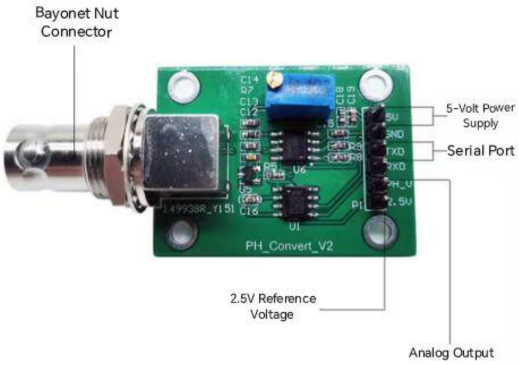


Fig. 5 Water pH detection module

(2) Water environment control

This system controls the water content, feed, temperature, light intensity, and dissolved oxygen in the water environment through a series of actuators. As shown in Fig6.

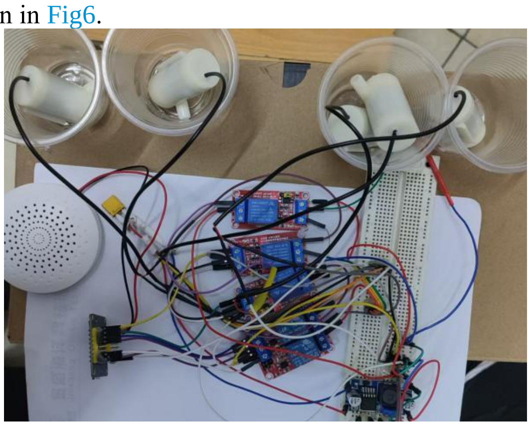


Fig. 6 Water control module

3.3. *Intelligent decision and control module*

The high-density breeding system for brine worms based on Internet of Things technology is designed with automatic water change, temperature control, feeding, oxygenation, and lighting functions in the brine worm breeding environment. Breeding users can monitor the breeding environment in real time through media such as PC and mobile APP.[4] In addition, the addition of voice assistants allows aquaculture users to issue commands to the system through voice, enabling them to regulate the environment of brine shrimp farming. (As shown in Fig7)

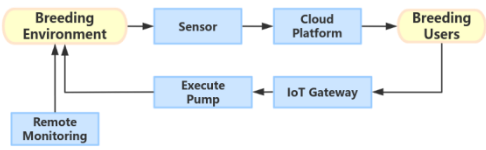


Fig. 7 Schematic diagram of the system operation

The various levels of the system cooperate with each other and work together. The system is equipped with advanced intelligent devices and sensors to monitor various parameters of the breeding environment in real time, and transmit the data to the platform backend for analysis and processing. The platform integrates, analyzes, and mines the collected data through big data management and analysis technology, providing real-time breeding data for breeding users; Through the Internet of Things gateway, the breeding environment can be intelligently controlled, achieving precise control over breeding environment, feed feeding, water quality testing, and other aspects, to meet the growth needs of brine worms to the greatest extent possible. At the same time, this system supports remote monitoring and remote operation functions, making it convenient for breeding users to control and manage the breeding environment anytime and anywhere. The overall schematic diagram of the system is shown in Fig8.

Fig. 8 Overall schematic diagram of the system

4. System Test

A high-density Artemia breeding platform based on Internet of Things technology integrates intelligent systems, data management, and aquaculture production. Through demand analysis and market research, the technical plan for the Artemia intelligent breeding platform was developed, encompassing the platform's basic structure, functional modules, and data processing flow. Following the overall system design, each functional module undergoes testing, which is primarily divided into two parts: the monitoring node test and the data transmission node test.

4.1. *Monitoring node test*

During the testing period, key environmental parameters such as dissolved oxygen, pH, water temperature, and salinity were successfully detected, and corresponding values were accurately obtained. Although there was a slight time lag when uploading data to the web end due to the influence of network transmission speed, the lag was exiguous and had negligible impact on the overall system performance and data accuracy.[4] The display module operated stably without any abnormalities and was able to display important information such as temperature, light intensity, and humidity at any time according to user needs, which providing great convenience for system debugging and daily management. Additionally, the USART serial port screen equipped with the system had excellent data processing capabilities, allowing for rapid and accurate processing of various data, so that fishermen and professional technicians could analyze the data timely and accurately, thereby making scientific and reasonable breeding decisions.

4.2. *Data transfer node test*

As a crucial hub connecting monitoring nodes and the cloud platform, the data transfer node is responsible for vital data transmission and command interaction tasks. In this system, the data transfer node employs Node MCU to achieve A/D conversion signal acquisition, the design that ensures the accuracy of data collection and lays a solid foundation for subsequent precise control. For data transmission, the system selects Wi-Fi for wireless transmission, enabling stable and efficient transmission of collected data to the network for storage. This ensures the timeliness and integrity of the data while also providing strong support for data processing and analysis on the cloud platform. Through rigorous testing of the data transfer node, it is ensured that it can stably and reliably achieve bidirectional data transmission during system operation, effectively guaranteeing the normal operation of the entire aquaculture platform.

5. Contrastive Analysis

Traditional Artemia aquaculture techniques mostly rely on existing salt pans and mudflat resources, which have the merit of certain cost advantages and relatively simple operations. However, this model faces numerous challenges, with significant impacts from uncertain factors such as climate and seasons, leading to large fluctuations in production and difficulties in ensuring stable Artemia quality. For instance, extreme weather conditions like heavy rainfall and drought can damage the growth environment of artemia, trigger pests and diseases, and severely affect both the yield and quality of artemia.

In comparison, the high-density Artemia aquaculture system designed by us demonstrates significant advantages. Equipped with precision sensors and advanced controllers, the system enables full-process and multi-dimensional precise monitoring and control of various parameters in the artemia aquaculture environment. From precise regulation of the aquaculture environment to delivery feed scientifically, and real-time water quality monitoring, every aspect is managed with fine precision, greatly enhancing the stability of Artemia production and quality.

6. Summery And Prospect

This project has developed a high-density Artemia aquaculture system relying on modern Internet of Things (IOT) technology, providing a novel platform for artemia aquaculture. Additionally, the system innovates in the application of artificial intelligence, the integrated use of detection equipment, and the overall aquaculture mode of artemia, promoting the overall transformation and upgrading of the industry. In the future, the high-density Artemia aquaculture system based on IOT technology will leverage its advanced technological advantages to propel the aquaculture industry towards intelligence and sustainable development.

References

1. Chen Xindan, Zhang Haiyan, Fan Yanwen, et al. Effect of temperature and salinity on the growth and survival of halogen worms; *proceedings of the Annual meeting of Shanghai Zoological Society 1999*, Shanghai, China, F, 1999.
2. Jiang Xiang, Wei Yitong, Xu Lele, et al. The influence of light intensity, temperature and seawater type on the egg hatching rate and juvenile survival rate of brine worm (Artemia saline) . Aquatic Science and Technology Intelligence, 2021,48 (04): 197-202
3. Kou Fu, Ma Zhanjun, Wang Shuo. Research on intelligent Aquaculture Monitoring System based on the Internet of Things . *Science and Technology Information*, 2024,22 (19): 79-81 +5
4. Tang Chaojing. Design of the aquaculture system based on the Internet of Things technology . *Internet of Things Technology*, 2024,14 (02): 82-5 + 9.

Authors Introduction

Ms. Wanying Zhang

She studied at Tianjin University of Science and Technology in 2022 and is currently a bachelor student in Marine Academy. She expects to receive a bachelor's degree in 2026.

Ms. Yicheng Wu

She is currently studying in Tianjin University of Science and Technology, she has mastered the knowledge of marine resources development and application, protein extraction and content determination.

Ms. Ziting Zhang

She was admitted to Tianjin University of Science and Technology in China in 2022 and is currently an undergraduate student in the School of Oceanography and Environment at Tianjin University of Science and Technology

Ms. Yumei Huang

She studied at Tianjin University of Science and Technology in 2021.She is now the first major in her faculty and is expected to receive her Bachelor’s degree in 2025.

Machine Vision-Based Chamfer Detection for Metal Parts

Shangying Han[1]*, Kaili Guo[1], Yanzi Kong[1], Yanliang Gong[1], Junjin Chen[2], Ce Bian[3], Mengfan Zhang[3]

[1] Tianjin University of Science and Technology, Tianjin, China

[2] SMC (Beijing) Manufacturing Co., LTD., China

[3] Tianjin Tianke Intelligent Manufacture Technology Co., LTD., China

*E-mail: * hsy_011023@163.com*

Abstract

This paper introduces a detection system specifically designed for chamfering in metal holes, aimed at achieving precise detection of the chamfers. Chamfering, as a process of beveling the edges or corners of metal parts, plays a crucial role in the subsequent machining and assembly stages. Through multiple experimental validations, this paper employs an industrial camera with a telecentric lens to capture images of the metal chamfers, achieving optimal results. This paper utilizes computer vision techniques to accurately identify the location of the chamfers and delineate their dimensions. A comprehensive analysis of the chamfer radius effectively determines the presence of defects.

Keywords: visual inspection, machine learning, image segmentation, contour curve

1. Introduction

Metal, as one of the main raw materials for industrial products, inevitably develops various defects during the manufacturing process, such as scratches, dents, thread hole residues, and lack of chamfering [1, 2]. These defects not only affect the appearance of the product, but more importantly, they also impact its performance. Therefore, defect detection of parts during industrial production is essential [3]. Traditional manual inspection methods are prone to missing or incorrect detections, and they are inefficient and costly, which does not meet the demand for automated and rapid inspection [4, 5]. Thus, this paper employs machine learning to process captured images for automated defect detection.

This paper explores the fundamental methods for automated chamfer defect detection. The approach involves capturing images using a light source and an industrial camera under varying angles and light intensity conditions to determine the optimal imaging setup. Computer software, combined with machine vision technology, is employed to segment the acquired images. Defect features within the region of interest are extracted, and the outer contour of the features is fitted into a circular shape. The radius of the feature circle is used to precisely determine whether the chamfer is defective. This ensures efficient and accurate automated detection [6].

Based The content of this paper is organized as follows. The second section introduces the methods for image acquisition and the main theories of image processing. The third section describes the three stages of this project: data acquisition, image processing, and data validation. The fourth section provides a summary and conclusion of the main content of the paper.

2. Methods

For camera preparation, the selection of the camera type must consider both practical engineering requirements and economic efficiency. In this project, the RS-A2300-GM60-M10 model was carefully chosen based on the required chamfer size range and defect depth specifications. This camera meets technical requirements with its suitable resolution and depth of field. As a CMOS camera, it also aligns with economic considerations. It provides high-quality image data for subsequent image processing and analysis, ensuring the accuracy and reliability of defect detection [7].

When selecting a light source, it is essential to enhance the contrast between defects and their surroundings to highlight target features. A clear distinction should be created between the inspected and non-inspected parts of the object to improve contrast. The chosen light source must provide sufficient brightness and stability when illuminating the object. Additionally, the effects of light source color, camera type, and light source position on image acquisition must be considered [8-10]. Based on multiple experimental tests, a white ring light source (XHJ-R-35-90-W) was selected for this study. The detailed placement method is shown in Fig.1.

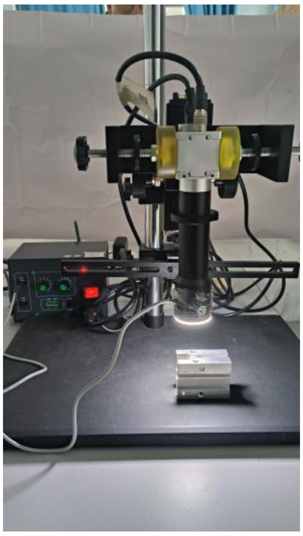


Fig.1 Photographic setup

In the chamfer inspection process, the core objective is to determine whether the target hole has undergone chamfering. Through photographic observation and analysis, it was found that the presence or absence of chamfering leads to significant differences in the radius of the outer contour circle. Holes with chamfering exhibit an increased outer contour circle radius. Based on this feature, various technical methods can be applied to accurately extract the radius data of the outer contour circle. By analyzing the size of this radius, it is possible to accurately determine whether the hole has been chamfered.

2.1. Region of interest (ROI) selection

In the field of machine vision and image processing, the critical area of interest within an image, defined by a rectangle, circle, ellipse, or irregular polygon, is referred to as the ROI. By explicitly specifying the image area to be processed, the ROI effectively reduces the amount of data handled by algorithms, significantly enhancing detection speed and efficiency. Focusing on specific areas also eliminates interference from irrelevant signals in the image, improving detection accuracy and reducing errors. In chamfer detection, creating ROI regions for key positions not only accelerates the detection process but also minimizes the influence of unrelated factors, making the detection process more efficient and precise.

2.2. Extraction method of detection images

The image is initially processed using a thresholding method, where the grayscale features of the image are used to calculate grayscale segmentation thresholds to segment the image. The theory involves comparing the grayscale value of each pixel with the threshold, and based on this comparison, the image is segmented. In this experiment, two fixed thresholds, T1 and T2, are used to binarize the image.

$$Out(x,y)\begin{cases}255 & T_1 \le p(x,y) \le T_2 \\ 0 & \end{cases} \qquad (1)$$

Grayscale inversion is a widely used technique in image processing that effectively changes the light and dark distribution patterns of an image. Through this operation, the bright areas of the image are transformed into dark regions, while the originally dark areas become bright [11]. Implementing grayscale inversion not only helps to more clearly highlight the ROI, but also provides significant convenience for subsequent image segmentation and morphological processing steps.

$$Out(x,y) = 255 - p(x,y) \qquad (2)$$

The image domain reduction operation limits the content of the input image to a specified ROI. It reduces the domain of the input image to the portion that intersects with the defined ROI, without changing the actual size or matrix dimensions of the image. This allows processing to focus on the target area while ignoring irrelevant elements, thus improving processing efficiency. By excluding the background region, it reduces the influence of background noise and focuses on the foreground area, which helps to enhance the accuracy of the algorithm.

$$I_{reduced}(x,y) = \begin{cases} I(x,y), & if\ R(x,y) = 1 \\ 0, & if\ R(x,y) = 0 \end{cases} \qquad (3)$$

2.3. Chamfer contour extraction method

In the algebraic method for circle fitting based on XLD contours, the contour of the circle is fitted by minimizing the algebraic distance between the contour points and the fitted circle. This process aims to find the best circle and output its radius and center coordinates [12]. The algebraic distance refers to the squared difference between a contour point and the corresponding point on the fitted circle. The method iteratively adjusts the circle's center coordinates and radius, optimizing the fitting process. This optimization minimizes the sum of the algebraic distances from all contour points to the fitted circle, resulting in the best circle fit. The radius of the fitted circle represents the radius of the chamfered contour. By evaluating the size of this radius, it is possible to detect whether the chamfer is present. The formula for the algebraic distance is as follows:

$$L_i = (x_i - a)^2 + (y_i - b)^2 - R^2 \qquad (4)$$

To find the minimum value, an iterative method is used. The values of a, b, and R (representing the circle's center coordinates and radius) are continuously adjusted until the target value becomes sufficiently small or converges to a stable solution. This iterative process refines the circle's parameters by minimizing the algebraic distance, ensuring an optimal fit. The algorithm proceeds until the change in the parameters is negligible or the stopping criterion is met, resulting in the most accurate circle fitting for the given contour.

$$F(a,b,R) = \sum_{i=1}^{n}[(x_i - a)^2 + (y_i - b)^2 - R^2]^2 \qquad (5)$$

3. Experimental Data and Experimental Procedure

This project consists of three core stages: the data acquisition stage, the image processing stage, and the data validation stage.

3.1. Experimental data collection

In this study, the RS-A2300-GM60-M10 industrial camera, along with a ring light source, was used to perform multiple systematic image acquisition tasks on various types of workpieces. The captured image data were immediately transmitted to the computer system for in-depth analysis and processing using Halcon machine vision software. Through comprehensive comparison and analysis of the datasets obtained under different shooting conditions, the optimal experimental data were selected. The purpose of this process was to adjust the shooting parameters and find the optimal position, providing a foundation for subsequent image processing.

3.2. Image processing steps

After the image is captured by the camera (Fig.2), it is transmitted to the computer for processing. The image is then analyzed using code to detect any defects.

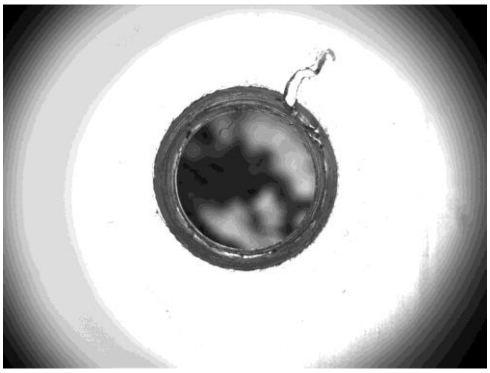


Fig.2. Captured image example

The specific flowchart is as follows (Fig.3):

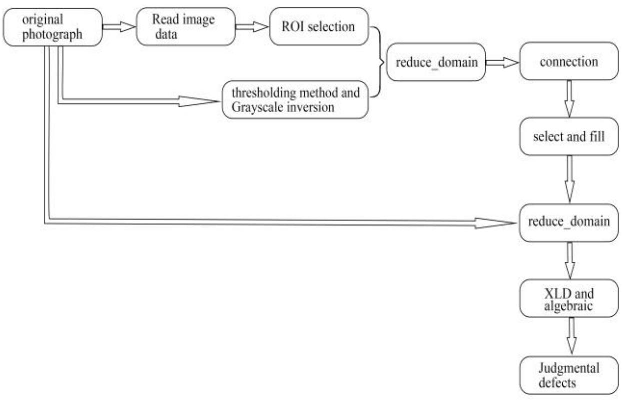


Fig.3. Image processing flowchart

(1) Selection of ROI.

The image is imported into dedicated software for subsequent precise analysis and processing. The first step is to extract the basic properties of the image, namely its size, which is represented by the total number of pixels. This step is crucial for positioning and setting the scale for the following operations. After determining the optimal shooting position, a circular ROI with a fixed radius is defined, centered on the middle of the image. This selection of the ROI not only simplifies the analysis but also focuses on the most informative part of the image, thereby improving the accuracy and efficiency of the analysis. Next, the image domain is reduced using the domain reduction operation, and the image (Fig.2) is cropped to the previously selected ROI. This effectively removes background or redundant information unrelated to the subsequent analysis, making the process more focused and efficient. The extraction result is shown in Fig.4.

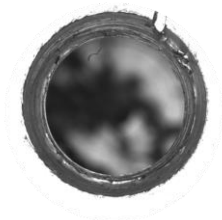


Fig.4. ROI extraction

(2) ROI Processing

After extracting the ROI, the next step is to further process this area. In the initial stage, since the selected radius is relatively large, it is necessary to remove unnecessary parts within the ROI to effectively eliminate potential interference. A thresholding operation is applied to the image (Fig.2) to identify and select the threshold range corresponding to the parts unrelated to the ROI. This step helps to distinguish the unrelated areas from the ROI. Next, a grayscale inversion is applied to the image. The purpose of this operation is to invert the grayscale values, which facilitates a subtractive effect during the subsequent domain reduction operation. This makes it easier to separate the ROI from surrounding unrelated parts more clearly. Finally, the grayscale-inverted area is combined with the original ROI for domain reduction. This step essentially performs an initial processing of the ROI. By reducing the domain, the interference from irrelevant areas surrounding the ROI is effectively removed, allowing for precise extraction and purification of the ROI. The result is shown in Fig.5.

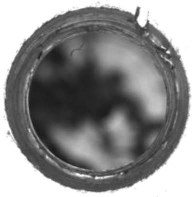
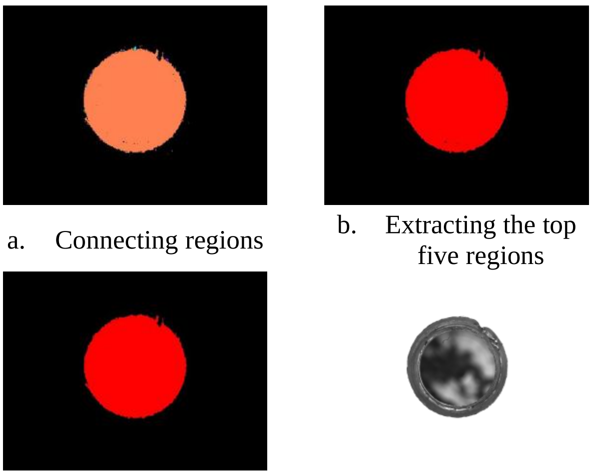


Fig.5. Captured image example

(3) Contour extraction preparation

After the initial processing of the Region of Interest (ROI), a connection region operation was applied to further process the image, separating the sub-regions within the ROI. This allowed each region to be processed individually (Fig.6). The area of each region was then calculated and sorted based on size. The area calculation not only provides information about the scale of each region but also serves as a basis for subsequent feature selection. To effectively exclude interference from unrelated areas, the top five regions were selected based on their area size as feature items (Fig.6). To facilitate the following contour extraction and feature analysis, the selected top five regions were filled and merged into a single unified region (Fig.6). This step not only simplifies the contour extraction process but also enhances the accuracy and efficiency of feature extraction. Finally, a domain reduction operation was applied again, reducing the domain of the image (Fig.2) and the filled regions. This allowed the features of the processed ROI to be extracted efficiently (Fig.6).



a. Connecting regions

b. Extracting the top five regions

c. Image filling

d. Feature extraction result

Fig.6. Preparation process

(4) Contour extraction

After the features are extracted, the next step is to obtain the boundary of the inner region. The boundary region is then used to define the contour points by selecting the outermost boundary pixels. An XLD contour is created based on these boundary points. Finally, an algebraic method is applied to fit a circle to the contour (Fig.7), and the relevant values of the fitted circle, such as the radius, are used as key parameters to determine whether the chamfer has any defects.

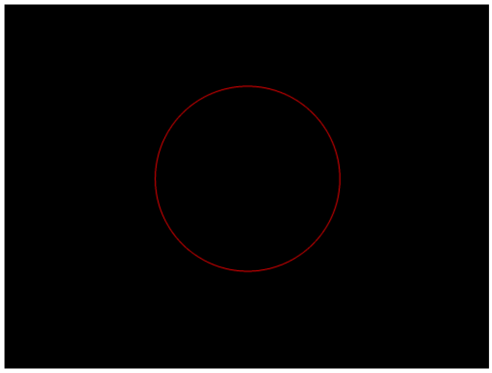
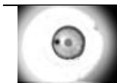
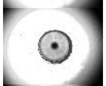
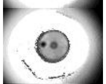
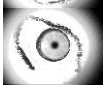
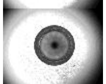
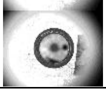


Fig.7. Captured image example

3.3. Extensive experimental data validation

After the code development was completed, it was crucial to validate its correctness and reliability. To ensure the comprehensiveness and rigor of the experiment, a large amount of experimental data was used for verification. The actual radius of the workpiece holes captured in the experiment was 2.8 mm (including the chamfer, Approx. pixels 302), with images taken using the RS-A2300-GM60-M10 camera at a resolution of 1600 × 1200 pixels. A subset of metal hole images was selected as test samples to assess the performance of the code in real-world applications. The experimental data used is detailed in Table 1, which provides an overview of the test samples

Table 1. Overview of selected experiments

Image	Theoretical Contour Radius (in Pixels)	Detected Contour Radius (in Pixels)	Chamfer Detection Results
	302	270.243	Without Chamfer
	302	267.669	Without Chamfer
	302	267.838	Without Chamfer
	302	265.173	Without Chamfer
	302	319.672	With Chamfer
	302	318.11	With Chamfer
	302	323.58	With Chamfer
	302	319.137	With Chamfer

4. Conclusion

This study explores the use of industrial cameras and ring light source technology to capture clearer images of chamfer defects. Advanced machine vision processing methods are then applied to these images for detailed analysis. By precisely defining the ROI, utilizing appropriate threshold settings and grayscale inversion techniques, key information is effectively extracted from the images. Finally, an accurate circle fitting algorithm is used to calculate the outer contour radius of the metal holes. This key parameter provides important support for determining whether chamfer defects are present.

This research holds significant practical value in the field of industrial vision inspection. It not only improves the accuracy and efficiency of chamfer defect detection but also provides strong technical support for the quality control of industrial products. By continuously optimizing the image acquisition and processing workflow, a more intelligent and automated industrial inspection system is expected to be achieved in the future, contributing to the transformation and upgrading of the manufacturing industry.

References

1. Yang, G., Zhu, L., Chen, W., Yu, X., & He, B. (2018). Initiation of Surface Cracks on Beam Blank in the Mold during Continuous Casting. *Metals*, *8*(9). p.712.
2. WU Lin, HAO Hong-Yu and SONG You. " A Review of Metal Surface Defect Detection Based on Computer Vision." *Acta Automatic Sinica 50.* 07 (2024): pp. 1261-1283.
3. Zhou Liang, Wang Zhenhuan, Sun Dongchen, et al. " Present situation and development of modern precision measurement technology " *Chinese Journal of Scientific Instrument 38.* 08 (2017): pp.1869-1878.
4. JIN Yi-jun, LI Zhen-yu and YANG Xu. " Surface Defect Detection of Beer Metal Covers Based on Machine Vision." *Packaging Engineering 44.* 11 (2023): pp.259-267.
5. LI Shao-Bo, YANG Jing and WANG Zheng, et al. " Review of Development and Application of Defect Detection Technology." *Acta Automatic Sinica 46.*11 (2020): pp.2319-2336.
6. Chen Haiyong, Qiu Ruina and Zhao Huifang, et al. " Vision Detection of Small Contour Defects of Chamfering Fine-blanking Parts." *Computer Measurement & Control 26.* 07 (2018): pp.32-37.
7. Chen Jingyuan, Liu Xiao and Du Lili, et al. " Image Noise Simulation and Verification of Area Array CMOS Sensor." *Acta Optica Sinica 44.* 12 (2024): pp.375-383.
8. HOU Yuan-shao. " Choice of Lighting in Machine Vision System." *Journal of Luoyang Normal University 33.* 08 (2014): pp.45-49.
9. HE Xin-yu, ZHAO Shi-lu and ZHANG Zhen, et al. " Development Trend of the Research and Application of Machine Vision." *Machinery Design & Manufacture.* 10 (2020): pp.281-283+287.
10. GAN Yong, YU Jiang-hao and ZENG Bo-qiao, et al. "Analysis and Optimization of Light Source Displacement Errors in Machine Vision System." *Machinery Design & Manufacture.* 03 (2024): pp.27-30+34.
11. GAN Yu-kun, DING Yue-jiao and LUO Zhen-wei, et al. " Research of Shaft Installation Clearance Extraction under Low Contrast Condition." *Computer Technology and Development 30.* 08 (2020): pp.183-187.
12. WANG Ning, DUAN Zhen-yun and ZHAO Wen-hui, et al. " Research on Method of Gear Outer Contour Segmentation in Vision Measurement" *Modular Machine Tool & Automatic Manufacturing Technique.* 04 (2016): pp.117-120.

Authors Introduction

Mr. Shangying Han

He is currently a university student in Tianjin University of Science and Technology. His research area is about mechanical learning and image processing.

Mr. Kaili Guo

He is currently a university student in Tianjin University of Science and Technology, studying mechanical automation and related fields. His research areas are machine vision and image processing.

Ms.Yanzi Kong

She is a lecturer at Tianjin University of Science and Technology. She received her PhD from the University of Chinese Academy of Sciences. Her primary research interests lie in machine vision and active perception.

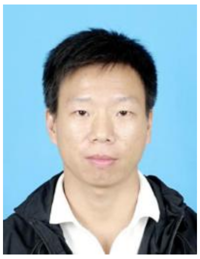
Mr. Yanliang Gong

He is currently a university student at Tianjin University of Science and Technology, with an interest in the field of robotics. His current research areas are machine vision and image processing.

Mr. Junjin Chen

He received his bachelor's degree in Mechanical Design, Manufacturing and Automation from Beijing Union University in 2014. His research interests are industrial automation and robotics. From 2005 to 2007, he worked at Beijing Aeronautical Manufacturing Engineering Research Institute. Since September 2007, he has been working at SMC (Beijing) Manufacturing Co., LTD., China.

Mr. Ce Bian



He received his master's degree in control engineering from Tianjin University of Science and Technology in 2019, focusing on machine vision technology. Since 2017, he has been working in Tianjin Tianke Intelligent Manufacture Technology Co., LTD., serving as the general manager.

Ms. Mengfan Zhang

She received her bachelor's degree in financial management from Zhengzhou College of Commerce in 2019. She specializes in machine vision technology. Since 2019, she has been working as a visual engineer in Tianjin Tianke Intelligent Manufacture Technology Co., LTD., China.

Deep Guard Dog - AI-Based Night Intrusion Detection Mobile Phone Software

Keming Chen*, Jiaxin Wang
Tianjin University of Science and Technology, Tianjin, China
*E-mail: * 18336462861@163.com*

Abstract

This article introduces an Android mobile app called "Super Electronic Watchdog", which aims to solve the problem of home security. The application utilizes Android Studio, NCNN framework and Opt2Ada night vision algorithm to realize humanoid object detection and night image enhancement. Users can switch the camera, select the humanoid detection model and CPU/GPU operation mode, and activate night vision through the app. The application has vibration and voice alarm functions to alert the user that someone has entered the monitored area. The software is divided into Native layer and Java layer, using C++ and Java development, the overall design structure is clear, efficient and practical.

Keywords: home security, Android Studio, humanoid object detection,voice alarm functions

1. Introduction

In today's society, with the improvement of people's living standards and the enhancement of safety awareness, home security[1] has become the focus of increasing attention. From whether the doors and Windows of the house are illegally broken into, to whether the incapacitated people such as children and the elderly stray into dangerous areas such as kitchens and balconies, these potential security risks threaten our daily life. Especially for people who rent houses, travel or travel on business, security cameras cannot be installed for various reasons, and home security problems are more prominent.

To address these challenges, an AI mobile app called "Deep Guard Dog" has been developed. It is developed using Android Studio, Tencent's deep learning reasoning framework NCNN, and our self-developed night vision algorithm Opt2Ada. Through the built-in artificial intelligence algorithm, this app can recognize the humanoid target in the coverage area of the mobile phone camera in real time, and has the function of enhancing the brightness of the mobile phone camera image, even at night can realize the clear recognition of the image.

When the app detects a humanoid object, it will immediately trigger a voice announcement and vibration notification to alert the user that someone is present in the target area. This function is not only suitable for daily home monitoring, but also plays an alarm role for illegal intrusion when the user is sleeping or busy with other things. Compared with the traditional pet dog guard method, "super electronic guard dog" is not only more intelligent and efficient, but also not limited by time, space and other factors, providing a more comprehensive and reliable guarantee for home security.

The rest of this article is organized as follows. The second part is the description of the software. The third part introduces the overall design of the software. The fourth part introduces the related modules of the software, including image acquisition module, night vision enhancement module, human object detection module, voice vibration alarm module, and the fifth part summarizes the main content of this paper.

2. Software description

The camera is used to collect video information, the night image is enhanced by the night vision algorithm deployed in the software, the humanoid target is detected by the depth model, and the mobile phone vibration or voice broadcast is performed according to the detection results. The running interface of the app is shown in the following Fig.1.

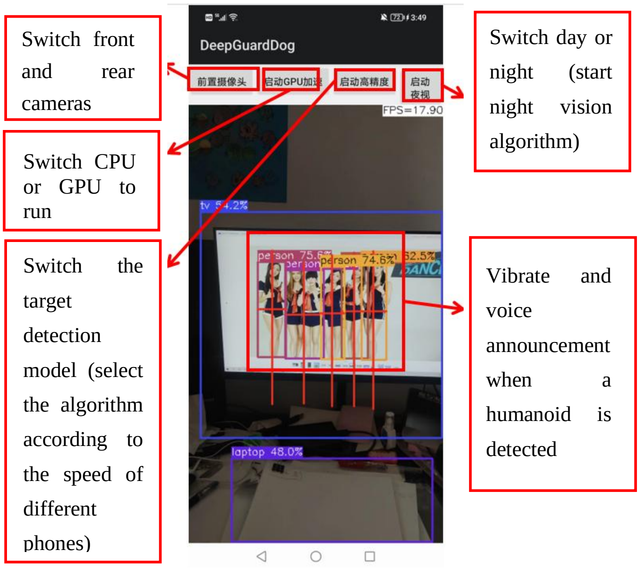


Fig.1 The running interface

2.1. Camera switch (front/rear)

The front/rear camera used to switch the mobile phone can be flexibly placed in the area that needs to be monitored when combined with the mobile phone bracket.

2.2. Humanoid detection model switch

Multiple artificial intelligence algorithms that can be used for humanoid target detection are deployed in the software. On the same phone, different algorithm models require different floating-point arithmetic (FLOPs), that is, the running speed and detection distance are different. Therefore, users can choose the optimal model according to their own mobile phone configuration and actual application scenarios.

2.3. CPU\GPU switch

Users can switch between using the CPU or GPU of the phone to run the software according to their mobile phone configuration. For some high-configuration mobile phones, the software runs faster on the GPU than the CPU.

2.4. Day \ night switch

Images taken by mobile phone cameras at night are too low in illumination to distinguish humanoid targets. To solve this problem, when using the software at night, you can choose to start the night vision algorithm[2], which can improve the illumination of the night image to a certain extent, so as to improve the detection rate and accuracy of humanoid target detection. The night vision function diagram is shown in Fig.2. In Fig.2,the night algorithm is not enabled on the left, while the night vision algorithm is enabled on the right. It can be seen that after the night vision algorithm is turned on, it can effectively detect previously undetected humanoid targets.

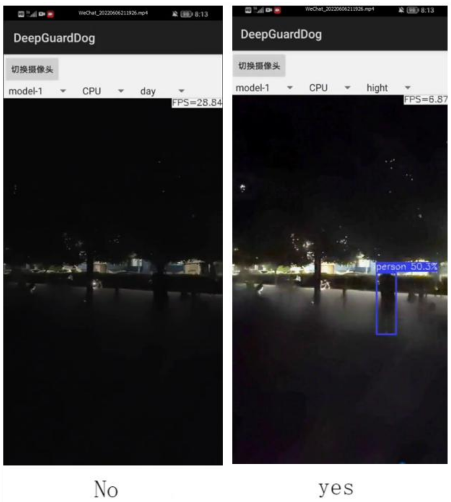


Fig.2 Schematic diagram of night vision function

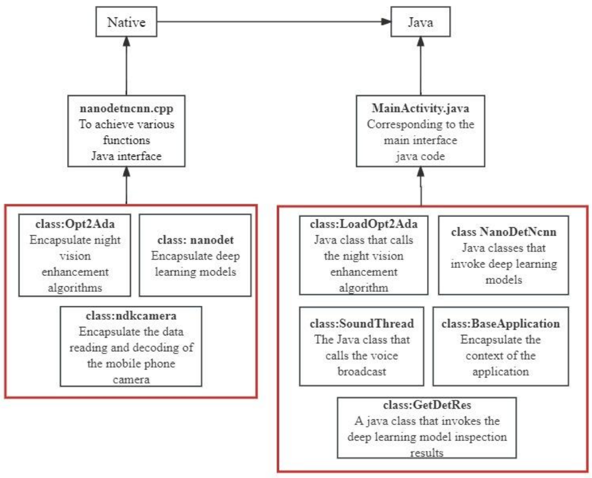
Based on the above design and implementation of the electronic watchdog function that can be realized by this app, it can identify and alarm personnel intrusion in the monitoring area.

3. Overall design

3.1. Design method

This app is developed using Android Studio. The structure of the software is divided into Native layer and Java layer. The Native layer first implements the call and reading of the mobile phone camera, and deploys deep learning algorithm and night vision algorithm, and uses cmake to manage and compile the code. The Native layer is developed using C++, in which the object detection algorithm is deployed using Tencent's open-source deep learning reasoning framework NCNN, and the night vision algorithm is implemented and deployed using C++. When compiled, the Native layer generates a shared library for the Java layer to call. In the Java layer, the deep learning algorithm and night vision algorithm in the Native layer are called, and the interface rendering and voice broadcasting functions of the app are realized. The design structure of the overall software is shown in Fig.3.

Fig.3 Schematic diagram of software overall code design



3.2. Overall structure

The overall structure of the software is to process the data of the mobile phone camera as a linear data stream. First, the software decodes the video captured by the mobile phone camera to obtain the RGB map. If the operating environment of the software is really daytime, it is not necessary to turn on the night vision algorithm, if the software is running at night or in a poor light environment, it is necessary to turn on the night vision algorithm. After the RGB image decoded by the camera is enhanced by the night vision algorithm, the next step is processed by the deep learning object detection algorithm to detect whether there is a humanoid target. According to the results of detection, the relevant vibration alarm and voice broadcast are carried out. The following describes the structure and flow of each processing module.

4. Correlation module

4.1. Image acquisition module

The schematic diagram of image acquisition module is shown in Fig.4. Here, the software uses the methods in the NCNN library to decode the image and convert the color space. This module directly reads the video stream collected by the camera of the mobile phone and outputs the image in RGB format.

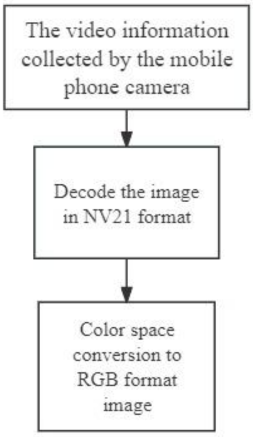


Fig.4 Structure diagram of image acquisition module

4.2. Night vision enhancement module

The night vision enhancement module is shown in Fig.5. The input of this module is the RGB format image output by the previous module, and the output image of this module is the image after illuminance enhancement and is used as the input image of the next module.

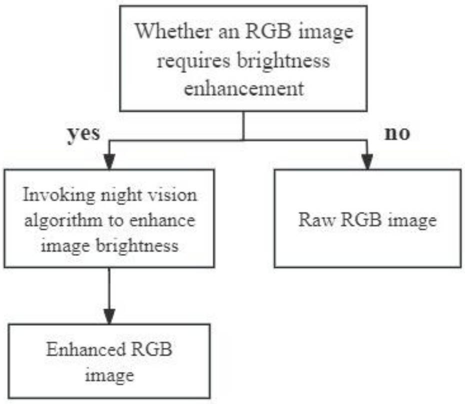


Fig.5Night vision enhancement module structure diagram

4.3. Humanoid target detection module

The humanoid target detection module[3] is shown in Fig.6. The input of this module is the RGB format image output of the previous module, and the output of this module is the detection result of the image. In this module, the NCNN library deployed the deep learning models NanoDet-RepVGG (input resolution 416*416) and NanoDet-EfficientLite (input resolution 512*512) for object detection. Users can choose which model to use according to the actual application scenario and phone configuration. Compared with traditional image processing algorithms, deep learning models can accurately identify humanoid objects from complex backgrounds, so they are suitable for home environments.

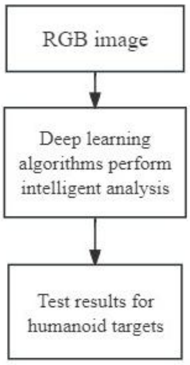


Fig.6 Humanoid object detection module

4.4. voice & vibration alarm module

The voice & vibration alarm module diagram is shown in Fig.7. This module is deployed in the Java layer and runs in a separate thread. The input of this module is the detection result of the deep learning model, and the detection result in the Native layer is retrieved through the data interface of the Java layer. If the result contains a humanoid target, the voice broadcast and mobile phone vibration alarm will be carried out. The voice broadcast content is: "Someone has entered", and the broadcast frequency is once every 2 seconds.

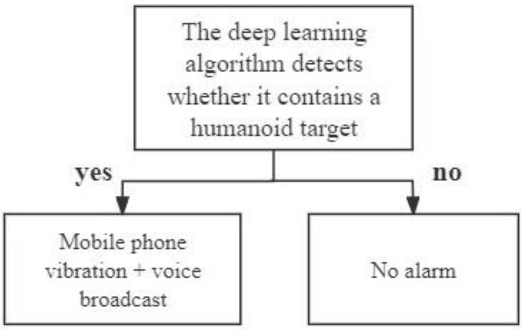


Fig.7 voice & vibration alarm module

5. Conclusion

"Deep Guard Dog" application uses smart phone technology and deep learning algorithms to achieve efficient and intelligent home security monitoring, solving the problem of illegal entry and incapacitated people into dangerous areas, more comprehensive and reliable than the traditional way, and will continue to optimize and improve user experience and identification accuracy in the future.

References

1. Yang Piao. Design and Implementation of Intelligent Video surveillance System in Security Field based on Deep Learning. *Materials for Information Records*, 2024, 25 (11):177-179.
2. Zou Peng, Chen Yuzhang, Chen Longbiao, et al. Light distribution prediction algorithm for night vision restoration based on neural networks. *Computer Science,* 2019, 46 (S2): 329-333+340. (in Chinese)
3. Lin Zhi-Bin, HUANG Zhi-Quan, YAN Lin-Ming. Research on Face detection based on Android platform. *Electronic Quality*, 2021, (11): 40-45.

Authors Introduction

Mr. Keming Chen

He is currently an undergraduate in Tianjin University of Science and Technology. He is majoring in automation. His research focuses on deep learning.He specializes in developing mobile phone software using C++ and Java.

Ms. Jiaxin Wang

She is currently an undergraduate in Tianjin University of Science and Technology. He is majoring in Robot engineering. Her research focuses on deep learning.She specializes in developing mobile phone software using C++ and Java.

Development of an Amphibious Surface Garbage Collection Robot and Its Applications

Yu Su *, Xin Wang, Long Shen, Zhenxing Liu, Xinrui Zhao, Xin Lin, Mengchen Huo, Yawen Qiao, Yan Zhang

College of Mechanical Engineering, Tianjin University of Science and Technology, 300222，China

*E-mail: * 1528140369@qq.com*

Abstract

This paper presents an amphibious water - surface garbage - collecting robot. It incorporates innovative technologies such as efficient garbage collection, accurate identification and classification, stable amphibious operation, and sustainable energy utilization. The double - four - bar linkage and "three - pipe" collection device ensure effective collection and classification. The amphibious crawler provides buoyancy and land - moving ability. Visual recognition technology has high accuracy. GPS automatic cruise and solar charging system are also included.

Keywords: amphibious, water - surface,GPS, remote control, monitoring

1. Introduction

With the acceleration of urbanization and the continuous expansion of human activities, it is more difficult to collect garbage in small and medium-sized waters. Algae reproduction, cigarette butts, plastic products discarded at will. In the long run, the ecological balance of small and medium-sized waters will be completely broken, and the biodiversity will continue to decline[1].

The investigation of the existing water surface garbage collection ships shows that large collection vessels occupy the majority and can quickly solve the situation of a large amount of garbage and water grass in the river. But it will take a lot of manpower and resources to clean up. In addition, large collection vessels are too large to enter small basins and closed basins, and require manual operation, and automation is not comprehensive enough. For small and medium-sized collection machines, most of the control accuracy is low, the garbage collection accuracy is not high, the efficiency is low, and the energy consumption is large.

Based on the above discussion, the focus of this design is the overall structure of small and medium-sized water surface garbage collection robot and the motion simulation of high-precision collection and low-energy design when the machine runs on the water surface under the GPS cruise state. The purpose of this design is to create a machine with amphibious, automatic garbage identification, sorting and positioning functions, so that small and medium-sized water garbage collection problems can be improved.

The rest of this paper is organized as follows. The second section introduces the structure of each part of the model, the third part introduces the design of visual recognition module, and the fourth part introduces the simulation of water surface operation to verify the effectiveness of the design. The fifth part summarizes the main content of this paper.

2. The mechanical structure

The water surface garbage collection robot is mainly responsible for garbage collection, sorting and automatic landing. Therefore, we use amphibious wheels, two sets of collection devices, one set of pull-back mechanism.

Amphibious wheel, using a specific foam material and crawler composite design. Foaming material can provide buoyancy and auxiliary power in water, and does not affect the land walking, has a strong amphibious mobility performance. The design of the amphibious wheel is shown in Fig.1.

Fig.1 The design of the amphibious wheel

2.1. Execution module

The execution module is mainly a push and pull double four-link mechanism, referring to the design of the packaging machine bag mechanism. Through visual identification, double four rod mechanism mainly carries out the collection of plastic products for recycling. When action, the motor drives the original parts to rotate. When the original driver rotates clockwise, the FEI rod moves to the right, the CBI rod moves to the left and the mechanism opens; when the original member rotates counterclockwise, the FEI rod moves to the left and the

CBI rod moves to the right to close.The mechanism schematic diagram is shown in the Fig.2.

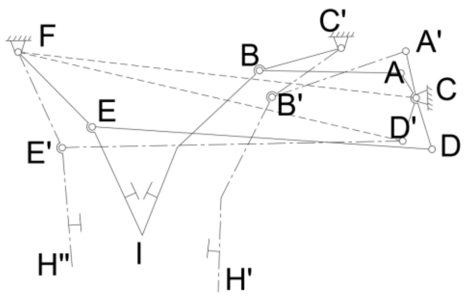


Fig. 2. Brief diagram of the double four-bar mechanism

2.2. Gathering unit

The design adopts dual-mode switching and three-tube collection. For plastic products, supplemented by visual identification, double four-bar mechanism, into the middle pipe, sent to the tail of the machine collection net; For floating debris, algae, etc., the suction method is used to enter the two side tubes and enter the collection bin inside the machine.The collection device is shown in Fig.3.

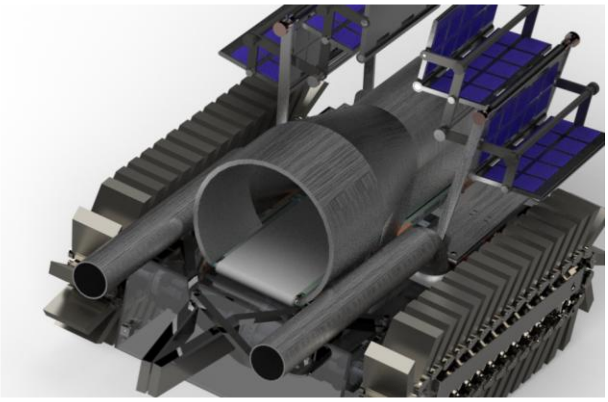


Fig.3. Collection device

3. Visual Recognition Module

In the design of the visual recognition module, we collected pictures of various types of trash and labeled them one by one, which were then imported into the model training as a dataset. The model was trained on the YOLOv5s and then compared with the validation set for model recognition. The visual recognition module is shown in Fig.4.

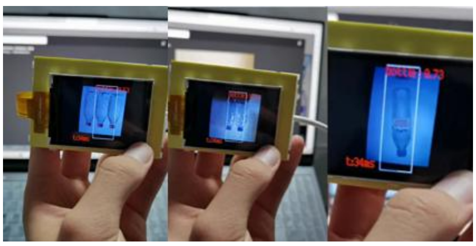


Fig.4. Visual recognition module

Sample identification data sheet is shown in Table 1.

Table 1. Sample identification data sheet

Sample name	Sample set	Successful number	success rate
Plastic	1000	963	96.3%
Cigarette butts	1000	951	95.1%
algae	1000	950	95.0%
leaves	1000	977	97.7%

4. Surface Operation Simulation

The model building process involves setting various parameters and dynamically adjusting the range of variables, referring to the setting and adjustment of the multi-agent system simulation platform to simulate the system performance under different working conditions [2]. At the same time, considering the CFD model of amphibious vehicles [3], the simplification of body shape follows certain principles (such as ignoring small scale structure, etc.) to adapt to the numerical simulation, in order to realize effective model construction and analysis in complex practical situations.

Model building and simulation are performed by using MATELAB software.

4.1. Mechanical simulation model

There are two main factors affecting mechanics. The first is buoyancy and weight on the surface of the water: the more objects collected, the greater the weight, the greater the buoyancy, and the mass of different objects is different. The second water resistance: related to buoyancy, the greater the buoyancy, the greater the resistance.

In some studies, the turbulent free surface flow of viscous liquid [5] or the influence of ship speed change on ship speed loss during deceleration were considered in the calculation of drag [6]. By comparing various optimization models, this paper approximately simplifies the resistance relation to a linear relation and sets the linear coefficient as K_x.At the same time,let the buoyancy F_1 the resistance F_2, the water uncertainty resistance F_x, the weight of the robot itself is G_0, and the inconsistent weight of each object is set to G_x.The formula can be obtained as follows.

$$\begin{cases} F_1 = K_1 \times (G_0 + n \times G_x) \\ F_2 = K_2 \times F_1 + F_x \end{cases} \quad (1)$$

- Buoyant simulation model: buoyancy F_1 is closely related to the weight of the robot itself G_0, the number of collected objects n and the weight of a single object G_x.The relationship is as follows.

$$F_1 = K_1 \times (G_0 + n \times G_x) \tag{2}$$

In the simulation, let $K_1 = 1$, $G_0 = 10$, and n values range from 0 to 50. In order to more fit to the actual working condition, with the increase of n the G_x value range is dynamically adjusted.

$$\begin{cases} \min\{G_x\} = 0.2 + 0.02 \times n \\ \max\{G_x\} = 1 + 0.1 \times n \end{cases} \tag{3}$$

For each calculation, G_x randomly takes values in this interval. At the same time, the robot submerged threshold was set to 100, and when $(G_0 + n \times G_x) > 100$, F_1 kept the former value unchanged. The F_1 value of each n was calculated to construct the buoyancy change curve with the number of collected objects.

- Resistance simulation model: resistance F_2 is affected by the buoyancy F_1 and the uncertain resistance F_x, the relation $F_2 = K_2 \times F_1 + F_x$, take $K_2 = 0.5$. When calculating F_2, G_x is determined within the range of n, and a random number of F_x (mean of 0, standard deviation of 5) is randomly generated, which obtains the corresponding F_2 value of each n and draws the change curve of resistance with the number of collected objects.

Simulation model of speed and energy relationship: speed v is closely related to energy E and resistance F_2, the expression is $v = K_4 \times (K_5 \times E - K_6 F_2)$, setting $K_4 = 0.1$, $K_5 = 2$, $K_6 = 0.2$. The grid n and E takes n to 50, E 0 to 100, F_1, F_2 and v values are calculated point by point, drawing the relationship surface between speed v and energy E and the number of collected objects n.

The simulation results of F_1, F_2 and n are shown in Fig.5.

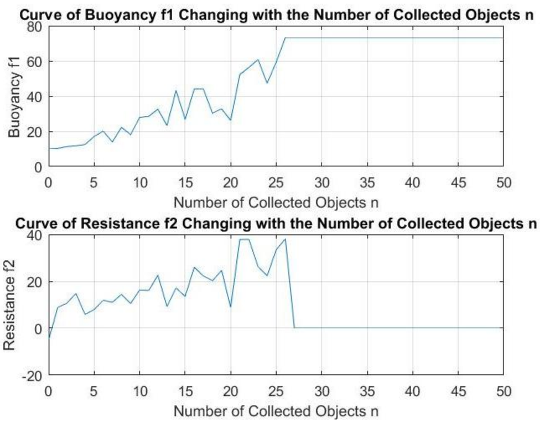


Fig.5. Simulation result diagram

Buoyancy characteristics: In the initial stage, n increases F_1 almost linearly, due to the increase of objects, and the buoyancy increases accordingly. When n exceeds a certain value and the total weight is close to the submerged threshold, the growth rate of F_1 slows down or even constant. This phenomenon indicates that the carrying capacity of the robot is limited and close to the critical state.

Resistance characteristics: The resistance F_2 curve shows that it increases with n rising and fluctuates significantly. This is attributed to $F_2 = K_2 \times F_1 + F_x$ increases the resistance base term with n growth, while the random fluctuation F_x increases the curve fluctuation, highlighting the complexity of the water surface environment on the robot motion resistance, indicating the direction for optimizing the robot shape and propulsion system design, to reduce the uncertainty resistance interference. The results are similar to the effect of various factors on resistance analysis of high-speed amphibious vehicle (HSAV) [4]. The resistance of HSAV includes friction resistance, viscous pressure resistance and wave resistance, and is affected by many factors such as vehicle shape and navigation speed. In this study, the resistance of water surface robot is affected by buoyancy and uncertain water surface resistance. The analysis of resistance characteristics can provide a basis for the optimal design.

4.2. GPS simulation model

The machine uses GPS module for positioning, and the positioning accuracy of GPS module affects the collection efficiency. Set the target point coordinates of the machine as (X, Y), the actual coordinate points of the machine at present as (X_1, Y_1), the GPS module output positioning coordinate points as, the expected turning Angle as (X_2, Y_2), the output heading Angle as θ_1, and the deviation Angle as θ_2,the formula can be obtained.

$$\begin{cases} \cos\theta_1 = \dfrac{X \cdot Y + X_1 \cdot Y_1}{\sqrt{X^2 + Y^2} \cdot \sqrt{X_1^2 + Y_1^2}} \\ \cos\theta_2 = \dfrac{X \cdot Y + X_2 \cdot Y_2}{\sqrt{X^2 + Y^2} \cdot \sqrt{X_2^2 + Y_2^2}} \\ \theta = |\theta_1 - \theta_2| \end{cases} \tag{4}$$

It can be seen from the formula that the smaller the absolute value of the deviation Angle, the higher the GPS positioning accuracy. The machine collection accuracy mainly depends on the GPS positioning accuracy. The positioning accuracy of GPS is not only related to its own control, but also positively related to its speed [7]. But at the same time, we should consider the system error, the fluctuation of water and the physical inertia impact of the double four-bar mechanism. Reducing the system error requires reducing the speed. Therefore, the relationship between accuracy and velocity obtained by curve fitting is shown in Fig.6.

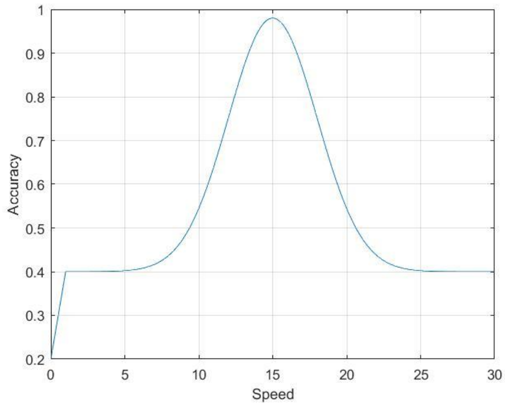


Fig.6. GPS fitted curve

4.3. Energy consumption simulation model

Energy consumption is related to speed, and the higher the resistance, the lower the speed. In addition, the speed is not constant, and the efficiency of collecting objects needs to be improved by accelerating or decelerating under certain conditions.

Simulation model of speed and energy relationship: speed v is closely related to energy E and resistance F_2, the expression is

$$v = K_4 \times (K_5 \times E - K_6 F_2) \quad (5)$$

Let $K_4 = 0.1$, $K_5 = 0.1$, $K_6 = 0.2$. The grid n and E takes n to 50, E 0 to 100, F_1, F_2 and v values are calculated point by point, drawing the relationship surface between speed v and energy E and the number of collected objects n.

The simulation results are are shown in Fig.7.

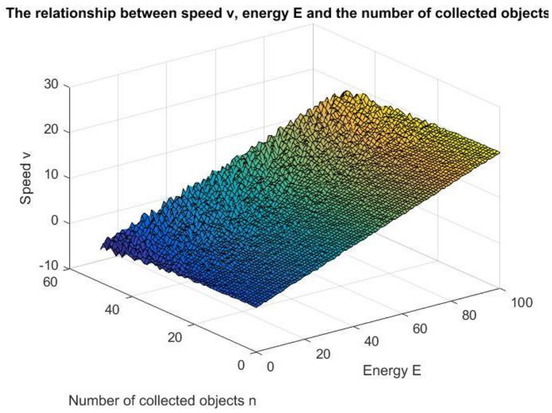


Fig.7. Simulation result diagram

Velocity-energy-object number relationship: the three-dimensional surface of velocity v intuitively presents its complex relationship with E and n. When n is fixed, E increases v increases, indicating that the energy supply can improve the speed and ensure the operation efficiency; E is constant, n increases v decreases, increasing the load of the robot, which is in line with the actual operation expectation.

4.4. Analysis and integration

In order to improve the accuracy of garbage collection and reduce the energy consumption under long working hours. The accuracy and energy consumption were simulated by MATELAB. Because the machine garbage collection is at full load, the inertia is maximum and the energy consumption is highest. In order to ensure more accurate simulation, it is considered to fit the machine under full load conditions.

When the speed is particularly large, considering the inertia and water fluctuation and other factors, the collection error will increase and the collection accuracy will be reduced. When the speed is particularly small, due to the very low speed, the gps accuracy will be lost , and the collection error will increase, so the entire collection accuracy can be approximately regarded as a normal distribution in the speed interval [7]. The gps accuracy curve can be divided into two sections, that is, it is a proportional function before the precise positioning speed threshold, and it is a constant beyond the speed threshold. After simulation by MATLAB, obtain Fig.8.

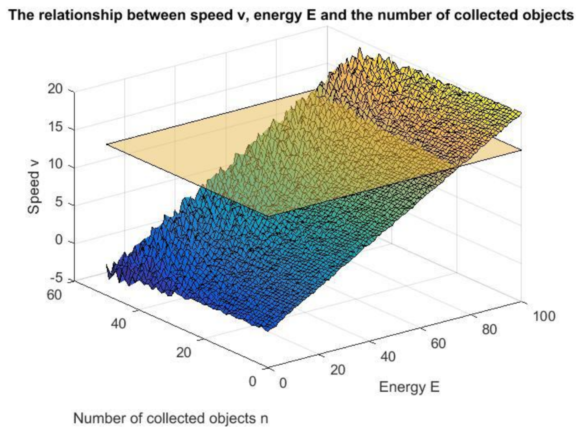


Fig.8.Speed threshold curve

5. Conclusion

Through the simulation of resistance buoyancy, GPS accuracy and energy consumption, it is found that in order to achieve high-precision collection and low-energy design of the machine, it is necessary to ensure that the value of the speed must be in a certain range, and the maximum threshold of the speed can be obtained under the full load state. Therefore, the speed is controlled within the interval, which can achieve high-precision collection while saving energy.

Acknowledgements

This part of the thesis is supported by the Innovation and Entrepreneurship Training Program (202410057142) of Tianjin University of Science and Technology in 2024, China.

References

1. Chang, H.-C., Hsu, Y.-L., Hung, S.-S., Ou, G.-R., Wu, J.-R., & Hsu, C. (2021). Autonomous Water Quality Monitoring and Water Surface Cleaning for Unmanned Surface Vehicle. Sensors, 21(4), 1102.
2. Chuang Zhang, Jichao Zhao, Fengzhi Dai. A Design of Multi-Agent System Simulation Platform Based on Unmanned Ground Vehicles and A Research on Formation Control Protocol. 2022 International Conference on Artificial Life and Robotics (ICAROB2022), January 20 to 23, 2022, pp. 808-812.
3. Xinmin Tian, Xiaochun Pan, Kai Yao, et al. Simulation of Amphibious Vehicle Water Resistance Based on FLUENT. International Conference on Materials Engineering and Information Technology Applications (MEITA 2015), January 2015, pp. 485 - 489.
4. Pan, Dibo, Xiaojun Xu, Bolong Liu. Influence of Flanks on Resistance Performance of High-Speed Amphibious Vehicle. Journal of Marine Science and Engineering, 2021, 9(11), pp. 1260-1269.
5. Kim, Mingyu, Olgun Hizir, Osman Turan, et al. A Study on Ship Speed Loss due to Added Resistance in a Seaway. Paper presented at the The 26th International Ocean and Polar Engineering Conference, Rhodes, Greece, June 2016. Paper Number: ISOPE-I-16-711.
6. Tabaczek, Tomasz, Jan Kulczyk, and Maciej Zawiślak. Analysis of Hull Resistance of Pushed Barges in Shallow Water. Polish Maritime Research, 2007.
7. Cao, Peter, et al. Application of GPS systems on a mobile robot. Intelligent Robots and Computer Vision XX: Algorithms, Techniques, and Active Vision, 2001.

Authors Introduction

Mr. Yu Su

He is currently studying at Tianjin University of Science and Technology, majoring in mechanical electronics as an undergraduate student. His research area is about mechanical design and manufacturing.

Mr. Xin Wang

He is currently studying at Tianjin University of Science and Technology, majoring in Communication Engineering as an undergraduate student. His research area is about communication technology and network engineering.

Mr. Long Shen

He is a junior student at Tianjin University of Science and Technology majoring in vehicle engineering, and his research field is new energy vehicles.

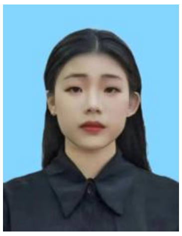
Mr. Zhenxing Liu

He studied in Tianjin University of Science and Technology, majoring in process equipment and control engineering. His research field is about mechanical design and manufacturing.

Miss. Xinrui Zhao

She is a junior student studying at Tianjin University of Science and Technology, her research field is about mechanical vehicle engineering.

Miss. Xin Lin



She is a senior student at Tianjin University of Science and Technology, majoring in mechanical manufacturing and its automation, and her research field is automated manufacturing.

Miss. Mengchen Huo

She is a senior majoring in mechanical and electronic engineering at Tianjin University of Science and Technology, and her research field is mechanical and electronic design.

Miss. Yawen Qiao

She is a sophomore student , studying at the College of Mechanical Engineering in Tianjin University of Science and Technology. Her field of study is about mechanical design, manufacture and automation .

Prof. Yan Zhang

She is a professor at the School of Mechanical Engineering, Tianjin University of Science and Technology. Her research interests include bionic mechanical design principles, biokinetic theory, and dynamic electricity integration technology

Design of an Intelligent Orbital Inspection Robot Based on Machine Vision and Ultrasonic Guided Waves

Xingwang Feng, Suqing Duan*

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

*Email: * SuqingDUAN@mail.tust.edu.cn*

Abstract

This paper introduces a wheel-foot switching track inspection robot based on machine vision and ultrasonic guided wave for enhancing the automation and intelligence of track inspection.The robot integrates the BeiDou positioning and autonomous traveling system, and combines image processing and ultrasonic guided wave technology to efficiently detect track cracks, settlements and other defects.The innovative wheel-foot switching structure and electro-hydraulic leveling platform improve the multi-terrain adaptability.The defect recognition algorithm based on particle swarm optimization algorithm and support vector machine, as well as the visual data processing platform, realizes the real-time transmission and analysis of inspection data.Tests show that the system has high inspection accuracy and stability, providing a reference for intelligent track inspection.

Keywords: Machine Vision, Rail Inspection, Ultrasonic Guided Wave, Defect Detection

1. Introduction

With the transformation of track inspection from informationization to intelligence, the inspection accuracy is improving, however, the development of autonomous motion, miniaturization and intelligent inspection instruments is slow and cannot meet the demand of intelligent development of rail transportation.Aiming at the above problems, the team members designed a wheel-foot switching track inspection robot, which can realize the application of multiple application scenarios, such as underground railroads, elevated light railways, ground tracks, mine tracks, etc. The product is equipped with multiple types of inspection equipment.The product is equipped with multiple types of inspection equipment, and the data processing platform realizes automatic processing and analysis of inspection data, improves the degree of automation, intelligence and informationization of track inspection, reduces the risk of personnel in hazardous scene inspection, and assists the development of intelligent track inspection industry.Model diagram of rail robot is shown in Fig.1.

Fig.1 Model diagram of rail robot

Intelligent track inspection robot based on machine vision and ultrasonic guided wave is a wheel-foot switching track inspection robot based on machine vision and ultrasonic guided wave intelligent detection, which is dedicated to detecting various defects such as cracks, settlement and unevenness of the track, and at the same time adapts to many kinds of complex terrains, with a wide operating range, and realizes intelligent full-automatic inspection.The track inspection robot applies Beidou positioning and automatic traveling, ultrasonic guided wave detection, image processing and other technical means to make the robot carry out inspection tasks according to the preset track traveling route, and utilizes the image processing technology and Beidou positioning system to accurately obtain the track image information and location information of the route, and then compares and identifies them through the track surface image enhancement algorithm of ACE and the data processing system, and transmits the location information of the inspection results to the rail transit system.The position information of the result is transmitted to the rail transportation platform.Meanwhile, ultrasonic guided wave equipment is added to carry out inspection, which can carry out high-precision inspection of internal defects of rails, determine the location, shape, size and other information of defects and cracks inside the rails, and solve the problems such as crack defects inside the rails which can not be detected by the machine's vision, so as to improve the accuracy of rail inspection.

Solving the speed inefficiency of manual inspection, leakage detection and misdiagnosis pressure low detection frequency is the primary goal of this product, through a more intelligent, efficient and safe way to replace manual labor, while saving manpower, material resources, financial resources, in order to achieve better operational results and economic benefits.Secondly, it is to strengthen the process of intelligent urban rail inspection and management, to build an intelligent urban rail inspection and management system by perfecting the integrated monitoring system, and to make the product walk in the forefront of the times by combining the modern 5G wave - Internet of Things and other advanced technologies.Through the continuous iteration of the product and the joint efforts of the company, the product has the advantages of high degree of integration, good compatibility and high applicability, and realizes the goals of high precision of track inspection, accurate on-site monitoring data, and good continuity and stability of system operation.

The rest of this article is organized as follows. The second section introduces the hardware components of the intelligent rail inspection robot, including the wheeled-leg switching structure, intelligent leveling platform, and the power system. The third section presents the defect detection system, which incorporates machine vision and ultrasonic guided wave technologies to identify and analyze rail defects. The fourth section discusses the innovative defect rating algorithm, which processes and evaluates defect data for effective classification and visualization. The fifth section describes the product's digital design, highlighting material selection, structural optimization, and system modeling. Finally, the sixth section provides the results of functional testing, including normal operating conditions and simulated fault scenarios, to validate the performance and reliability of the proposed system.

2. Hardware Part

2.1. Wheel-foot switching structure

The main body of the team adopts mechanical dog mechanical structure, which can realize precise control and movement, accurate positioning, strong stability, and be able to realize a variety of complex actions and movement modes, and the track inspection robot adopts quadrupedal, five-degree-of-freedom design and wheel-foot switching model, which can adjust the leg posture and gait according to the needs of the terrain conditions and environmental changes in which it is located, and realize smooth walking and and, satisfy the multi-topography movement, with a very strongAdaptability.It can solve the problem that conventional inspection robots need to be manually carried to the rails for inspection.Physical diagram of track inspection robot with wheel and foot switching is shown in Fig.2.

Fig.2 Physical diagram of track inspection robot with wheel and foot switching

Placing the robot in a variety of complex external environments, as the end rail wheel can be freely switched to paving and iron wheel movement modes for walking, running, jumping and other tests, the leveling platform can realize a variety of complex actions and movement modes, and it can be used to walk on the complex railroad roads, so that the robot has a good function of obstacle-crossing.

2.2. Intelligent leveling platform

In the field of high-precision measurement, it is crucial to ensure that the measurement equipment is in a stable state, which directly affects the accuracy of the data.At present, the traditional leveling method has the problem of low efficiency.The electro-hydraulic leveling platform developed in this project is designed to generate corresponding movements by driving the connecting rods connected to the electro-hydraulic rods through the telescopic movements of the electro-hydraulic rods.This design utilizes a notch fit mechanism between the rods to convert the horizontal movement of the electro-hydraulic rods into vertical adjustments in the X and Y axes.In addition, the platform incorporates a PID control algorithm to provide fast and intelligent adjustment of the leveling module in the robot system.

Key components of the system include four miniature gyroscopes on the auto-leveling base plate that automatically monitor and evaluate the platform's flatness.These gyroscopes compare the attitude information collected with the ideal level and quickly relay the information to the microcontroller via the data transfer module.The microcontroller then controls the rotational speed of the gimbal servos to adjust the attitude of the instrumented gimbal to achieve precise leveling.Model of leveling platform is shown in Fig.3. Engineering diagram of leveling platform is shown in Fig.4.

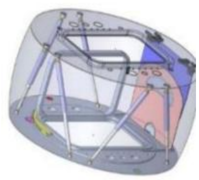


Fig.3 Model of leveling platform

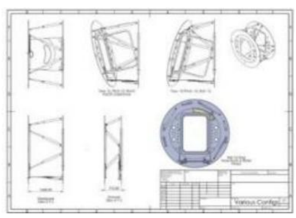


Fig.4 Engineering diagram of leveling platform

2.3. Power system

The intelligent track inspection robot based on machine vision and ultrasonic guided wave is powered by 4S lithium battery, which has higher energy density and can provide longer endurance time.It has a low self-discharge rate and can maintain a high state of charge when stored for a long time.With good charging and discharging performance, it supports fast charging and high-speed discharging, which improves the efficiency of the device.With proper use and charging, the lithium battery has a long service life, which can reduce the frequency of battery replacement and lower maintenance costs.

In a variety of complex external environments, the test track inspection robot in the foot structure speed can be up to 0.5m/s, on the track after the wheel foot switch to the wheel structure for detection, detection speed up to 20km/h. Robot built-in Li-ion battery pack + magnetic coupling resonance type without automatic charging, the efficiency of transmitting and receiving device distance of 0.3m is as high as 95%, compared with traditional wireless charging equipment, has a Compared with the traditional wireless charging equipment, it has the advantages of long transmission distance, high transmission efficiency and high transmission power.In practical applications, the use of lithium phosphate batteries, charging 5h can be filled, in normal operation can work continuously for 8 hours, range 140km, compared with the traditional artificial detection, greatly improving the detection speed and unit detection mileage.[1]

3. Defect Detection System

3.1. Internal and external recognition inspection system

The track inspection robot utilizes advanced machine vision technology to perform comprehensive crack detection, deformation analysis and other defect and abnormality identification on the track surface.During the inspection process, ultrasonic waveguide technology is used to identify potential defects in the track and ultrasonic signals are collected to determine the exact location of the defects in the axial and circumferential directions of the track.Subsequently, the robot collects a structured light image of the detected defect area using machine vision technology to form a rectangular image.

Through image processing technology, the robot is able to extract the centerline of the light bar in the structured light image, and use the powerful function of MATLAB software to realize the three-dimensional visualization of the local area of the track.By combining the ultrasonic guided wave inspection results and full-focus imaging technology, the robot is able to accurately locate the defects on the track and comprehensively assess the severity of the defects. Polar coordinates for circumferential localization is shown in Fig.5.

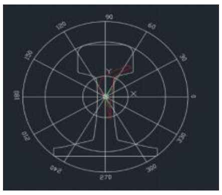


Fig.5 Polar coordinates for circumferential localization

The image processing system built by the team realizes accurate track image processing. In order to retain the characteristic information of the inspection items completely, the team chooses 24-bit RGB model color camera as the image acquisition equipment for the product, meanwhile, since the target surface of the CCD camera is a real light-sensitive component, there is no algorithm to fill in the simulation of the color situation, which can reduce the processing time and retain the contrasting characteristics of the original data, as well as carry out the image model conversion to convert the three-dimensional data into the image model.image model conversion, the three-dimensional data information is proposed to become single-channel data.[2] Rail seam image model conversion is shown in Fig.6.

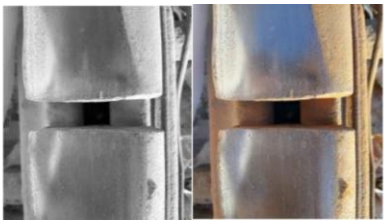


Fig.6 Rail seam image model conversion

When the intelligent rail inspection robot based on machine vision and ultrasonic guided wave acquires image data, it encounters a variety of factors that cause the image quality to be impaired, such as fluctuations in the detection speed, vibrations during operation, and changes in the ambient light, which may introduce noise points on the image.In order to improve this problem, an adaptive weighted median filtering technique is used to denoise the image, effectively eliminating the disturbing noise.By adjusting the size of the filtering window, different processing effects can be obtained, which also provides a strong support for accurate identification of track defects.[3] Improved weighted median filtering track image processing results is shown in Fig.7.

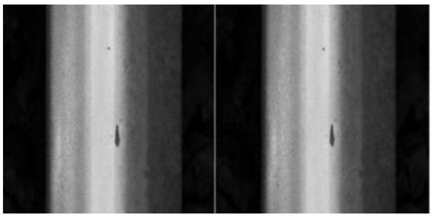


Fig.7 Improved weighted median filtering track image processing results

3.2. Roadbed settlement detection

The vibration characteristics of the vehicle will change significantly under the condition of roadbed settlement.Based on this phenomenon, the vertical coupling model of vehicle - track - roadbed is constructed.The model shows that when the roadbed settles, the track structure will have corresponding deformation under the influence of its own gravity, which will cause the track surface to be uneven.Through the application of this model, the influence of roadbed settlement on track and train operation can be predicted, which provides scientific basis for the maintenance and management of railroad engineering. Railroad bed structure model is shown in Fig.8.

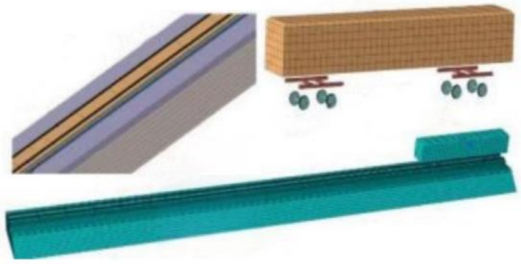


Fig.8 Railroad bed structure model

By analyzing the vibration response curve of intelligent rail inspection robot based on machine vision and ultrasonic guided wave under different settlement conditions, the uneven settlement of roadbed can be effectively identified.This method not only improves the feasibility of monitoring, but also can visualize the impact of settlement on train operation by extracting and analyzing the vibration data. Vehicle vertical acceleration time course curve under different settlement amplitude is shown in Fig.9.

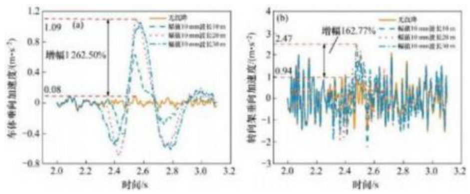


Fig.9 Vehicle vertical acceleration time course curve under different settlement amplitude

According to the analysis of the data presented in the graph, with the increase of settlement amplitude from 0mm to 30mm, the maximum vertical acceleration of the vehicle body, bogie and wheelset when passing through the settlement area increased significantly by 754.55%, 234.02% and 16.73%, respectively.[4] This phenomenon indicates that the vertical acceleration of the car body and bogie has a high sensitivity to the degree of roadbed settlement, which provides an important basis for judging whether there is uneven settlement of the roadbed.In classifying and identifying the track roadbed settlement, an identification method based on particle swarm optimization algorithm (PSO) and support vector machine (SVM) is used.This method improves the performance of the identification model by optimizing the algorithm, and then realizes the accurate identification of different types of roadbed settlement. PSO-SVM identification process of roadbed settlement is shown in Fig.10.

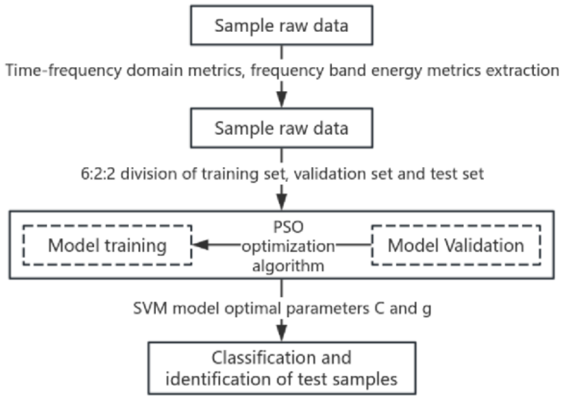


Fig.10 PSO-SVM identification process of roadbed settlement

After collecting the time series data of vehicle vertical acceleration under different working conditions, we performed feature extraction for each set of data, including feature indicators such as time domain and frequency band energy.These feature metrics constitute the sample set, where X represents the feature metrics and Y represents the corresponding working condition labels.The sample set is divided into training set, validation set and test set in the ratio of 6:2:2.

Using the particle swarm optimization algorithm, we trained the model on the training and validation sets to find the optimal key parameters C and g, in order to construct an efficient ballasted track subgrade settlement identification model.After several rounds of training and parameter optimization, we obtained a classification model with the best performance.After inputting the test set into this model, according to the output results of the model and the corresponding labels, we can accurately classify and recognize different types of roadbed settlement conditions. Recognition rate of settlement wavelength/amplitude based on PSO-SVM is shown in Fig.11.

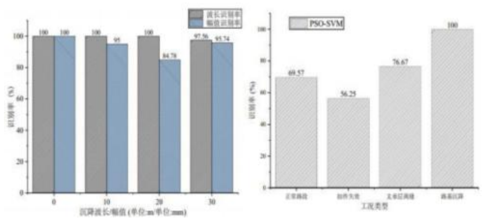


Fig.11 Recognition rate of settlement wavelength/amplitude based on PSO-SVM

The recognition accuracy reaches 95% and 84.78% when the settlement amplitude is 10mm and 20mm, respectively.This shows that the algorithm is able to maintain a high accuracy in the case of small settling amplitude, although the recognition ability of the algorithm decreases.[5] For the case of no settling and different settling wavelengths, the algorithm's recognition accuracy is almost perfect.These results show that particle swarm optimization combined with support vector machines is effective in identifying uneven settlement of roadbeds.With these data, we can conclude that the algorithm shows high reliability in dealing with the settlement problem, especially when the settlement magnitude is large.The algorithm is still able to provide reliable identification results even when the settlement magnitude is small.

4. Innovative rating Algorithm

The visualization data processing platform receives and processes a large amount of raw data, adopts independent innovative algorithms and automatically analyzes and rates the rail crack data, and after the defect information is processed, the platform shows the defect data in the form of charts, maps, etc. Firstly, the rail defect detection based on ultrasonic guided wave and machine vision is carried out.

First of all, the platform will be based on ultrasonic guided wave and machine vision rail defect detection, through the ultrasonic guided wave detection of rail defects, ultrasonic guided wave signals are collected, the axial coordinates of the defects and axial positioning polar coordinates, the location of the defects are determined, and the advanced visualization equipment is used for three-dimensional visualization, and then the location of the rail defects are located, and the degree of defects is assessed by fusion of the rail defects.Using defect classification algorithm, image processing, analysis, real-time, non-contact points can be very good luck for rail detection.

Team members constructed a comprehensive database of rail defect data collected through field research.Using this database, we trained and developed an algorithmic model.On the PyQt5 software development platform, we encoded images captured by ultrasonic guided waves and machine vision techniques for data analysis of rail defects.The process first recognizes the ultrasound guided wave and machine vision images, extracts the features of the track defects, and identifies them, providing possible identification results and relevant attribute information.Through target identification and classification, we were able to determine the specific type of defect.In the actual test, we measured a defect with a circumferential length of 7 cm and calculated its circumferential angle to be approximately 59.44°. By further analyzing the polar plot, we determined the angle of the circumferential defect to be 61.4°, with an error of only 3.3%. Compared with the traditional detection method, the detection precision is improved by 2.3% and the accuracy is 95.2%. Defect database is shown in Fig.12.

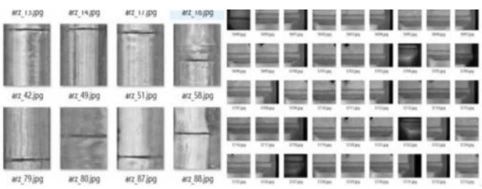


Fig.12 Defect database

The defect types obtained by the software processing platform are then rated by the team's self-developed rating algorithm based on the depth, length, width and other information of the defects to determine the severity of the defects and classify them for subsequent processing.The analysis uses the OpenCV library to read the images uploaded by the user and convert them into a processable format.The XML file is parsed using the ElementTree library to extract the category and location information for each object.Then, based on the information in the XML file, the system draws a rectangular box for each object on the image and displays its category label inside the box, and saves the annotated image to a specified folder for subsequent use or sharing. It has good scalability, which helps the user to cope with the annotation needs of different tasks. Innovative Defect Rating is shown in Fig.13.

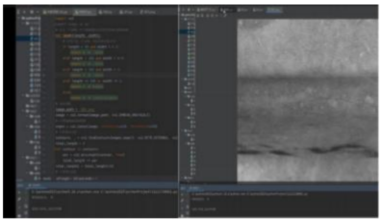


Fig.13 Innovative Defect Rating

The visualized data platform for railroad track crack detection provides a window for real-time information transmission between the robot and the track service personnel, transmits the inspection data to the data processing platform, and the defect rating algorithm of the platform efficiently receives and processes a large amount of raw data, including extracting and analyzing the defect location, the amount of settlement of the roadbed, the shape of the track cracks, the size, the density and other characteristics, and automatically analyzes and rates the track crack data.Rating.It provides users with high quality data support, helps users to quickly and accurately identify the type, size, location and other information of defects, provides important reference for subsequent maintenance and repair work, and helps to make maintenance plans and decisions in a timely manner.

5. Product Digital Design

5.1. Robot structure design

The team chose aluminum alloy, which has low density, high strength, excellent corrosion resistance and long service life, as the main component material of the product body.One-piece stamping technology is utilized to manufacture the various components of the body, a method that not only improves the efficiency of material use, but also enhances the strength and rigidity of the body.In addition, the surface of the aluminum alloy is treated with electrolysis to form an oxide film, which gives the car body excellent waterproof performance.The design of the vehicle body adopts a streamlined drum-shaped aluminum alloy structure, which significantly reduces the air resistance of the vehicle during driving, effectively improves the stability of the vehicle, reduces the risk of rollover, and enhances its wind resistance. This significantly improves the performance and durability of the product. Robot density analysis is shown in Fig.14.

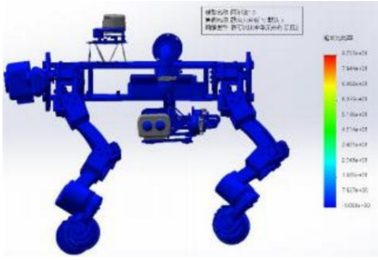
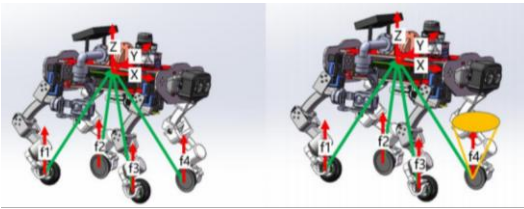


Fig.14 Robot density analysis

Pattern members digitally design the MPC controller for the center-of-mass motion analysis of the robotic dog. The purpose of the MPC controller is to allow the center of mass of the robotic dog to track the latest center-of-mass reference trajectory as much as possible, and then calculate the reaction force required at the end of the foot in the coming period of time, and ultimately generate the system state variables: the position, velocity, acceleration, rotation angle, and angular velocity of the robotic dog, and the system input variables: the quadrupedal touchdown force.The single rigid body model ignores the effect of the legs of the quadruped, and simplifies the motion analysis by considering the robot directly as a whole, i.e., a rigid body, with the ground force on the end of the leg acting directly and equivalently at the center of mass.[6] Dynamics modeling analysis is shown in Fig.15.



Fig,15 Dynamics modeling analysis

5.2. Hardware feasibility test

In order to ensure that the orbital inspection robot maintains a high and stable level of measurement during the measurement inspection, the team members tested and verified the structural aspects of the robot's machine.The overall stress situation was analyzed by SolidWorks software.During the simulation and verification process, the team members analyzed the stress on the main joints of the robot and optimized its structure to meet the transformation and motion strength.This not only ensures that the overall shape of the robot will not be deformed due to the force, but also meets a certain load capacity.Enhance the adaptability.SolidWorks Deformation Analysis of Major Components is shown in Fig.16.

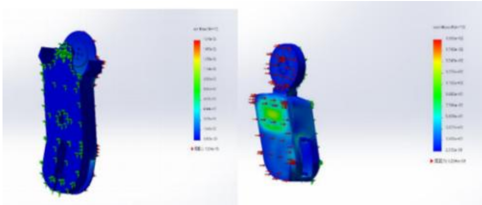


Fig.16 SolidWorks Deformation Analysis of Major Components

In the field of SolidWorks model quality digital design, the integrated application of simulation tools is a key step in achieving comprehensive evaluation and optimization of the design.These tools enable in-depth analysis of designs in a virtual environment, covering mechanical properties such as stress, strain, fatigue, and nonlinear response, as well as multiphysics field issues such as heat transfer and fluid dynamics.This analysis helps ensure the reliability and efficiency of the design in real-world applications.

Through exhaustive quality analysis, the design phase can anticipate potential structural defects, performance limitations, and the many problems that may be encountered during manufacturing and assembly.Avoiding costly modifications and rework at a later stage effectively reduces the overall cost and cycle time of product development.

Model quality analysis also supports interdisciplinary design optimization for comprehensive design solution evaluation.This approach not only accelerates the transition from concept to finished product, but also reduces the cost and time of development by reducing the reliance on physical prototypes and testing.Mass Density Analysis of Inspection Robot is shown in Fig.17.

Fig.17 Mass Density Analysis of Inspection Robot

5.3. Control analysis

Each mechanism must be precisely coordinated with each other and the force is reasonable in order to show the expected function, which requires high machining precision and coordination.And then combined with the knowledge of material mechanics on the limit strength, stiffness, stability analysis of these parts and consider the overall structure and weight of the car body and the comprehensive analysis of the applicable places. Simulation and optimization data of Adams optimized servo parameters is shown in Fig.18.

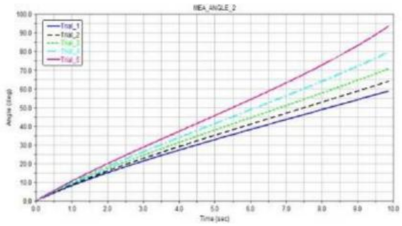


Fig.18 Simulation and optimization data of Adams optimized servo parameters

Based on the above foundation, the team analyzed its operation performance and continuously improved the structure to meet the motion performance. In order to meet the automation of unattended, multi-morphic robots, team members carried out circuit design and circuit simulation simulation, circuit feasibility and can fulfill our preset functions and operating standards.Body Circuit Design is shown in Fig.19.

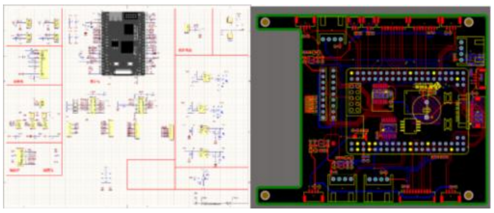


Fig.19 Body Circuit Design

6. Functional Testing

6.1. Track testing under normal operating conditions

● ***Preparation stage***

Before the test, the robot is subjected to a comprehensive system check and calibration to ensure that its sensors, cameras and navigation equipment are in the best working condition.The test route is also planned, and various track types such as straight sections, curved sections and turnout areas are selected as test objects.Robot completes testing preparation is shown in Fig.20.

Fig,20 Robot completes testing preparation

● ***Data acquisition***

Start the robot, drive along the predetermined route automatically, use the omni-directional laser scanner, side gimbaled surface array camera, ultrasonic guided wave and autonomous cruise obstacle avoidance navigation to collect the geometric size data and image information of the track synchronously, and transmit the collected information and data in real time to the back-end processing platform to carry out the preliminary pre-processing and screening of the data.

● ***Data analysis***

The back-end processing platform applies advanced image recognition and data analysis algorithms to analyze the collected data in depth, and automatically generates track condition assessment report, including geometric deviation, surface defect location and severity and other information.The evaluation results are compared with the railroad maintenance standards to identify the areas that need to be paid attention to or potential safety hidden dangers, and the above data are shown in the form of visualized charts and pictures. Detection Data Interface is shown in Fig.21.

Fig.21 Detection Data Interface

6.2. Defect detection under simulated failure scenario

● ***Scene setting***

Set up simulated fault points on the test route, such as artificially created track surface cracks, gauge expansion, unevenness, triangular pit fasteners missing, screws loose and other defects, to ensure that each simulated fault point is representative of a comprehensive test of the detection ability of the rail intelligent inspection robot.Simulated fault points is shown in Fig.22.

Fig.22 Simulated fault points

● ***Inspection process***

The robot drives according to the established route, scans and photographs the whole section of the railroad with high precision, and transmits the data to the back-end data processing platform in real time, processes and analyzes the images through the defect grading algorithm, and evaluates the degree of defects of the railroad in a fusion manner.[7] Railway Intelligent Identification and Detection is shown in Fig.23.

Fig.23 Railway Intelligent Identification and Detection

- ***Emergency response***

According to the grading standard of the innovative grading algorithm, the robot should classify the detected defects into different grades and trigger the corresponding early warning mechanism according to the logic program of the algorithm, and provide detailed location information and defect description as well as remedial measures to the railroad staff through the visualized data processing platform.The robot's emergency response speed and report accuracy are evaluated to verify its practicality and reliability in emergency situations. Defect Rating is shown in Fig.24.

Fig.24 Defect Rating

7. Conclusion

This paper presents an intelligent track inspection robot integrating advanced technologies such as machine vision, ultrasonic guided wave detection, and Beidou positioning. The system achieves high-precision rail defect detection and efficient data analysis, addressing the limitations of traditional methods. Its robust design and innovative defect rating algorithm ensure reliability and adaptability in complex environments. Functional testing confirms its effectiveness in enhancing automation and safety, contributing to the intelligent development of rail inspection.

References

1. Zhang Hongwei. Automatic inspection method for surface defects of ballasted rail track based on visual characteristics [J]. Automation Application, 2023, 64(15):144-146.
2. Zhang Xiaoming, Li Zhi, Wang Xianhui, Yu Jiangong. The fractional Kelvin-Voigt model for circumferential guided waves in a viscoelastic FGM hollow cylinder. Applied Mathematical Modeling[J]. 2021(89) :299-313.
3. Ren Juanjuan, Zhang Yizhi, Liu Wei, et al. Recognition of ballasted track subgrade settlement damage based on PSO-SVM[J]. Journal of Railway Science and Engineering, 2023, 20 (11): 4400-4411.
4. Gao Jialin, Research on quantitative detection method of rail fastener defects based on deep convolutional network [D]. Beijing University of Architecture, 2023.
5. Wang Yuanping, Zha Mingan, Dai Qiang, et al. Development and application of intelligent equipment inspection system [J]. Information Record Material, 2019, 20(04):105-107.
6. Lin Qiao, Xin Zhulin, Kong Lingbo, et al. Development status of China's chili pepper industry and breeding countermeasures[J]. Journal of China Agricultural University, 2023, 28 (05): 82-95.
7. Feng Dan, Zhang Haiyan, Yu Yongling, et al. Propagation characteristics of guided waves in railroad tracks with cracks[C] Acoustical Society of Sichuan Province, Physics and Ultrasonic Electronics Division of Acoustical Society of China, Shanghai Acoust.

Authors Introduction

Mr. Xingwang Feng

He is currently pursuing his undergraduate degree at the School of Electronic Information and Automation, Tianjin University of Science and Technology. His research field is embedded system.

Ms. Suqing Duan

She is currently pursuing her undergraduate degree at the School of Electronic Information and Automation, Tianjin University of Science and Technology. Her research field is embedded system.

Intelligent Temperature Control System for Chip Soldering Station

Huahao Li [1] *, Junjin Chen [2], Ce Bian [3], Mengfan Zhang [3]
[1]*College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China*
[2] *SMC (Beijing) Manufacturing Co., LTD, China*
[3] *Tianjin Tianke Intelligent Manufacture Technology Co., LTD, China*

*E-mail: *1983618432@qq.com*

Abstract

Chip is the general name of semiconductor component products, mainly by the semiconductor material, solid state electronic devices, silicon wafers and other materials processed through a number of responsible processes. According to the functional requirements, the intelligent temperature control system of the chip welding bench is designed with PLC as the control core. In the PLC system design using PID instructions, when the temperature is close to the specified temperature using low-power heating, when the temperature difference is large high-power heating. In this paper, we design the temperature control system of chip soldering bench with Siemens S7-200 PLC as the control core.

Keywords: MCGS, Temperature Control, Siemens S7-200 PLC, PID

1. Introduction

Due to the needs of industrial process control, higher and higher requirements have been placed on automation equipment. With the miniaturization, integration and functional diversification of electronic components, it makes the integrated circuit (IC) manufacturing process increasingly complex. And the density of IC packages is increasing, thus making the problem of thermal deformation of chip solder joints particularly prominent [1]. However, with the development of science and technology and the application of computer technology, the research technology of temperature control system has also risen rapidly. To ensure the highest efficiency and production quality of the equipment, and has been widely used in various fields [2].

The design of this temperature control incorporates the practical requirements of a patch soldering workstation. S7-200PLC was utilized to complete the study of the temperature control system: temperature acquisition, operation and analysis. This controller is designed primarily for temperature control devices for patch soldering stations in modern industrial automated production processes. The use of PID control to replace the more traditional temperature control methods, so that the temperature control can be utilized more effectively. Finally, through the configuration monitoring screen for simulation and auxiliary display [3].

Information collection and study on the process flow of chip soldering station temperature control system. Make the overall program of the temperature control system for the chip soldering station. Utilize PLC-200 model to complete the control of the system. The ladder diagram temperature control program is written for PLC, and PID algorithm is adopted to complete the temperature control.
Use MCGS configuration software to simulate the monitoring screen of the temperature control system of the chip soldering station to monitor and simulate the real-time temperature.

The rest of this article is organized as follows. The second section introduces the main research techniques. In the third part, a description of the control algorithm is presented. The fourth section describes the design of the screen. The fifth part summarizes the main content of this paper.

2. Introduction to Research Techniques

Intelligent temperature control system for chip soldering station adopts S7-200 PLC, CPU model 224, and extends EM231 module. For the study of MCGS software, the establishment of a complete chip soldering station intelligent temperature control system simulation. The main system design is as follows:

Programmable Logic Controller (PLC) is a kind of industrial digital operating system. PLC is characterized by small size, strong anti-interference ability, simple and flexible programming, easy to expand and easy to network with computers. In the field of automatic control has been widely used. PLC is structurally very similar to the general control machine. Its main features are good versatility, strong anti-interference ability, easy maintenance, short debugging cycle and so on.

The PLC main unit can be matched with individual I/0 expansion units. And each extended unit has its own use function and purpose, and connected to the host computer through the bus. Each expansion unit can be expanded or used in combination as needed to meet the performance requirements of controllers for different occasions. The basic structure of PLC is shown in the figure Fig.1 below.

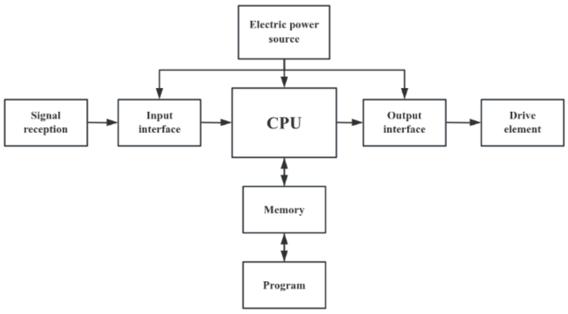


Fig.1 PLC Infrastructure Diagram

Configuration software provides testers with visual monitoring screens for real-time site monitoring. In MCGS configuration software, the monitoring screen can be created conveniently through the graphic editing function. Various controls in the screen are graphically displayed, and the control devices are animated with alarm windows, historical curves, real-time trend curves, reports and other functions. The composition of the MCGS configuration system is shown in Fig.2 below.

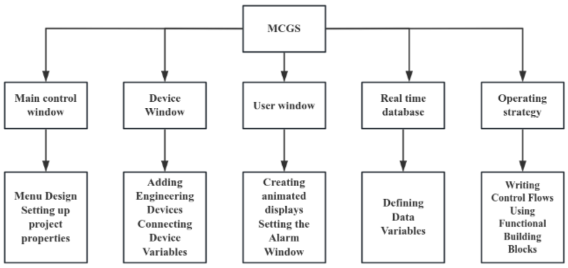


Fig.2. MCGS Structure Diagram

In the selection of the module EM231 is used, which has high accuracy. Since the PLC can only receive digital signals, this module is required to convert the detected temperature analog signals to digital signals. To create an analog input channel, an EM231 module equipped with 4 analog input channels is used to complete this design.

3. Description of the Control Algorithm

PID control is a relatively simple and convenient method, and has been widely used in industrial production control systems to achieve precise control of the production process on site [4]. The main PID descriptions are as follows:

PID program control module: this controller uses a new type of fuzzy control technology for control, which can automatically adjust the operating state of the controlled object according to the objective function and constraints set by the system.

PID function instruction: the combination of analog input module and analog output module can do and PID process control similar effect, compared with the lower cost, more affordable.

PID closed-loop control of their own programming: due to the lack of functional instructions and process modules, you can use their own throughout some of the program to control.

First the PID controller is designed, then the PID controller in digitized form is rewritten into PID equations in discrete form and finally the program is written.

In a closed loop control system with typical PID, Sp(t) is the input quantity for a given value, Pv(t) is the feedback quantity of the system and C(t) is the output quantity of the system. This is shown in Fig.3 below.

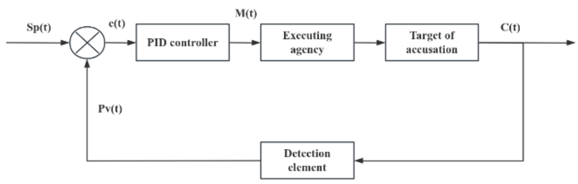


Fig.3. Typical PID closed-loop control system

Adjustment of proportional control: We should change the proportionality coefficient, slowly from small to large. At the same time, observe the corresponding system response until we get a curve with fast response and small overshoot.

Trimming the integration link: Reduce the proportionality coefficient of the first screening step to 50% to 80%, and then set the value of the integration time to a relatively large value. Observe the corresponding response curve.

Under the condition that this condition remains unchanged, the experimental study is carried out by changing different variables, the integration time is reduced, the integration effect is increased, and then the corresponding proportionality coefficient is adjusted by a small margin. After many repeated experiments and adjustments, a relatively ideal ratio parameter and a more ideal integration parameter are finally achieved.

Adjustment of the differential link: For the differential time, first set it to zero, and then slowly increase it. Several experiments and adjust the corresponding proportionality coefficient and integration time, to find the most compatible with the experimental effect of the PID control parameters.

4. Design of Configuration Screen

The system of this design is composed of S7-200PL and CMCGS configuration software, which can realize the following functions:

System control process: create relevant configuration screen in the system, which can clearly and intuitively observe the whole process of temperature control.

Temperature real-time monitoring: temperature changes during system operation are displayed and detected in real time through curves. System Fault Alarm: If the system temperature is too high or too low, the temperature high limit alarm and low limit alarm will be carried out.

When the system is running and working, it can monitor the process production situation and modify the control parameters in real time. When there is a fault or unexpected situation in the system, it can receive alarm signals and deal with it in time. The main screen monitoring diagram is as follows Fig.4 below.

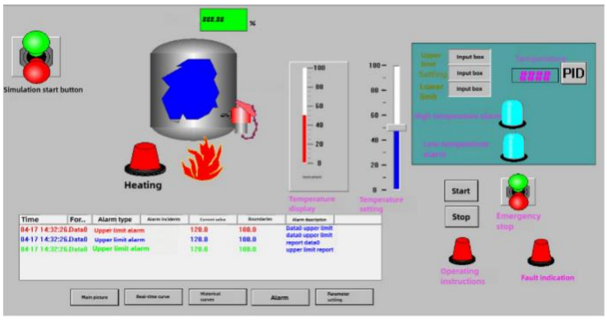


Fig.4. Main screen monitoring diagram

The process of animation screen monitoring display can be updated in real time the state of the temperature, so that the process and temperature control process more clearly show.

Starting and stopping the temperature control system of the chip soldering station through MCGS configuration software, you can view the current temperature, trend curve, alarm data and modify the PID parameters. When the temperature of the chip soldering station is higher or lower than the set alarm temperature, the fault light on the screen will be lit, and the alarm window will display the alarm data to facilitate processing. Configuration of the parameters of the temperature system is shown in Fig.5.

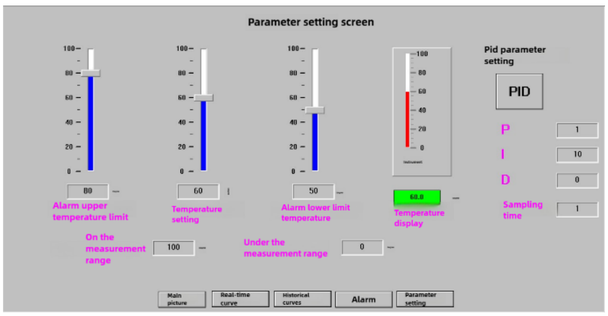


Fig.5. Parameterization of the temperature system

Open the MCGS configuration software and open the Patch Welding Station Intelligent Temperature Control System project monitoring interface. Entering the operation mode, you can display the temperature and status of the Patch Welding Station and the real-time and historical curves of the temperature in the monitoring screen. We can set the temperature according to the process requirements in the system control interface. PID temperature adjustment for the SMD station, its working status will change and the corresponding indicator light will be on. Fault Alarm Display: If the system temperature is too high or too low, the temperature upper limit alarm and temperature lower limit alarm will be carried out. The historical trend graph is shown in Figure Fig.6.

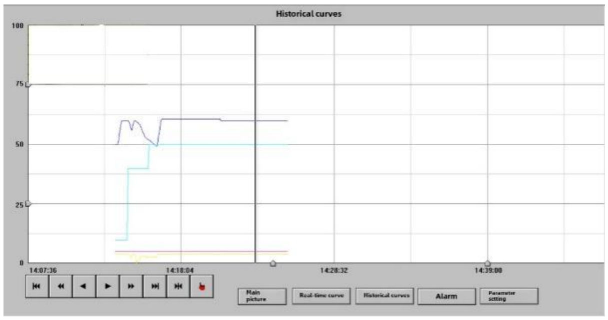


Fig.6. Historical Trend Graph

5. Conclusion

A design that combines temperature control with the practical needs of a chip soldering station. The use of S7-200PLC as the core control to complete the temperature control system. The controller is mainly for the modern industrial automation production process in the chip welding station temperature control device and design, the use of PID control instead of the more traditional temperature control. This can make the temperature control can be used more effectively. Finally, through the configuration monitoring screen for simulation and auxiliary display.

References

1. L. Zhou, M. Mo, Z. Dong, et al., "Design of Intelligent Monitoring System for Greenhouses Based on PLC and Landscape King," *2024 International Symposium on Intelligent Robotics and Systems (ISOIRS)*, Changsha, China, 2024, pp. 170-174.
2. H. Zhou, Y. Liu, G. Li, et al., "Research on PLC Controlled Refrigeration Equipment under Intelligent Electrical Control System," *2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI)*, Changchun, China, 2024, pp. 1470-1474.
3. S. Li, X. Li and Z. Yang, "Research on Temperature and Light Control System of Vertical Farm Based on PLC," *2023 International Conference on Networking, Informatics and Computing (ICNETIC)*, Palermo, Italy, 2023, pp. 320-324.
4. E. E. C. Ames, B. J. L. Quilca, E. N. U. Taipe, et al." Control and Monitoring of Thermographic Chambers by Means of PLC and HMI," *2022 2nd International Conference on Robotics, Automation and Artificial Intelligence (RAAI)*, Singapore, Singapore, 2022, pp. 272-276.

Authors Introduction

Mr. Huahao Li

He is a second-year master candidate in Tianjin University of Science and Technology, majoring in Intelligent multibody robot.

Mr. Junjin Chen

He received his bachelor's degree in Mechanical Design, Manufacturing and Automation from Beijing Union University in 2014. His research interests are industrial automation and robotics. From 2005 to 2007, he worked at Beijing Aeronautical Manufacturing Engineering Research Institute. Since September 2007, he has been working at SMC (Beijing) Manufacturing Co., LTD., China.

Mr. Ce Bian

He received his master's degree in control engineering from Tianjin University of Science and Technology in 2019, focusing on machine vision technology. Since 2017, he has been working in Tianjin Tianke Intelligent Manufacture Technology Co., LTD., serving as the general manager.

Ms. Mengfan Zhang

She received her bachelor's degree in financial management from Zhengzhou College of Commerce in 2019. She specializes in machine vision technology. Since 2019, she has been working as a visual engineer in Tianjin Tianke Intelligent Manufacture Technology Co., LTD., China.

A Review of Object Detection Techniques Applied to Pest Images

Hongshuo Zhai [1*], Fengzhi Dai [1], Lijiang Zhang [2], Qiang Wang [3]

[1] Tianjin University of Science and Technology, Tianjin,300222, China

[2] Xinjiang Shenhua Biotechnology Co., Ltd, Xinjiang, China;

[3] Easy Control Intelligent Technology (Tianjin) Co., Ltd., Tianjin, China

*E-mail: * 1936130557@qq.com*

Abstract

In agricultural information management, crop pest control has always been an important topic, and the image detection technology of small target pests is particularly critical in this process. At present, the technology faces challenges such as difficult data collection and insufficient robustness. This paper first introduces the development of object detection technology and its application in the field of agriculture, then analyzes the challenges of information-based pest control, discusses the research progress of pest dataset construction, image data augmentation technology and object detection algorithm, and finally points out the future research direction in this field.

Keywords: image detection technology, data collection, insufficient robustness, agriculture

1. Introduction

The application of target detection in the field of pest detection is very important. Since the acquisition of pest images can be affected by environmental conditions such as weather, light and occlusion, this increases the difficulty of detection. In addition, the diversity of surface textures and shapes of pests also poses challenges for detection. Even objects of the same class can vary from Angle to Angle and from individual to individual, adding to the complexity of detection1.

In the traditional methods, sliding window based search and artificial feature extraction face the challenge of efficiency and accuracy in practical applications. For example, the sliding window method requires traversing every possible position of the image, which is time-consuming and prone to creating a large number of redundant Windows. To overcome these problems, deep learning techniques have been introduced into the field of pest detection. Deep convolutional neural networks (CNNS) reduce manual intervention and improve the robustness of models by automatically learning features from large amounts of data. Deep learning models such as YOLO and SSD are trained end-to-end to make predictions directly from image pixels to target categories and bounding boxes, simplifying the detection process and increasing efficiency.

This paper mainly introduces the development of target detection in the field of agricultural pests, analyzes the existing problems and solutions, and finally introduces the improvement and lightweight of the model.

2. The Phase Detection Algorithm and Its Application in Agriculture

2.1. Two-stage deep learning algorithm

Two-stage deep learning algorithm plays an important role in the field of object detection, and its workflow usually includes two main steps: First, the candidate region is generated through the region proposal network (RPN); Then, the convolutional neural network (CNN) is used to classify these candidate regions and perform boundary box regression. Representative algorithms of this method include R-CNN, Fast R-CNN, and Faster R-CNN. Fig.1 represents the framework of the two-stage algorithm.

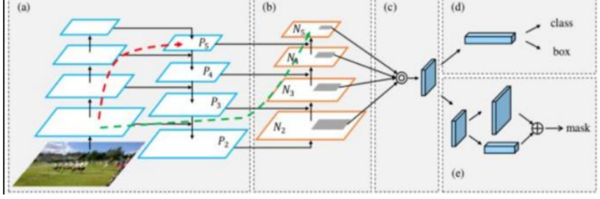


Fig.1 Two-stage algorithm

R-CNN is one of the earliest two-stage object detection algorithms. It identifies multiple candidate regions that may contain target objects in the original image by selective search algorithms, then scales these regions to a fixed size, and extracts features using a pre-trained CNN model. Finally, support vector machine (SVM) is used to classify the extracted features, and the boundary box is adjusted by regression model. Although R-CNN has made significant progress in accuracy, it is less computationally efficient because each candidate region needs to be processed independently2.

Fast R-CNN significantly improves computational efficiency by feeding the entire image into the CNN for one-time feature extraction. It introduces an RoI Pooling layer to unify the feature sizes proposed by different regions, thereby reducing double counting. However, Fast R-CNN still relies on selective search to generate candidate regions, resulting in higher computational costs. Fig.2 represents the fast R-CNN algorithm.

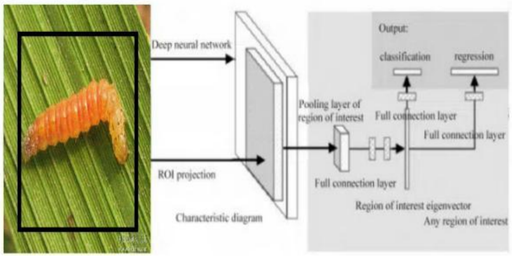


Fig.2 Fast R-CNN algorithm diagram

2.2. *Single-stage deep learning algorithm*

The single-stage deep learning algorithm uses an end-to-end approach to identify the category and location of the object using only one network on a detected image. Compared with the two-stage method, the single-stage method simplifies the detection process, greatly improves the efficiency, and has better real-time performance. A typical single-stage algorithm is YOLO3.Fig.3 represents the framework of the single-stage algorithm.

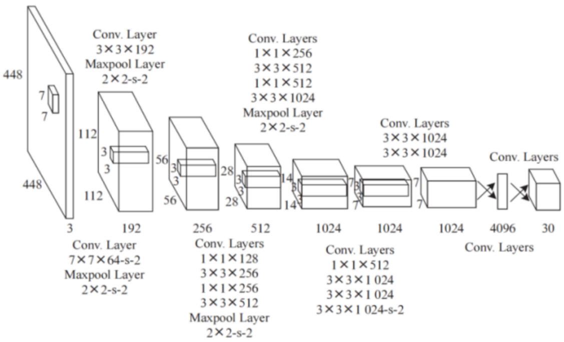


Fig.3 Single-stage algorithm

The YOLO series algorithm works by dividing the image into multiple grids, each of which predicts the border and category information of the target. The YOLO series of algorithms works by dividing the image into multiple grids, each of which predicts the boundary and category information of the target. In pest detection, YOLO algorithm is widely used in real-time monitoring and rapid response. For example, the lightweight YOLO-MobileNet-CBAM model performed well in the detection of small crop pests and diseases, with an accuracy of 92.38%.Fig.4 shows the verification results of IP102 data set after the training of YOLOv10 algorithm.

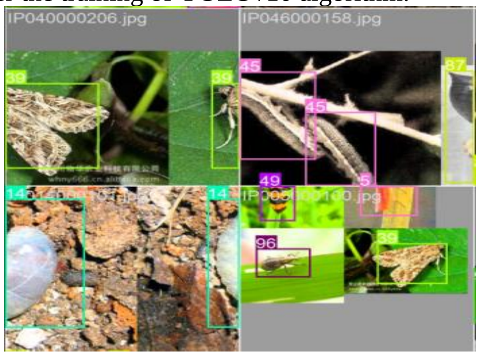


Fig.4 Verification results

3. The Main Existing Problems and Solutions

3.1. *Major problem*

In the process of applying target detection technology to automatic pest identification, although it has improved significantly compared with traditional methods, it still faces multiple challenges in practical application.

Insufficient training samples: Deep learning models, especially CNNS, require large amounts of training data to learn complex features. In a field environment, collecting enough tagged pest images is difficult because shooting in field conditions can be affected by a variety of factors, such as leaf occlusion, lighting changes, and so on. In addition, the diversity and number of insects is limited, resulting in a scarcity of samples available for training.

Pest targets are small in size: Images of pests taken in wild environments tend to be small, perhaps only a few pixels, which makes them difficult to identify in images. With the increase of the number of CNN layers, the feature information of these small targets may be lost, which increases the difficulty of detection.

High similarity between pests: Different pests can be very similar to each other in appearance, which makes it difficult for models to distinguish between them. In addition, pests of the same species may also be difficult to identify due to individual differences, attitudinal changes, or partial occlusion. Fig.5 shows the image of a small sample of insects in the actual agricultural scene.

Fig.5 Small pest targets

3.2. Solution

For the above problems, we usually have some to solve.

Data enhancement and synthesis: Generate composite images through data enhancement techniques (such as rotation, scaling, color adjustment), increase the diversity and number of data sets, and improve the generalization ability of models. Fig.6 shows the image of the original data after partial data enhancement.

Fig.6 Image after data enhancement

Multi-scale feature fusion: The feature pyramid network (FPN) and other technologies are used to fuse features of different scales to enhance the detection ability of the model for small targets.

Deep learning model optimization: single-stage detection algorithms such as YOLO and SSD are adopted to simplify the detection process and improve the detection speed. At the same time, the algorithm structure is iteratively optimized, such as the YOLOv10 model, and the accuracy and efficiency of detection are improved by improving the network structure and loss function.

4. Datasets and Data Enhancement

4.1. Datasets

In the field of deep learning-driven object detection, the quality and diversity of data sets have a decisive impact on model performance. The construction of pest image data set is the first step to achieve efficient target detection. Since the acquisition of pest images is limited by environmental factors and acquisition techniques, it is essential to build a comprehensive, diverse and accurately labeled data set. In recent years, researchers have documented a large number of insect photos, but there are still insufficient samples. For example, the ip102 dataset contains 102 insect species, but some samples have far fewer images than others. Fig.7 visualizes the number of categories in the IP102 dataset.

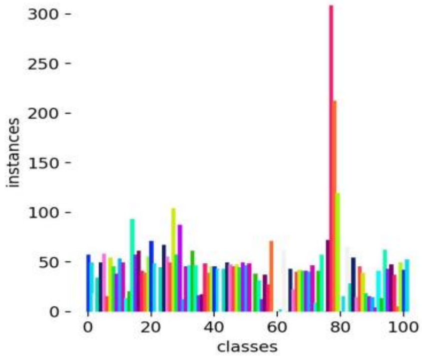


Fig.7 IP102 data set category number visualization.

4.2. Data enhancement

In view of the problem of insufficient samples in pest image datasets, data enhancement technology has become the key to improve model generalization and detection performance. Data enhancement manually extends the training set by applying a series of transform operations, including but not limited to rotation, scaling, color adjustment, blurring, and noise addition. These operations can not only increase the diversity of samples, but also improve the adaptability of the model to pest images under different environmental conditions.

In addition to traditional data enhancement methods, the researchers also explored more advanced techniques such as:

Mixup: By mixing two images and their labels in a certain proportion, a new training sample is generated to enhance the robustness of the model to input perturbations.

CutMix: Similar to Mixup, but cuts and pastes local areas of the image while adjusting the mix ratio of the label. Fig.8 shows the CutMix mixing operation.

Fig.8 Demonstration of CutMix

5. Detection Methods

After the creation of specific pest data sets and the application of data enhancement techniques to form large-scale data sets, the problem of high similarity of pest images remains a challenge in detection. Therefore, enhancing the performance of the algorithm and improving the accuracy of the model is the key to solve the problem of automatic pest image recognition. Although the object detection technology of deep learning has made remarkable application and progress, there are still many challenges in small object detection. The main problem is that the features of small-size targets are gradually lost after passing through the pooling layer of CNN, which may lead to missing detection of some targets in multi-layer networks, resulting in poor effect of existing target detection methods in processing small-size pest images.

5.1. Improvement based on network structure

On the one hand, the detection performance can be improved by increasing the depth of the network. The deep network can learn more complex feature representations, helping to capture the subtle features of small targets, but attention should be paid to solving problems such as gradient disappearance or explosion that may occur. Hu et al. studied the image classification task, based on the traditional convolutional neural network, and improved the performance by increasing the network depth. For example, the depth of the original network is increased appropriately and the residual connection technique is used to detect the tiny deformed digits in handwritten digital images. The results show that the classification accuracy of these small target numbers with subtle differences is improved by about 5%, which proves that increasing network depth combined with appropriate technology can improve the detection effect of small target.

On the other hand, changing the network architecture is also an effective approach, such as using the feature pyramid network (FPN) structure, which can comprehensively utilize different levels of feature information, and fuse shallow high-resolution feature maps with deep semantic feature maps, so that the model can detect small targets with both the required resolution for positioning and semantic information for classification. Lin et al. proposed FPN architecture for target detection. When detecting small objects in the COCO dataset, the Faster R-CNN model based on FPN architecture is used. This model integrates the feature maps of different levels, which improves the positioning accuracy of small objects by about 8% compared with the original model without FPN, fully demonstrating the advantages of FPN in the improvement of network architecture in the detection of small objects.

5.2. Improvements based on feature enhancement

The first is to carry out multi-scale feature fusion. Since small targets have different performance at different scales, it is difficult to describe a single scale feature comprehensively. By fusing feature maps output by different convolutional layers, which contain multiple features ranging from details to semantics, the model can obtain more comprehensive target information.

The second is the application of attention mechanism, such as spatial attention mechanism can guide the model to focus on the region where the small target is located, and channel attention mechanism can adapt to adjust the importance of channel features, and improve the detection accuracy by highlighting the features of the small target. Hu et al. proposed SENet to apply SENet's channel attention mechanism to the target detection model when detecting small animals in ImageNet data sets. By increasing the channel weights that contain key features of small targets, the detection accuracy of small animals is improved from about 60% to about 68% without using this mechanism, highlighting the improvement effect of attention mechanism on small target detection.

5.3. Lightweight model

The application scope of automatic pest identification should not be limited to large-scale farms, small-scale family farms are also in urgent need of it. To this end, it is crucial to create a lightweight system that can directly photograph and recognize pest images, thereby helping farmers quickly and accurately identify pest species4.

The realization of system lightweight can be started from many aspects. On the one hand, by reducing the parameter scale of the model, pruning is carried out on those unimportant weights and convolutional kernels in the CNN model to remove redundant parts and make the model "slim". Fig.9 shows the schematic diagram of pruning weight reduction.

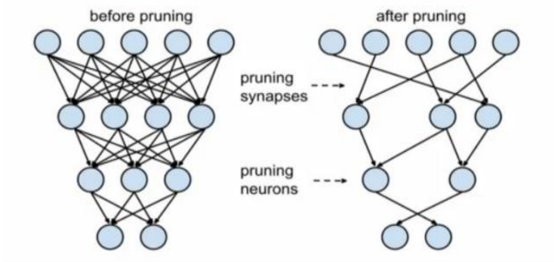


Fig.9 Pruning schematic diagram

Distillation technology can also be introduced to use the soft label information of the teacher model to guide the learning of the student model, so that the student model can maintain high performance while the number of parameters is greatly reduced and the model is further lightweight.

References

1. Popescu Dan, Dinca Alexandru, Ichim Loretta & Angelescu Nicoleta. (2023). *New trends in detection of harmful insects and pests in modern agriculture using artificial neural networks. a review* . Frontiers in Plant Science 1268167-1268167.
2. Kai Run Li, Li Jun Duan, Yang Jun Deng, et al.,(2024).*Pest Detection Based on Lightweight Locality-Aware Faster R-CNN*.Agronomy(10),2303-2303.
3. Ajantha Vijayakumar & Subramaniyaswamy Vairavasundaram. (2024).*YOLO-based Object Detection Models: A Review and its Applications.Multimedia Tools and Applications*(35),83535-83574.

4. Tingyao Jiang & Shuo Chen. (2024).*A Lightweight Forest Pest Image Recognition Model Based on Improved YOLOv8.*Applied Sciences(5),

Authors Introduction

Mr. Hongshuo Zhai

He is a second-year master candidate in Tianjin University of Science and Technology.

Dr. Fengzhi Dai

He received M.E. and Doctor of Engineering (PhD) from the Beijing Institute of Technology, China in 1998 and Oita University, Japan in 2004 respectively. His main research interests are artificial intelligence, pattern recognition and robotics. He worked in National Institute of Technology, Matsue College, Japan from 2003 to 2009. Since October 2009, he has been the staff in College of Electronic Information and Automation, Tianjin University of Science and Technology, China.

Mr. Lijiang Zhang

He has been chairman of Xinjiang Shenhua Biotechnology Co., Ltd, China and Zhongnong Guoyou Technology (Tianjin) Co., Ltd, China since 2019. His research interest is deep processing of agricultural products.

Mr. Qiang Wang

He was graduated in Tianjin University of Science and Technology in 1997. Since 2022, he has been the chairman of Easy Control Intelligent Technology (Tianjin) Co., Ltd., China, focusing on the development of KRMotion: A PC-Based Software Motion Control Platform.

Design of an Intelligent Pet Feeding System Based on STM32

Shuhuan Peng [1]*, Qiang Wang [2]

[1] *Tianjin University of Science and Technology, Tianjin, China*

[2] *Easy Control Intelligent Technology(Tianjin) Co., Ltd., Tianjin, China*

*E-mail: *2879431759@qq.com*

Abstract

This design aims to develop an intelligent pet feeding system based on the STM32F103RCT6 microcontroller. By integrating multiple sensors and actuators, the system achieves automated feeding and watering, while also offering a user-friendly interactive interface and remote monitoring capabilities. This paper details the system's hardware and software design, implementation processes, functional testing, and result analysis. The system's stability and reliability have been verified through rigorous testing. Although there are areas for improvement in practical application, this design has achieved the basic functions of intelligent feeding and possesses good potential for expansion.

Keywords: STM32F103RCT6, intelligent pet feeding, sensors, user interaction, remote monitoring

1. Introduction

Nowadays, with the improvement of people's material living standards and the diversification of spiritual needs, the pet industry has been unprecedentedly developed.

However, with the acceleration of the pace of modern life, people's increasingly busy daily life has gradually transformed the affectionate care and proper care of pets at home into a topic that cannot be ignored[1].

The design and development of intelligent pet feeding systems provides an effective way to solve the above problems. Through the Internet and artificial intelligence, the system is designed with integrated sensors, timers and network interfaces, which can realize the function of caring for pets while fully taking into account the convenience and life rhythm of pet owners[2]. When pet owners need to go out for a long time, they can remotely control the feeding time and amount from their mobile device. In people's busy lives, the system can assist pet owners with feeding and watering, ensure that pets receive nutrition as planned, and provide appropriate and effective care for pets in a timely manner[3].

The design purpose of the intelligent pet feeding system is based on STM32F103RCT6 microcontroller, integrating a variety of sensors to realise the automatic completion of pet feeding and water supply tasks, and realise the function of interaction and remote monitoring through the support of host computer software to meet the needs of pet owners and pets[4].

The rest of this article is organized as follows. The second section introduces the hardware design of the system. In the third part, the software design of the system is introduced. In the fourth part, the results of the system tests and the analysis of the results are presented. The fifth part summarizes the main content of this paper.

2. Hardware Design

The design uses STM32F103RCT6 as the main control chip. As shown in Fig.1, in the intelligent pet feeding system, when the perception module of the system detects that the pet needs to eat, this information is transmitted to the control module for processing and analysis, and then the control module transmits the release instruction derived from the result analysis to the servo and the water pump of the execution module, so as to realise automatic water feeding. In addition, pet owners can feed through the LCD touch screen of the human-computer interaction module, and at the same time, they can also use the Bluetooth HC-05 module in the communication module to control the system through the host computer software for feeding. Fig.2 shows the actual system diagram.

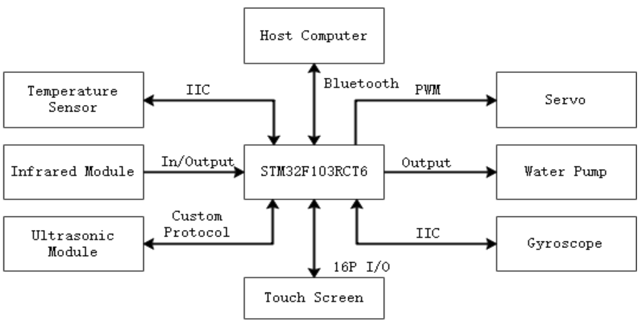


Fig.1 System control logic block diagram

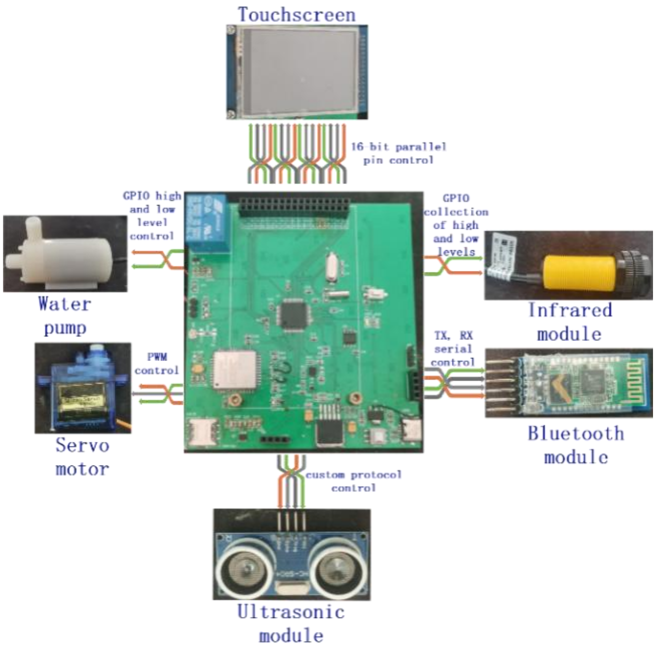


Fig.2 Physical diagram of the system

2.1. *Main control chip*

In this system, the STM32F103RCT6 microcontroller is used as the main control chip, as shown in Fig.3, which has powerful processing power, rich peripheral support and high cost-effectiveness, and has very high applicability in complex control and data processing tasks.

STM32F103RCT6 microcontroller has many functions in this system. First of all, the STM32F103RCT6 microcontroller has the function of information collection, which effectively communicates with various sensors through interfaces to collect the operating status of feeding and drinking mechanisms and environmental information in real time. In addition, based on data processing, the STM32F103RCT6 microcontroller can issue instructions to each actuator based on the analysis results, for example, send instructions to the servo and water pump, control the feeding and drinking of pets, and ensure timely response to pet needs. At the same time, the microcontroller is also linked with the host computer software to allow pet owners to remotely monitor the status of the system through the interactive platform to help pet owners understand the number of times their pets eat.

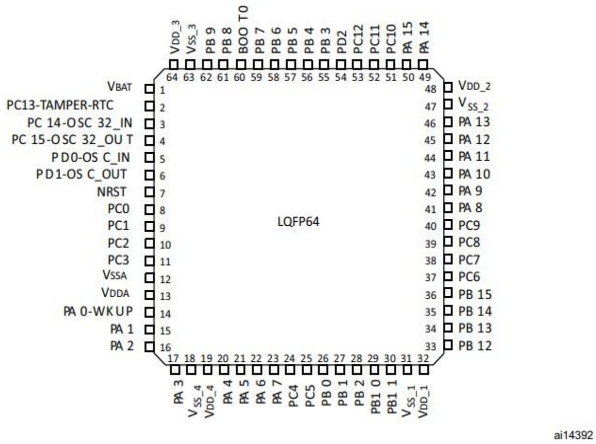


Fig.3 Pin diagram of the microcontroller

2.2. *Sensor modules*

The sensor module of the pet feeding system integrates an infrared sensor, an ultrasonic sensor, a gyroscope, and a temperature sensor to accurately detect and respond to pet behavior.

The infrared sensor, E18-D80NK, initially detects the pet's presence within a 70cm range, triggering the ultrasonic sensor, HC-SR04, to ensure precise distance measurement, activating the feeding mechanism only when the pet is at the optimal distance of 30cm. The gyroscope, MMA8452, monitors the system's stability, alerting the owner if the device is tipped or moved, thereby preventing feeding errors. Additionally, the TMP102 temperature sensor maintains optimal food storage conditions by monitoring ambient temperature, ensuring the pet's food remains fresh and consumable. Together, these sensors work in harmony to ensure efficient and safe operation of the feeding system, promoting the pet's health and the system's reliability.

2.3. *Execution module*

In this smart pet feeding system, the execution module includes a servo and a water pump, and the function of the module is to ensure that the system can carry out the instructions for feeding and water supply.

(1) The servo

When performing the feeding function, the servo can adjust the opening and closing of the feeder mouth according to the instructions of the control module, so as to accurately control the amount of food fed and the feeding time. The servos are able to respond quickly to commands in a short period of time, providing discrete or continuous feeding, allowing the system to meet the dietary needs of pets at different times.

The servo used in the system is the SG90, which is a miniature position servo drive. The working principle of SG90 is to accurately control the rotation angle by receiving control signals, which is widely used in model aircraft, robots and other fields.

(2) The water pump

The water pump is responsible for supplying the water required by the pet when performing the water feeding function, and similar to the servo, the water pump also operates according to the instructions issued by the control module to ensure that the pet has a stable and hygienic water source.

2.4. The human-computer interaction module

The human-computer interaction module uses the LCD touch screen as an interactive platform, providing an intuitive and easy-to-operate interface for pet owners. The interface allows pet owners to easily set up various functions and customise a personalised feeding plan for their pet.

The interactive interface of the system uses a 2.8-inch resistive touch screen, which can be operated directly by pet owners. When the pet owner touches different areas of the screen, the resistance value of the screen changes to enable control of the system.

2.5. The communication module

The communication module used in the system is the Bluetooth HC-05 module, which has the function of wireless connection and data exchange with the user's smart devices (such as mobile phones or tablets).

In this system, the Bluetooth HC-05 module works in slave mode, and the STM32F103RCT6 microcontroller connects with the Bluetooth module through the UART communication protocol, so as to realise wireless communication with the mobile phone software. When the system performs feeding or water supply functions, the relevant data will be sent to the Bluetooth HC-05 module via the control module, and then transmitted to the communication device via Bluetooth. As a result, pet owners can monitor their pets' feeding and drinking conditions in real-time on communication devices. Through this communication module, the system realises the connection between the system and the communication equipment, and also greatly improves the user's operation experience and the convenience of the system.

2.6. Design of PCB boards

This smart pet feeding system is mainly used in home life, so the volume should not be too large. Therefore, in this system, we use 9*10 PCB boards for hardware design.

PCBs are designed to make circuits miniaturised, intuitive, and play an important role in the optimisation of electrical layouts. The design of the PCB board is shown in Fig.4.

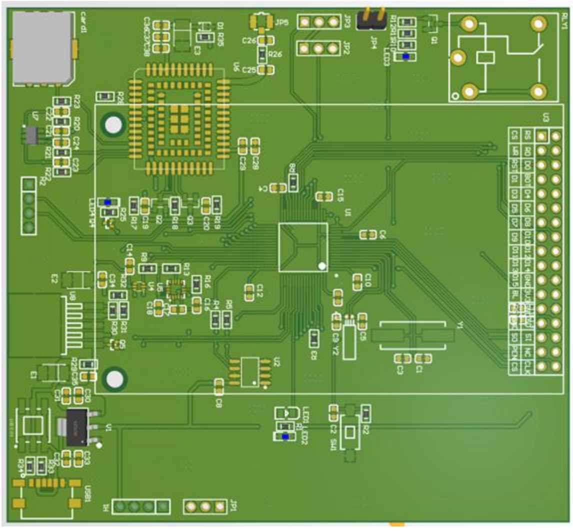


Fig.4 3D view of the PCB

3. Software Design

The software design of the system uses Keil 5 MDK version 5.37.

The total programme flow chart of the system is shown in Fig.5, through the software design, the system realises the functions of infrared/ultrasonic detection and feeding control, dumping detection and alarm, temperature monitoring and so on.

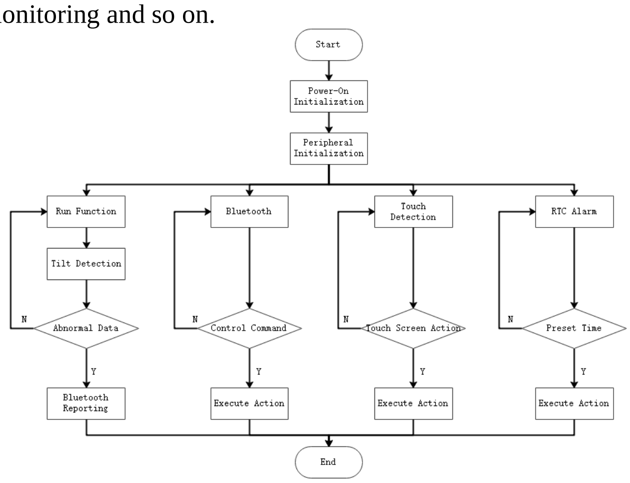


Fig.5 Overall software flow chart of the system

4. System Testing and Analysis of Results

In this design, the five core functions of the system (timed feeding, infrared/ultrasonic detection, tipping detection, temperature detection, and control and feedback of mobile app) are tested.

The experimental results show that the intelligent pet feeding system has basically realised the functions of automatic water feeding, ambient temperature monitoring, equipment tipping detection, Bluetooth control and so on.

4.1. Timing feeding function test

Table 1 records the test results of the timed feeding function. The data show that the deviation between the trigger time of the actual feeding and the preset time of the system is within the acceptable error range, which proves the stability and reliability of the timed feeding function.

Table 1 Timing feeding function test results

Test number	Preset feeding times	Actual trigger time	Deviation	Outcome
01	08:00:00	08:00:02	+2s	Pass
02	12:00:00	12:00:04	+4s	Pass

4.2. *Infrared/Ultrasonic Detection Testing*

Table 2 records the test data of the infrared/ultrasonic detection function. The table shows that the system is able to accurately activate the feeding and feeding procedures when the infrared sensor detects an object within a set range of 70 cm and the ultrasonic sensor ensures that the object is within 30 cm of the feeding programme.

Table 2 Infrared/Ultrasonic test results

Test number	IR presets Detection distance	Ultrasound presets detection distance	Physical distance	Outcome
01	70 cm	30 cm	20 cm	Feeding
02	70 cm	30 cm	50 cm	Do not feed

4.3. *Tipping detection test*

Table 3 shows the performance of the tip detection function of the system. The test verifies that the system is able to detect and issue an alarm in time to prevent misoperation of the equipment in the event of a tipping event.

Table 3 Dump detection test results

Test number	Simulate the dumping angle	Test results	Whether or not to trigger an alert	Outcome
01	15°	Not detected	No	Not passed
02	45°	Detect	Yes	Pass

4.4. *Mobile device functionality testing*

When the serial port assistant of the mobile app is used to send WATER to the Bluetooth module, the device starts to run the pumping work, returns WATER when the command is received, and returns happen water when the action is executed, as shown in Fig.6.

Fig.6 Bluetooth water feeding function test

When the serial port assistant of the mobile app is used to send FEED to the Bluetooth module, the device starts to run the feeding work, returns OK when the command is received, and returns the happen feed when the action is executed, as shown in Fig.7.

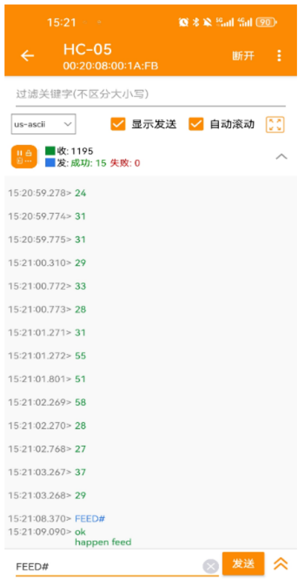


Fig.7 Bluetooth feeding function test

5. Conclusion

This paper presents the design and implementation of an intelligent pet feeding system utilizing the STM32F103RCT6 microcontroller. The system integrates various sensors and actuators to achieve automatic feeding and watering, while providing user-friendly interfaces and remote monitoring capabilities. The hardware design involves sensor modules, including infrared, ultrasonic, gyroscopic, and temperature sensors, to accurately detect and respond to pet behavior, ensuring efficient and safe operation. The execution module, consisting of a servo and water pump, executes feeding commands precisely. The communication module facilitates remote control via Bluetooth, enhancing user convenience. Rigorous testing demonstrates the system's stability and reliability in key functions like timed feeding, detection accuracy, and mobile control. While there is room for improvement, the system successfully fulfills its primary functions and offers potential for future expansion, addressing pet care challenges in modern busy lifestyles.

References

1. Zhang, Jianbao. "Design and application of intelligent pet feeding management control system." 3rd International C onference on Internet of Things and Smart City (IoTSC 2 023). Vol. 12708. SPIE, 2023.
2. Liu, Ningning, et al. "Design of an Intelligent Pet Feeding System." 2023 International Conference on Electronics a nd Devices, Computational Science (ICEDCS). IEEE, 202 3.
3. Ma, Baijie, and Ning Guo. "Design of remote pet feeding system based on ARM." 2020 Chinese Automation Congr ess (CAC). IEEE, 2020.
4. Hasni, Helmi Che, and Suhazlan Suhaimi. "Pet Feeding S ystem." International Journal of Recent Technology and A pplied Science (IJORTAS) 4.2 (2022): 111-118.

Authors Introduction

Ms. Shuhuan Peng

She received her B.S. degree from College of Electronic Information, Tianjin University of Science and Technology, China in 2024. She is currently a Master course student in Tianjin University of Science and Technology.

Mr. Qiang Wang

He He was graduated in Tianjin University of Science and Technology in 1997. Since 2022, he has been the chairman of Easy Control Intelligent Technology (Tianjin) Co., Ltd., China, focusing on the development of KRMotion: A PC-Based Software Motion Control Platform.

Design of Teaching Attendance System Based on Image Processing

Shengyu Wang[1*], Ce Bian [2]

[1] *Tianjin University of Science and Technology, Tianjin, China*
[2] *Tianjin Tianke Intelligent Manufacture Technology Co., LTD., China*

*E-mail: *wsy972117935@163.com*

Abstract

Traditional classrooms often require teachers to roll call one by one in class, which not only affects the length of the class but also affects the quality of the class, and there is no timely data feedback, resulting in the lag of the work of many college students. This topic mainly uses computer simulation software for algorithm research, which is mainly divided into four parts, the first is the initial establishment of the face library, the second is the use of PCA algorithm for face image dimensionality reduction in face recognition, the Euclidean distance is used again for face closest matching, and finally the function is realized by GUI interface.

Keywords: image recognition, face recognition, MATLAB, PCA algorithm

1. Introduction

In recent years, there has been an increasing emphasis on disciplines in the size of the university student population problems of students, students absenteeism, difficulties in failing to pass, difficulties in graduation and other phenomena occur frequently.

At present, the commonly used forms of human identity verification mainly include magnetic cards, IC cards and identification numbers, and these technologies have now become mature, but this information is easy to be imitated and falsified, and there are many problems. However, face recognition is based on the inherent characteristics of the human face, which is difficult to imitate and has strong immutability, which is the mainstream scientific research direction. Face recognition technology has occupied a wide field in research that is still continuing in this work. [1]

The design is developed on the basis of the analysis of the technical characteristics of the PCA principal component analysis face recognition algorithm and the research status and difficulties at home and abroad, and the PCA algorithm based on principal component analysis is applied to reduce the dimensionality of the face image to obtain the face feature value, and the near-neighborhood K-L algorithm is used to match and recognize the face. In this dissertation, a face recognition system has been built using MATLAB. The camera is retrieved, the picture is collected, the face is positioned and segmented, and the face is compared with the sample database to obtain the comparison face. And the number of recognitions, time, etc. are displayed through the GUI interface. [2, 3]

The rest of this article is organized as follows. The second section introduces the algorithms related to image processing. In the third part, introduced facial feature localization, establishment of facial database, and optimization of algorithm usage. In the fourth section, simulation example is given to verify the effectiveness of the designed protocol. The fifth part summarizes the main content of this paper.

2. Algorithms Related to Image Processing

With the steady improvement of computer performance, the algorithm of image processing is also gradually following up, and the main algorithms of this system are PCA (Principal Component Analysis) algorithm based on face recognition and K-L (Karhunen-Loeve) transform.

2.1. Principal component analysis algorithm

The PCA algorithm is currently the most widely used data dimensionality reduction algorithm. The main idea of PCA is to transform r-dimensional features into k-dimensions, which are also known as principal components as a new orthogonal vector.

PCA work is to sequentially search for a set of mutually orthogonal coordinate axes in the original space. From Fig.1 can be observed that. The two axes mainly refer to two-dimensional data, and the black dots represent two-dimensional data points. If the data of one axis is projected onto the space of other axes, it can be reduced in dimension. The use of PCA method can greatly increase computational efficiency.

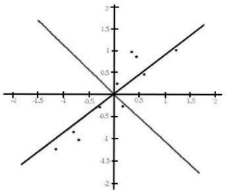


Fig. 1 Schematic diagram of two-dimensional data projection

2.2. Principle of k-l transformation

The K-L transform is essentially a method applicable to a wide range of random images. When the K-L transform is applied to N images, the images in the set composed of the transformed N new images are not correlated with each other. Obtaining a finite number of images M (M<N) restored from the transformed result image set will become the optimal approximation of the original image in statistical sense.

Assuming Y is an n-dimensional random variable, it can be represented by a weighted sum of n basis vectors.

$$X = \sum_{i=1}^{n} \alpha_i \varphi_i \quad (1)$$

Find the self coherence matrix of the random vector X.

$$R = E[X^T X] \quad (2)$$

Find the eigenvalues λ_j and eigenvectors $\emptyset_j$ of the autocorrelation matrix or covariance matrix R.

$$\alpha = \emptyset^T X \quad (3)$$

2.3. Classifier recognition

Introduce the training sample data after dimensionality reduction of the transformation matrix into the classifier, complete classifier data training after specific classifier classification, match the reduced facial image data with the data trained by the original classifier, import the facial image to be detected into the system, and complete facial recognition based on the classification criteria of each classifier.

The Nearest Neighbor Classifier is currently one of the most theoretically sound algorithms in classifiers. It requires finding the closest m points, calculating the distance between this point and the other points, and then counting these m points to find the point with the highest weight and determine which category this point belongs to. The most important condition for the use of nearest neighbor classifier algorithms is to measure datasets of M-size that are close in distance. The formula used is Euclidean distance:

$$D(x, y) = \|x - y\| = \left| \sum_{i=l}^{n} (x_i - y_i)^2 \right|^{\frac{1}{2}} \quad (4)$$

3. Facial Feature Localization

In this design, it is necessary to locate the face. During the face detection process, the features of the eyes and mouth can be recognized and converted into specific vectors, which are then added to the PCA algorithm for subsequent processing.

3.1. Face detection

The main implementation process is to input any frame of image through the laptop camera, and the system automatically uses PCA face detection algorithm to distinguish whether there is a face in the input image. If a face is detected, it is then located, the relevant position information of the face is recorded, and the face is cropped to obtain an effective face image.

Facial detection scenarios can be divided into two types. One is to use computer cameras to directly capture facial images for detection and capture facial images; The second is to perform face detection on the images cached in the computer. If a face is detected, the contour information and position information of the face are recorded, and the relevant information of the person is displayed accordingly.

3.2. Facial image preprocessing

Due to the influence of different factors such as lighting, shooting angle, facial expressions, during the process of shooting in the same environment, image preprocessing is required, such as grayscale processing, binarization processing, histogram equalization, geometric normalization, in order to process the obtained facial photos.

For histogram equalization transformation, it can be divided into two types: normalized histogram and equalized histogram. Normalized histogram and equalized histogram are shown in Fig.2 and Fig.3.

(a) (b)

Fig. 2 Original image and corresponding histogram

(a) (b)

Fig. 3 Balanced effect and corresponding balanced histogram

3.3. Facial detection process

In the process of face detection, the main features of the face, such as eyes, mouth, or color, can be used for detection. After morphological processing, the target can be effectively segmented, and then each feature can be read, combined with PCA algorithm processing. The specific process is as follows:

Collect facial images using computer cameras or download facial images from websites. After image segmentation processing, the face image is obtained.

According to the requirements, perform facial region detection and collect partial feature data, such as facial color and other features. Collect facial features. Facial detection process is shown in Fig.4.

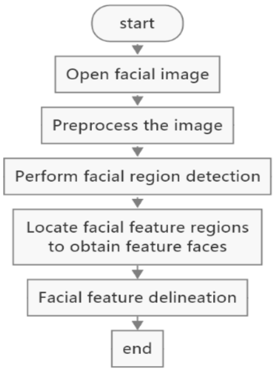


Fig. 4 Facial localization process

3.4. *Face location*

For the facial localization of this system, there may be multiple faces in one screen or other types of problems, which may mislead the collection of facial information.

Using a face detector to reduce the coefficient enables faster retrieval of faces in images. In the detection process, for possible multiple faces, the largest face is selected as the sample, and after the sample is captured, it is enlarged. Facial localization process is Fig.5.

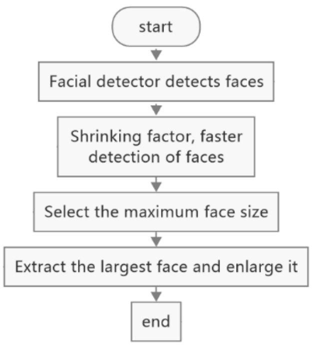


Fig. 5 Facial training sample library

Find the corner points in the detection area, where bbox is a matrix sequence. Form a rectangular box by framing the face, calculate the size of the rectangular box area, and select the maximum area rectangle as the face output.

3.5. *Establishment of face training sample library*

First, establish a TrainDatebase face training sample database, which can copy and paste foreign face samples. The same people try to choose several pictures with large differences, so that there is more room for future face selection. You can also use a computer camera to capture faces using a GUI interface. After that, image processing will be performed to enlarge the face, select suitable photos, and arrange them in the order of numbers 1, 2, 3, etc. into the face training sample library. Facial training sample library is Fig.6.

Fig.6 Facial detection process

3.6. *Detailed process of image improvement based on PCA algorithm*

For PCA, its basic idea can be understood as assuming that a set of 2D points shown in Fig.7 can be approximated as a single row, that is, the size of the points can be reduced from 2D to 1D.

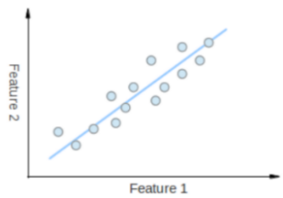


Fig. 7 2D point chart

We can see that these points change the most along the blue line, more than along the feature1 or feature2 axis. This means that if we can know the position of a point along the blue line, we have more information about that point than just knowing its position on the Feature1 or Feature2 axis.

The main optimization process of this system is to obtain the average vector of the sample, calculate the vector difference between each vector and the average vector, synthesize it into a feature matrix, and solve for the intelligent perception and feature vectors. Select the feature vectors for dimensionality reduction, and finally obtain the feature face.

Firstly, calculate the average vector of the sample, and its main formula is:

$$\bar{x} = \frac{1}{n}\sum_{\alpha=1}^{n} x_{(\alpha)} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_P) \qquad (5)$$

Calculate the difference vector between each sample and the average vector, subtract the average face from each sample, and finally combine the vectors into an n*n matrix.

After forming the matrix, calculate the covariance matrix, whose main formula is:

$$S \triangleq \sum_{\alpha=1}^{n}(X_{(\alpha)}-\bar{x})(X_{(\alpha)}-\bar{x})' = (S_{ij})_{p*p} \qquad (6)$$

Solve the eigenvectors and eigenvalues through the covariance matrix, select the eigenvectors, and finally perform dimensionality reduction to obtain the feature face. For eigenvectors and eigenvalues, it can be understood that if A is an n-order square matrix and there exists a constant λ and a non-zero n-vector x, then

$$Ax = \lambda x \quad (7)$$

So λ is the eigenvalue of matrix A, and x is the eigenvector of A belonging to eigenvalue λ.

The comparison of the image before and after reconstruction is shown in Fig.8.

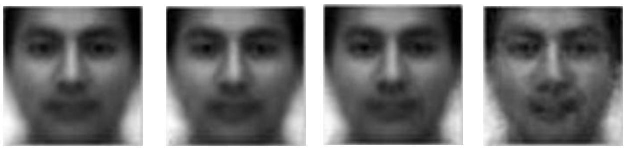


Fig. 8 Comparison of its image before and after reconstruction

3.7. *Feature face matching based on euclidean distance*

Euclidean metric (also known as Euclidean distance) is a common distance concept that refers to the true distance between two points in m-dimensional space, or the natural length of a vector (i.e. the distance between a point and the origin of coordinates). The Euclidean distance in 2D and 3D space is the true distance between two points. The two-dimensional formula is:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (8)$$

The three-dimensional formula is:

$$d = \sqrt{(x_1 - x_2)^2 + (y_i - y_2)^2 + (z_1 - z_2)^2} \quad (9)$$

In n-dimensional space, two image matrices form a point, and then the Euclidean distance formula in mathematics is used to calculate the distance between the two points. The smallest distance is the most matching image. The process is shown in Fig.9.

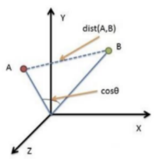


Fig. 9 Euclidean distance diagram

After combining the PCA algorithm with Euclidean distance. The process is shown in Fig.10.

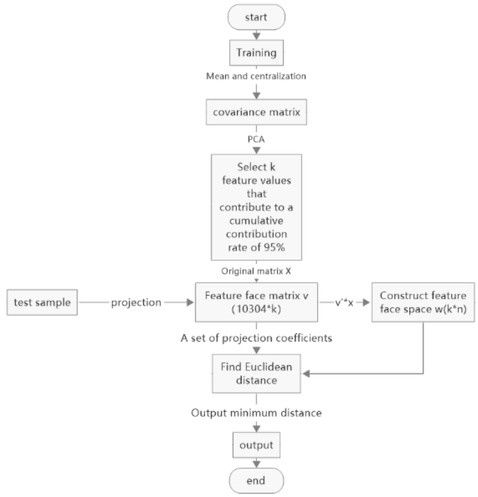


Fig. 10 Facial recognition and output flowchart

4. System design and simulation implementation

4.1. *System detailed design*

The recognition process of this system is shown in Fig.11, This system mainly has two modules, namely dynamic camera processing and static image processing.

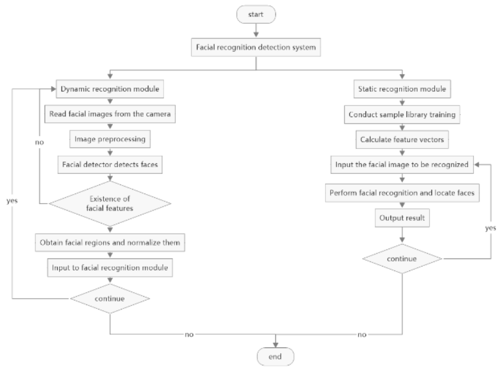


Fig. 11 Facial recognition detection system

4.2. *Establishment of system GUI interface*

The GUI interface of this system mainly includes facial recognition, positioning, testing, and camera shooting functions. After successful recognition, personal information will be displayed, such as class name, check-in times, check-in times, etc., and the check-in times of all people this week will be counted, which is shown in Fig.12.

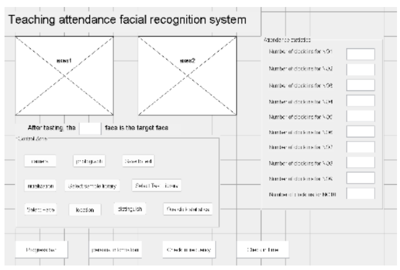


Fig.12 Initial GUI interface

4.3. Simulation results of the operation of each module in the system

Firstly, using the camera function of MATLAB, borrow a computer camera to take facial photos. After confirming the shooting, the image will enter the first box, which is shown in Fig.13.

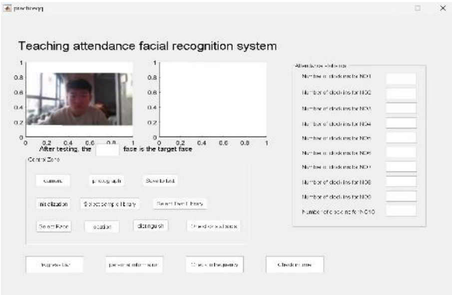


Fig.13 Shooting results

After shooting, click save to save it to the test data database. Then initialize the page and select the sample library for training, which involves extracting feature faces and identifying feature values. After successful training, it will prompt that the facial database training is complete, as shown in Fig.14.

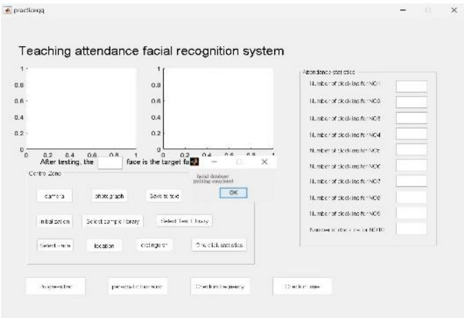


Fig.14 Sample library testing

Further test the test library. After all tests are completed, select the desired face and perform localization. The system will locate the face area based on the image. After clicking on recognition, the system will compare the face information contained in the image with the database and finally select the face that is closest to it. After recognition, a one click statistics will be performed, and the attendance count will be summarized on the right side of the diagram. Since the people recognized it once and it was number one, the system counted the number of clock ins as one. The final result is shown in Fig.15.

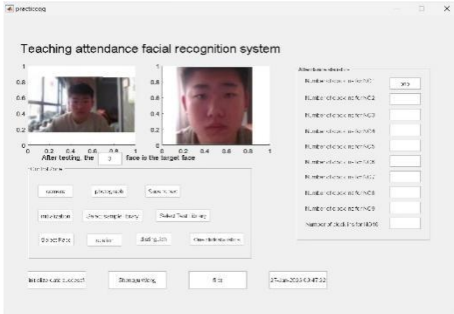


Fig.15 final result

The static recognition of ORL face database testing is also carried out in the same way. As shown in Fig.16.

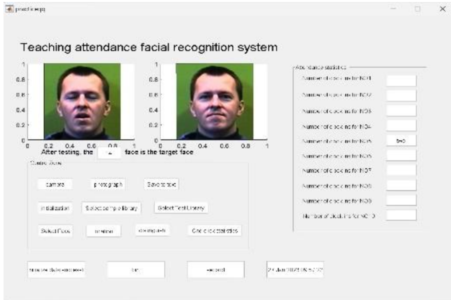


Fig.16 ORL facial database testing

5. Conclusion

According to the test results, there are also some incorrect recognitions, most of which are caused by extremely similar facial features. And due to the recognition accuracy of the PCA algorithm being only 85%, there may be a certain degree of error based on the component weights of each person in the sample library, which requires continuous testing.

Reference

1. Abd L S A ,Abdulelah A A ,Tahir S F A .*Face Detection for Color Image Based on MATLAB*.Journal of Physics: Conference Series,2021,1879(2).
2. Choudhary S ,Moriwal R ,Bansal S P D .*Face Recognition System Based on PCA*.Journal of Research in Science and Engineering,2020,2(11).
3. Yan X ,Xi Y ,Gaqiong L .*Realization and optimization of face recognition system based on MATLAB*.Journal of Physics: Conference Series,2020,1650(3):032091.

Authors Introduction

Mr. Shengyu Wang

He received his B.S. degree from College of Automation, Tianjin Sino-German University of Applied Technology, China in 2023. He is currently a Master course student in Tianjin University of Science and Technology. His research area is about deep learning and image processing.

Mr. Ce Bian

He received his master's degree in control engineering from Tianjin University of Science and Technology in 2019, focusing on machine vision technology. Since 2017, he has been working in Tianjin Tianke Intelligent Manufacture Technology Co., LTD., serving as the general manager.

Semi-automatic Leek Harvester Based on Multi-angle Adjustment

Hongpi Zhao *, Xuefeng Jia, Wenqi Fu,Yizhun Peng
Tianjin University of Science and Technology, Tianjin, 300222,China

*E-mail: * cxdz107@163.com*

Abstract

The team is committed to solving the domestic leek harvesting process of excessive human input, high cost, intelligence, low level of mechanization, to provide users with a diversified range of high-performance semi-automatic leek harvester equipment and solutions, which can make the leek production safer, time-saving, labor-saving. The team independently researched and developed leek harvester which can walk independently and harvest automatically, adopting new adjustable mechanical structure, artificial intelligence algorithm and human-computer interaction software application, which makes the domestic automatic harvesting gradually become possible.

*Keywords:*leek harvesting, intelligent, mechanical innovation, self-propelled

1. Introduction

This product from the farm planting leek growers of the actual problem, in-depth for their leek harvesting for labor intensive, high cost problems, to understand their real needs, and then determine the leek harvester design positioning. After sufficient market research, after analyzing and comparing several typical leek harvesters on the market, we chose the multi-angle adjustment; through ergonomic analysis to determine the main body size of the harvester, and then carried out the selection of materials, structural design, parts design and strength check; through the button operation start-stop worm gear DC gear motor control machine forward, DC gear motor drive synchronous toothed belt transmission, and 3D printing is used to slope the design for the grain support device.[1,2]

The rest of the paper is organized as follows. Part II presents the overall mechanical design. Part III presents the light source following system design. Part IV gives the finite element analysis to verify the effectiveness of the designed mechanism. Part V summarizes the main points of this paper.

2. Overall Mechanical Design

This product is an intelligent leek harvesting robot based on multi-angle adjustment. It can harvest accurately, and the stubble height level is controlled within 3% error, and the whole machine movement speed is 0.3m/s, the speed is smooth, and it can meet the requirements of leek planting with row spacing of 30-40cm and plant spacing of 15-20cm. The whole machine size in length × width × height = 1670mm × 470mm × 470mm, the total quality of 65kg, by the transmission mechanism, supporting Harvesting institutions, harvesting institutions, the three main institutions. The overall organizational structure and physical diagram is shown in Fig.1.

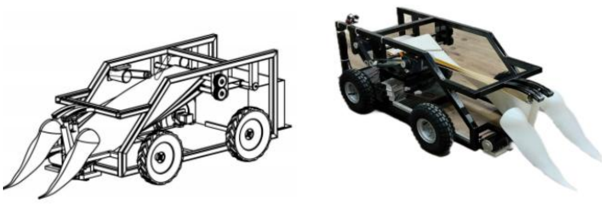
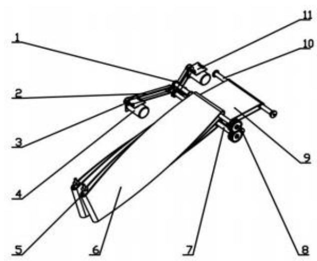


Fig.1 Overall organizational structure and physical map

2.1. Delivery mechanism

In order to complete the recycling of leeks after harvesting,so that leeks fall into the collection device in a smooth and orderly manner,we utilize a rotary conveyor belt and a horizontal buffer conveyor belt to form a conveyor belt group,and the degree of wrapping reaches 90%.

The mechanism includes mechanical components such as shafts,gears,pulleys,DC motors,motor drives,buttons,and so on.The conveyor mechanism is shown in Fig.2.

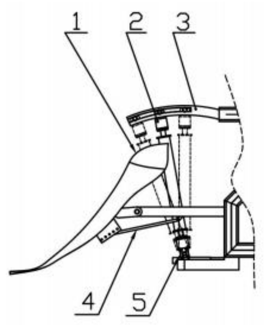


1—driven pulley；2—synchronous belt；3—drive pulley；4—DC motor；5—Vertical follower shaft；6—Sponge Conveyor Belt；7—Horizontal with main shaft；8—spur gear；9—horizontal conveyor belt；10—Horizontal follower shaft；11—Motor connection plate

Fig. 2. Conveyor mechanism

2.2. Fuhe organization

The product of the Harvest Supporting Device is symmetrical setting, Harvest Supporting Plate operation is a continuous progressive process, the lowest end is almost in contact with the ground and parallel to the ground, the height from the ground is less than 2cm, and the vibration error is controlled at 6.3%. The shape of the cross-section of the Harvest Supporting Device is based on the progressive smooth arc shape of the plant with minimum force and less mechanical damage, the front end is 10mm wide, the back end is 100mm wide, showing a progressive trend, preventing the Harvest Supporting Angle for different kinds of leeks is different, and the organization can adjust the position of the connecting rod to make a small degree of angular adjustment. The composition of the mechanism is shown in Fig.3.[3]



1—propane；2—upper connection；3—indexing bar；
4—connection bar；5—bottom bracket
Fig.3. Harvest support mechanism

The mechanism can adjust the angle of the connecting rod corresponds to a number of holes, and then adjust the angle of the Harvest Board, in addition to adjusting the upper and lower connectors to adjust the clamping angle, the connecting rod mechanism schematic diagram shown in Fig.4.

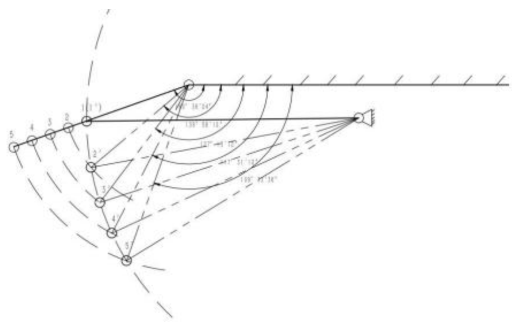


Fig.4.Schematic diagram of connecting rod mechanism

2.3. Harvesting mechanism

This harvester is single row harvesting, compact structural space arrangement, and leeks in the clamping assistance to complete the cutting, so the choice of disc cutter. The disk cutter is shown in Fig.5.

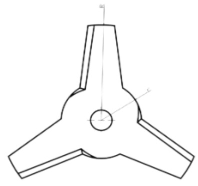
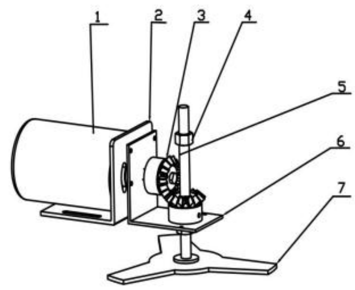


Fig.5. STM32F103ZET6 chip

The whole structure of the cutting device is shown in Fig.6, which consists of cutting motor, rotating shaft of the cutter, bevel gear, relaxation nut, etc. The cutting device can be adjusted by adjusting the height of the rotating shaft to ensure the cutting of leeks. Cutting device by adjusting the height of the rotary axis to ensure that the stubble height. The cutting disk is a one-piece disk knife, and each side edge cuts the leek.Blade parameters are shown in Table 1.

Table 1. Blade parameters

makings	edge angle	InnerBore Diameter	rotary diameter	Base circle
65Mn	17°	16mm	120mm	67mm



1—DC motor；2—Motor Bracket；3—pinion gear；
4—Loosen the screws.；
5—cutter shaft；6—Gear connection plate；7—fretting knife
Fig. 6 Cutter mechanism

3. Dual-axis light source following system

In order to solve the problem of insufficient power generation efficiency of traditional solar panels, the team has developed a multi-functional two-dimensional tracking solar power generation system especially for the possible power shortage of the harvester, which is capable of tracking intelligently according to the intensity of light and the sun's altitude angle, so as to significantly improve the power generation efficiency, and the physical object is as shown in Fig. 7.

Fig. 7 Physical solar charging system

The control system is composed of an Arduino microcontroller L298N motor drive module and a solar panel with four built-in analog photoresistors to ensure the smooth operation of the system.

3.1. Principle of operation

When the sun's light intensity changes, the built-in photoresistor of the solar panel will automatically recognize the light intensity, after reading the data will be fed back to the Arduino microcontroller for processing, and finally the high and low levels will be output by the L298N motor driver module, so as to control the motor, in order to achieve the effect of adjusting the angle of the two directions, horizontally and vertically.

In order to guarantee the stable operation of the system as well as the energy saving and environmental protection indexes, the horizontal and vertical angle change rates during the operation of the mechanism are 15° per second and 10° per second, respectively, and at the same time, the program compiled inside the microcontroller system has additional limitations, so that the system is activated to run only when the difference in the intensity of the sensed light reaches a certain level.

4. Finite element analysis and verification

Modal analysis is an efficient way to study the structural vibration characteristics, mainly applied to analyze whether the machine structure design is reasonable, through modal analysis, can determine the cutter's intrinsic frequency, vibration pattern and vibration mode and other information, so as to derive the structural strength and stability of the cutter, as well as to predict possible resonance phenomenon, resonance can be avoided, to provide theoretical guidance for the design and improvement of the later stage.

According to the results of modal analysis, the vibration frequency of the cutter can be obtained in the range of 201.55~320.13 HZ, and the vibration pattern is mainly concentrated in the position of the cutting teeth. The maximum amplitude of the cutter is 162.3mm in the 4th order, and the vibration pattern mainly shows the twisting and deformation of the cutting edge of the cutting teeth. According to the requirements of the machine production operation cutter speed up to 1800r/min, its resonance frequency is about 28.66Hz, through the modal analysis of the cutter, the minimum frequency of the cutter is 201.55Hz, so the cutter will not resonate in the process of the work and the phenomenon of local early destruction, to meet the cutter in the work of the safety and stability requirements. The modal analysis is shown in Fig.8.

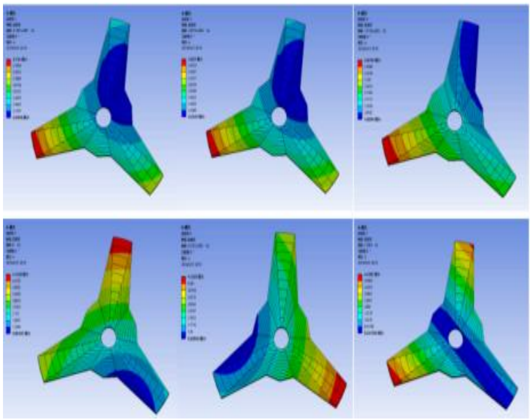


Fig. 8 Sixth order modes of cutter

5. Conclusion

Based on the high labor cost and low efficiency problems faced by leek growers, this paper designs and develops an intelligent leek harvester, which systematically solves a number of pain points existing in the traditional harvesting process. Through market research and demand analysis, this paper establishes the design concept of multi-angle adjustment as the core, specifically including the modular design of the transmission mechanism, the harvesting mechanism and the harvesting mechanism. Among them, the transmission mechanism adopts the combination of rotary conveyor belt and horizontal buffer conveyor belt, which improves the stability and efficiency of leek collection; the harvesting mechanism effectively reduces the plant damage by adjusting the angle of the connecting rod and optimizing the clamping angle; and the harvesting mechanism uses disc cutter, which, combined with the compact structural layout, ensures the precise control of the harvesting height. The whole machine is designed to achieve a stable movement speed of 0.3m/s and meets the requirements of row spacing and plant spacing for leek planting, providing users with an efficient and reliable mechanization solution.

In addition, in order to solve the problem of insufficient power that may occur when the equipment runs for a long time, this paper innovatively designs a dual-axis light source following system. It provides a reference for the subsequent application of solar energy in agricultural machinery.

The strength and stability of the cutter and other key components are verified through finite element modal analysis to ensure that the equipment will not resonate or be partially damaged under high-frequency operating conditions.

With intelligence, energy saving and high efficiency as the core, this design explores the new path of leek harvesting mechanization and provides new technical solutions and innovative ideas for the development of agricultural mechanization and intelligence.

References

1. ZHANG Wendong,YANG Xuelan,GUO Hong'en,et al. Design and test of tracked self-propelled leek harvester[J]. *Agricultural Equipment and Vehicle Engineering*,2024,62(11):1-6+13.
2. ZHANG Xuming,YANG Ying,ZHANG Hao,et al. Experimental study on the application of electric leek harvester[J]. *Agricultural Machinery Quality and Supervision*,2024,(08):18+20.
3. LI Jiwei,BIAN Lina,ZHAI Joy,et al. Technology and development of new device for supporting harvesting device of leek harvester[J]. *Anhui Agricultural Science*,2021,49(21):223-226+230.

Authors Introduction

Mr. Hongpi Zhao

He is currently pursuing his undergraduate degree at Tianjin University of Science and Technology. He is currently studying in the School of Mechanical Engineering.

Mr. Xuefeng Jia

He is currently pursuing his undergraduate degree at Tianjin University of Science and Technology. He is currently studying in the School of Mechanical Engineering.

Ms. Wenqi Fu

She is currently pursuing her undergraduate degree at Tianjin University of Science and Technology. She is currently studying in the College of Electronic Information and Automation.

Dr. Yizhun Peng

He is an Associate Professor of Faculty of Engineering at Tokyo Polytechnic University in Japan. He graduated from the Department of Mechanical Engineering,Fukui University, in 2000. He received his D. Eng. degree in System Design Engineering from Fukui University in 2005. His research interest is Robotics.

Smart Inspection Guard - Inspection Robot for Unattended Plants

Liangyu Wang *, Yanhong Yu, Yizhun Peng
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300457, China
*E-mail: *2697909953@qq.com*

Abstract

This document describes an intelligent inspection robot, based on Arduino and Raspberry Pi, with autonomous navigation and video surveillance. The robot uses infrared sensors for tracing, RFID for localization, and uploads the captured video to the server. The innovative integration of wireless charging technology realizes the unattended function, as well as the self-designed camera clamping mechanism. Key technologies cover differential control, wireless charging, data communication and server design. Tests show that the robot can improve inspection efficiency and quality, save labor costs, and comply with the trend of intelligent manufacturing.

Keywords: Intelligent Inspection Robot, Autonomous Navigation, RFID, Infrared Tracing, Wireless Charging

1. Introduction

With the rapid development of the times, the level of automation is also increasing, the application of robots is becoming more and more widespread, industrial, educational, medical and service robots are also specifically applied in various industries, replacing numerous positions, but also for people to bring great convenience, and some of the work of repetitive, boring traditional positions will also be replaced by robots. [1]

For large-scale unattended plant, the traditional inspection method is not enough to meet the demand, in order to reduce the work intensity of laborers, the use of robots instead of people to complete those repetitive and boring work, improve the quality of equipment inspection. This paper designs an intelligent inspection robot for unattended machine rooms, relay stations and other locations, which has the functions of autonomous patrol, network video monitoring and so on.

The rest of this article is organized as follows. The second section introduces the overall design of intelligent inspection robot. The third section introduces the hardware design of intelligent inspection robot. The fourth section describes its software design.The fifth part summarizes the main content of this paper.

2. The Overall Design

2.1. Overall structural design

The inspection robot system has four components: an Arduino-based control system for movement and data, a Raspberry Pi-based logic system for tasks like path planning and image capture, a server for data storage, and a client for accessing data. It operates in a distributed manner with a control center, data transmission, and a robot terminal layer equipped with sensors. The robot can inspect autonomously, avoid obstacles, and send alerts when needed.

2.2. Hardware structure of the robot

The robot's chassis is designed with two active wheels and two universal driven wheels for agile turns. The active wheels are powered by motors controlled by an Arduino2560, which takes commands from a Raspberry Pi. A 4S battery with explosion-proof and power detection features supplies power, regulated to ensure reliable operation of the electronics. The robot also has a wireless charging receiver and eight infrared sensors for navigation. It includes temperature and humidity sensors and a height-adjustable camera bracket with three cameras and LED lights, all connected with 3D-printed parts.The appearance and physical drawing of the inspection robot is shown in Fig.1.

Fig.1 Exterior and physical view of the inspection robot

2.3. The working process

This design inspection robot uses Raspberry Pi as the logic processor, its functional modules are: image acquisition (Open CV), path planning, motion control, maintenance and support module, each module communicates with each other through the MQTT protocol, and the motion control module communicates with the underlying motion controller Arduino through I2C to send control commands.

The robot operates via an Arduino that executes commands from a Raspberry Pi and uses infrared sensors for path tracing, ensuring it follows a predetermined route. An RFID reader under the chassis prompts the robot to stop for image capture, which are then uploaded to a server for client access. [2]It runs on a 4S battery with a voltage regulator to power the Arduino and Raspberry Pi, and it can wirelessly recharge at designated charging points when the battery is low.

3. Hardware Design of Robot

3.1. Design of the navigation system

The navigation algorithm for inspection robots is designed for swift and stable movement along a guide line, utilizing infrared sensors to determine the robot's position and orientation. This design employs the TCRT5000 reflective infrared sensor for the trajectory module, which includes a high-power infrared transmitter and a high-sensitivity receiver. [4]When the transmitted infrared signal is reflected back to the receiver, the resistance in the receiver changes, leading to a voltage change in the circuit. This change is then processed by an ADC converter module to yield output results. The resistance change, and thus the voltage signal, depends on the intensity of the infrared signal received, which is influenced by the color of the reflective surface and the distance from the receiver to the emitting surface .Infrared sensor shown in Fig.2.

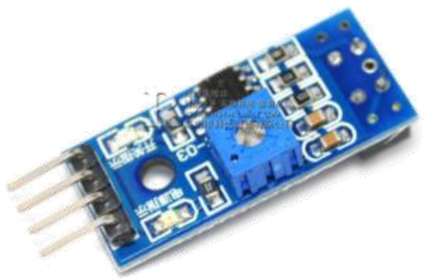


Fig.2 Infrared sensor

3.2. Fixed-point image acquisition

(1)Fixed-point identification:

RFID technology enables the inspection robot to identify its position and capture images by detecting positioning cards along its route through radio frequency communication. The RFID system, which includes tags, readers, and a communication network, works by having the reader send out a signal that the tags receive and respond to, allowing the reader to decode the tag's information and send it to a computer for processing and control actions.

(2)Image acquisition:

The inspection robot is equipped with a high-definition camera that activates and captures images at designated locations during its tasks. The captured image data is then analyzed and processed, with the final results sent to a server for storage. This system allows for early detection of equipment failures, enabling timely maintenance and preventing major safety incidents. When the environment is poorly lit, a fill light module consisting of six LEDs is used to ensure clear image capture, and this module can be configured to be normally open or closed based on the robot's working conditions.[3]

3.3. Automatic obstacle avoidance system design

For distant objects, the robot uses visual recognition and ranging. Up close or when passing obstacles, cameras might miss the full view, so photoelectric sensors on the robot's corners are used for obstacle avoidance. If the robot strays off course, it retreats, turns 30 degrees, and repeats until it finds the navigation line again, ensuring it doesn't damage any equipment.

3.4. Underlying motion controller

The Arduino2560 is an open-source electronics platform consisting of hardware (Arduino boards) and software (Arduino IDE). It's based on a simple I/O interface and uses a Java/C-like language similar to Processing/Wiring.[5]Arduino can interact with the environment via sensors and control devices like lights and motors. The board's microcontroller is programmed using the Arduino language, compiled into binary, and uploaded to the controller. Projects can involve just Arduino or combine it with other software on a PC for communication. The hardware architecture is shown in Fig.3.

Fig.3 Arduino2560

4. The Software Design

4.1. Design of the tracing algorithm

A tracing algorithm based on eight infrared sensors was designed to enable the robot to recognize and follow the black line for navigation.The feedback values from the infrared sensors determine the robot's motion states, including stopping, traveling straight, small adjustments and large adjustments.Through the coding approach, the robot is able to perform six different types of motion to suit different tracing needs.

4.2. Software design of navigation and localization module

RFID technology is utilized for localization, and the card reader reads the RF card information to enable the robot to achieve precise positioning on the inspection route.Communication between the reader and the Raspberry Pi ensures that the robot is able to stop at a specific location for image acquisition.The RFID system consists of tags, readers and a communication network to ensure accurate data transmission and processing.

4.3. Path planning algorithm design

Python language is applied on Raspberry Pi to automatically generate maps and navigate according to the path planning algorithm.The patrol mode of the robot is categorized into fully automatic patrol and point-to-point patrol to adapt to different inspection needs.Through the recursive function and path table, the robot is able to record and optimize the path to achieve efficient patrol.

4.4. Image acquisition module software design

The robot is equipped with three cameras for collecting panoramic images of the equipment and uploading the data to the server via wireless network.The image acquisition module is triggered when the RFID card is read by the RFID reader, which ensures image acquisition at a specific location.The captured image data is packaged and uploaded along with the power and location information of the robot for remote monitoring and analysis.

5. Conclusion

There is a growing demand for safety and environmental monitoring in many warehouses, logistics and production lines in China. Although inspection robots have been used in the power industry, they have not yet been popularized in the field of ground inspection. With the development of science and technology, the shortcomings of manual inspection are gradually being compensated by intelligent inspection robots, which are of great significance in terms of social and economic benefits. This paper introduces an intelligent inspection robot, which integrates infrared, photoelectric, temperature and humidity sensors and cameras, and is able to record the site conditions. The robot is based on Raspberry Pi and Arduino, runs logic processing and motion control programs, and uploads video information to the client via wireless network. We also designed a low-battery automatic return wireless charging function, and verified the robot's detection accuracy and functional reliability through lab tests. This robot is able to monitor the status of items, environmental parameters and power level in real time, and automatically plan the path and avoid obstacles to ensure the safety and efficiency of inspection.

References

1. Xing S-Y, Jiang W. Research on communication quality of substation inspection robot. Internet of Things Technology, 2018.
2. Meng Xiangzhong, Wang Baolei. Research and design of RFID-based wireless charging system for substation inspection robot. Industrial Instrumentation and Automation Device, 2017
3. Ma Yiming. Application of intelligent inspection robot in unattended substation . North China Electric Power University, 2017.
4. Zhang Qiang, Hu Lifu, Li Peng. Design of intelligent substation inspection robot based on WIFI. China Science and Technology Information, 2016
5. WU Gongping, YANG Zhiyong, Wangwei, et al. Autonomous charging and docking control method for inspection robot. Journal of Harbin Institute of Technology, 2016

Authors Introduction

Liangyu Wang

Liangyu Wang is currently pursuing his undergraduate degree at the School of Electronic Information and Automation, Tianjin University of Science and Technology.

Yanhong Yu

Yanhong Yu is currently pursuing her undergraduate degree at the School of Electronic Information and Automation, Tianjin University of Science and Technology.

Yizhun Peng

He is an associate professor at the School of Electronic Information and Automation, Tianjin University of Science and Technology, with research interests in neural network control.

Deep Learning Based Integrated Removable Smart Waste Sorting Device

Yanhong Yu, Liangyu Wang, Yizhun Peng
College of Electronic Information and Automation, Tianjin University of Science and Technology, 300457，China
*E-mail: * 2782921778@qq.com*

Abstract

In this study, an intelligent waste sorting device based on Inception v3 and migration learning is developed to achieve fast and accurate waste recognition and sorting through deep learning and sensor fusion techniques. The device is designed to be detachable and adaptable to existing bins, with the ability to continuously learn new waste types. Through real-time data transmission, the device supports remote monitoring and management, which effectively improves the efficiency of waste classification and is important for urban environmental protection.

Keywords: Intelligent waste sorting, Inception v3, sensor fusion, detachable design, environmental monitoring

1. Introduction

The problem of municipal waste is serious, with China producing more than 150 million tonnes per year , with an annual growth rate of 8-10%, and a cumulative stockpile of 7 billion tonnes. This waste leads to resource wastage, and the resource recovery rate is less than 5%, which is much lower than that of developed countries [1] . Waste classification is the key to achieving waste minimisation, resource recovery and safety, and is crucial for sustainable urban development.

National and local governments attach great import ance to waste classification and have introduced man y policies.20-24 The General Office of the State Cou ncil issued the Opinions on Accelerating the Constru ction of a Waste Recycling System, which further e mphasises the importance of improving the efficiency of resource utilisation and realising the fine managem ent, effective recycling and efficient utilisation of wa ste [2] .

To improve the efficiency of waste disposal, this s tudy proposes an intelligent waste sorting device base d on Inception v3 and transfer learning, which solves the problem of irregular waste recognition and enabl es the bin to learn new types of waste. The device can be retrofitted to existing bins and the hardware a nd software work together to achieve automatic sorti ng. The product is suitable for a variety of indoor lo cations and is equipped with a management terminal and a mobile phone terminal, which makes it easy to monitor the status of the bin and encourages the pr actice of waste classification.

The rest of the paper is organized as follows, Sect ion 2 describes how the device implements the proce ss of waste sorting through image recognition and se nsor-assisted detection. The key techniques in Section 3 involve machine vision, sensor fusion and intellige nt route planning to improve the accuracy and efficie ncy of waste sorting. Part 4 details the physical desi gn of the four-classified and two-classified waste bin s, as well as the software design of the web and mo bile applications. Part 5 shows with tests that the de vice achieves 94% image recognition accuracy and th e power system ensures continuous power supply und er different conditions. Part VI summarizes the main features of the device, including easy modification, a utomatic sorting, reminder function, migration learnin g and diversified power supply system.

2. Overall Design

Once the waste has been thrown out, the first step is to capture and recognise the image of the waste being thrown out. At the same time, capacitive proxi mity sensors, metal sensors, humidity sensors and V OC odour sensors are used to aid detection. The det ection information is then transferred to the STM32 microcontroller for summary analysis, and the microc ontroller then controls the self-designed automatic sor ting device to complete the waste classification.The o verall flow chart is shown in Fig.1.

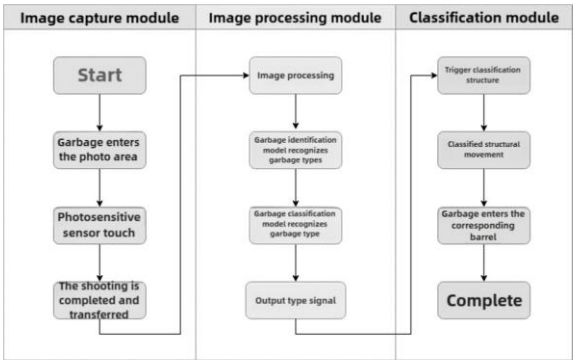


Fig.1 Overall flow chart

3. Key Technologies

3. 1. Machine vision technology

Machine vision technology is at the core of this project, and a deep learning model based on Inception v3 is used for image classification. The device is based on Raspberry Pi 3B+ image processing technology combined with camera recognition and is written in Python [3] . The process of recognising plastic water cups and the recognition results are shown in Fig.2.

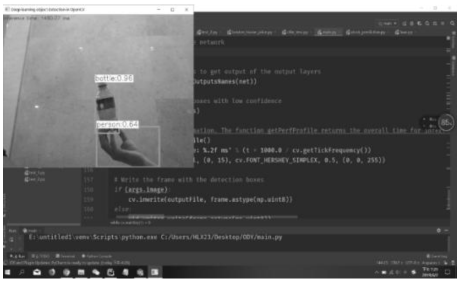


Fig.2 Identification of plastic water cups

To improve the recognition performance of the waste classifier, deep learning techniques are fused with Haar feature integral map and Adaboost algorithm to train simple and strong classifiers. For target detection, YOLO algorithm is used to combine SSD and MobileNet to achieve efficient target detection on Raspberry Pi. The target detector is built using OpenCV, the labels are set, the images are processed, the neural network is trained, and the parameters are adjusted to achieve optimal detection.

3. 2. Sensor fusion technology

In order to improve the recognition capability of the system, various sensors such as capacitive proximity sensors [4] , metal sensors, humidity sensors and VOC odour sensors are integrated. These sensors provide additional information to help the system identify and sort the waste more accurately.

3. 3. Intelligent Recycling Vehicle Planning Route

There are GPS module and networking module (ESP8266, ESP-12) connected to the main control board of all intelligent bins, which stores and sends the map data to the network according to the tile algorithm, and obtains the GPS positioning information through the serial port and analyses the latitude and longitude coordinates from it. Based on the coordinates, the corresponding map data is read and displayed to the recycler. After all the coordinate points are uploaded to the network, the dynamic planning algorithm is used to automatically plan the optimal paths for the recyclers to reach each point, which facilitates the recycling of rubbish, as shown in Fig. 3.

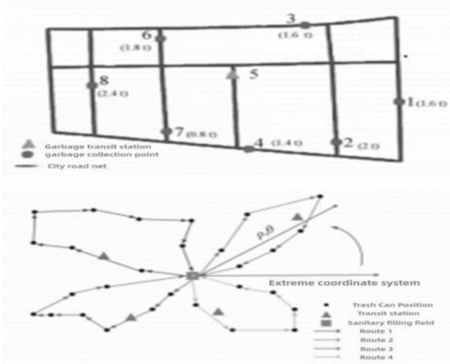


Fig.3 Schematic diagram of multi-circle combination of random collection points

4. Device Design and Implementation

4. 1. Physical end design

The physical end design consists of two bin config urations: a four-compartment bin and a two-compartm ent bin. The four-compartment bin is suitable for the Chinese waste classification standard, while the two-c ompartment bin is suitable for the classification of re cyclable and other waste.

- Sorting Structure for Four Types of Waste

A solar panel is installed above the top panel of the trash can, and the sorting structure is installed in the middle entrance area of the four categories of trash cans. A camera is mounted on the bottom of the roof panel. The said box is equipped with a number of sorting chambers for the collection of different types of garbage. The sorting chambers follow the parallel arrangement of the common garbage bins, so they have a stronger applicability and the sorting device can be removed and directly retrofitted to the existing four-port garbage bins. There is an electrically controlled platform in the middle of the upper part of the sorting compartments, which can be moved left and right, as well as turned down and dumped, and the garbage turning and dumping plate is a groove structure that can receive garbage,as shown in Fig. 4.

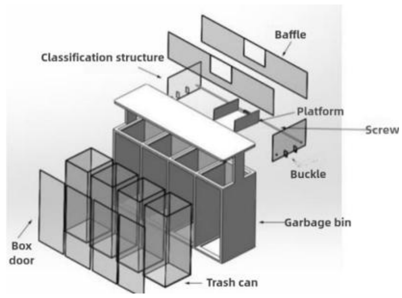
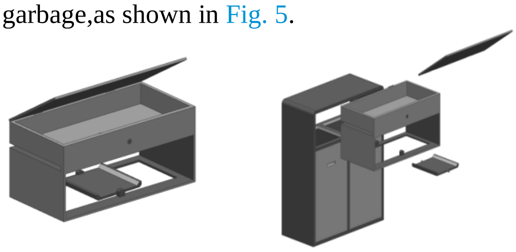


Fig.4 The structure diagram of the four types of garbage sorting device

- Two Types of Waste Sorting Device

The device is fully integrated and can be regarded as two parts, and the top plate of the trash can can be directly inserted into the center when it is retrofitted. There is a hardware box on the top of the trash can, and the sorting structure is designed in the middle entrance area of the trash can. The lower box has a camera installed on the upper surface, and the middle part is equipped with an electrically controlled platform that can move left and right, as well as turn down and dump, and the garbage turning and dumping plate is the same groove structure that can receive garbage,as shown in Fig. 5.



（a）Two types of classification devices （b） combination charts

Fig. 5 Exploded view of installation of two types of garbage sorting devices

Both structures take into account the compatibility of existing bins and can be intelligently upgraded with simple modifications.

4. 2. Software design

The software comprises web and mobile applications. The web side displays real-time information about the status of the bins, such as location, storage status and battery level. The mobile side provides navigation and information retrieval functions for refuse vehicle drivers to optimise recycling routes and improve efficiency.

5. Device Testing and Evaluation

5.1. Image recognition accuracy test

The image recognition accuracy of the device was tested[5]. The results show that the recognition accuracy of the optimised system reaches 94%, which is higher than similar products on the market. During the testing process, we considered different lighting conditions, garbage shapes and colours to ensure the stability and accuracy of the system.

5.2. Power continuity test

The unit uses a power supply system that combines solar energy and lead-acid batteries to ensure energy self-sufficiency. Under sunny conditions, the solar panels are able to provide sufficient power for the system, while at night or on cloudy or rainy days, the batteries are able to ensure the continuous operation of the system.

6. Conclusion

The team has designed a detachable IoT waste sorting device with the following features:

- It can be directly retrofitted to the existing rubbish bin. It saves the cost of intellectualisation, adopts modular design in both hardware structure and software programming, and reserves some interfaces, so that users can easily add these functions if they have requirements such as user login and accumulated points.
- Solve the problem of misplaced rubbish in public places. Applying related technologies such as IoT, machine vision and machine learning to public bins, it realises automatic classification of rubbish and solves the problem of misplacing rubbish due to people being in a hurry, not knowing how to classify and having a low awareness of rubbish classification.
- Overflow reminder function. It can notify relevant people to collect rubbish in time, effectively solving the problem of overflowing rubbish without anyone disposing of it. At the same time, it provides intelligent route planning function to achieve the optimal path to recycle rubbish, realise the effective use of personnel and vehicles, and greatly save the cost of manpower and material resources.
- With the ability of migration learning. When encountering rubbish that cannot be identified, the bin will upload the data to the cloud, and the relevant personnel will identify and mark the type in the background. When the next time to meet the same kind of rubbish can be automatically classified. At the same time all the data between the bins is shared, in short, has a brain, but there are a lot of split, so you can learn at the same time, the speed of learning will be exponential.
- Diversification of power supply system. The use of solar energy and battery combined power supply system or underground wiring, to achieve self-sufficiency of energy, to ensure that the use of the location is not restricted. At the same time strive to maximise the environment without pollution, low power consumption, environmental protection and reliability.

References

1. Shen, Yingqing. (2011). Status quo and countermeasures for garbage classification in China. Northern Environment (08), 13-14.
2. The General Office of the State Council. Opinions on Accelerating the Construction of a Waste Recycling System. Beijing: The General Office of the State Council, 2024.
3. Dai, Congling, Tang, Linjie, Xie, Zhenghua,et al.2011 A new species of the genus Pterostilbene (Araneae, Pterostilbene) from China. (2021). Intelligent waste classification system in scenic spots based on Easy-DL platform + Raspberry Pi. Internet of Things
4. Xu, Ruichao, Line, Chen, Hu, Yankai, ,et al. (2021). Multi-sensor based intelligent trash can. Equipment Machinery (04), 42-44.
5. Zizheng, Cheng Hui,He Jia. 2011 A new species of the genus Pseudococcus (Coleoptera, Staphylinidae) from China. (2023). Intelligent waste classification technology based on deep learning. Science, Technology and Innovation (21), 46-49.

Authors Introduction

Yanhong Yu

Yanhong Yu is currently pursuing her undergraduate degree at the School of Electronic Information and Automation, Tianjin University of Science and Technology.

Liangyu Wang

Liangyu Wang is currently pursuing his undergraduate degree at the School of Electronic Information and Automation, Tianjin University of Science and Technology.

Yizhun Peng

He is an associate professor at the School of Electronic Information and Automation, Tianjin University of Science and Technology, with research interests in neural network control.

Enhanced Deep Reinforcement Learning for Robotic Manipulation: Tackling Dynamic Weight in Noodle Grasping Task

Prem Gamolped, Ja Sin Yon Pang, Vjosa Bytyqi, Eiji Hayashi
Department of Creative Informatics, Kyushu Institute of Technology,
680-4 Kawazu, Iizuka, Fukuoka, 820-8502, Japan
Email: gmp.prem1997@gmail.com, yon-pang.sin-ja413@mail.kyutech.jp,
vjosabytyci@gmail.com, haya@mse.kyutech.ac.jp

Abstract

Handling food items with dynamic weight changes over time, which alter physical properties such as shape, size, and weight, poses significant challenges, particularly when precise output weight is required. This study introduces an enhanced deep reinforcement learning framework for robotic manipulation, focusing on the task of spaghetti grasping. Building on prior research, we propose a data augmentation strategy that simulates diverse environmental conditions, including variations in image observations and the physical properties of spaghetti, to improve models. The model is validated using metrics such as grasp success rate, average grasp time, and generalization score under varying environmental conditions. This work advances the robustness of robotic models in previously unseen environments.

Keywords: Robotic Manipulation, Deep Reinforcement Learning (DRL), Data Augmentation, Grasping Task

1. Introduction

The automation of complex food handling tasks has emerged as a critical focus in robotics, addressing the growing demand for precision and efficiency in food production lines. Building on our previous research [1] in robotic manipulation for assembling bento boxes, this study extends the exploration of deep reinforcement learning (DRL) techniques to address the challenges of handling food items with dynamic weight changes. While our earlier work focused on grasping and placing spaghetti with a specific weight under a controlled environment, this study will tackle the broader problem: how robotic systems can adapt to dynamic changes in the physical properties of food items, such as weight, shape, and size, over time. Handling deformable food items, such as spaghetti, presents unique challenges in robotic manipulation. Spaghetti's properties are affected by factors such as humidity, temperature, and preparation methods—which necessitate precise grasping strategies to achieve consistent grasping weight. In our previous study, a DRL-based robotic system demonstrated significant success in grasping spaghetti with a Mean Absolute Percentage Error (MAPE) of approximately 19%. This success defines the potential of DRL-based in tackling complex food-handling tasks but also reveals the need for more robust solutions capable of adapting to environmental variations.

This study introduced an enhanced DRL framework specifically designed for dynamic food-handling scenarios. The framework incorporates a data augmentation strategy to simulate diverse environmental conditions, enabling the robotic system to generalize under diverse environmental conditions. By integrating these data augmentation techniques to make the environment more varied in the training process, the proposed aims to improve the robustness and adaptability of DRL models in real-world environments. In this study, we will use the metric to assess the model's performance in dynamic settings. These include grasp success rate (GSR), average grasp time (AVT), task efficiency (TE), and generalization score (GS) under varying conditions. Such metrics can provide a sufficient evaluation framework, highlighting the system's ability to adapt and perform reliably despite the fluctuations in the target item's properties. The experimental validation of our approach builds upon the methodology and setup outlined in our prior work, which included a custom-designed gripper, an RGB-D camera for visual input, and a digital scale for weight measurement. Extending the prior system, the current study incorporates data augmentation [2], [3] techniques, leveraging diverse datasets to enhance the DRL model's learning process. The result is a system capable of achieving consistent grasping performance even when confronted with previously unseen environments.

2. Related Works

This section reviews previous studies in robotic food manipulation, and deep reinforcement learning (DRL) applications in dynamic environments. These provide the foundation for our proposed framework and highlight the gaps it seeks to address.

2.1. *Robotic Manipulation in Food Handling*

The integration of robotics into food processing lines has seen significant advancements over the past decade, with applications ranging from ingredient sorting to packaging. However, handling deformable objects [4], such as spaghetti, dough, or fruits, poses unique challenges due to their dynamic and unpredictable properties. Research has investigated diverse grippers and sensing technologies to enable precise handling, but achieving generalization across environmental conditions and object variability remains a significant challenge.

In a previous study [1], a DRL-based robotic system for spaghetti grasping in bento box assembly achieved a weight accuracy MAPE of around 19% using a custom gripper, RGB-D sensing, and a digital scale. While effective in controlled settings, it struggled with variations in spaghetti texture, weight, and humidity, highlighting the need for better generalization strategies to handle dynamic food properties.

2.2. *Deep Reinforcement Learning for Robotic Manipulation*

Deep Reinforcement Learning (DRL) has shown significant success in training robots to perform complex tasks in unstructured environments. Algorithms such as Soft Actor-Critic (SAC), Proximal Policy Optimization (PPO) [10], and Deep Q-Networks (DQN) [9] have been widely adopted for robotic manipulation tasks, leveraging the ability to learn from trial-and-error interactions with the environment.

Several studies have applied DRL to robotic grasping tasks, focusing on both rigid and deformable objects. For instance, this study [5] proposed a hybrid approach combining DRL and soft grippers for grasping objects of varying shapes and sizes, achieving robust performance in cluttered environments. Similarly, another study [6]. Demonstrated the potential of DRL in dexterous manipulation tasks, showcasing the ability to handle objects with complex geometries. Although these studies highlight the efficacy of DRL in grasping tasks, they largely assume static object properties, reducing their relevance for dynamic scenarios like food handling. In the context of food manipulation, similar work [7] employed reinforcement learning to pick and place cluttered objects using dense object descriptors. However, the study did not address challenges related to changes in object properties over time. This work extends these approaches by incorporating data augmentation strategies that simulate variations in environmental and object conditions, enabling the model to adapt to dynamic changes in food properties.

2.3. *Gaps and Opportunities*

Despite significant advancements in robotic manipulation, several gaps remain in the field. First, existing research predominantly focuses on static or semi-static environments, overlooking the dynamic changes in object properties observed in real-world scenarios. Second, while DRL has shown promise in robotic manipulation, its application to tasks involving deformable objects with unpredictable properties remains underexplored. Third, the evaluation metrics used in prior studies are often limited to success rates, neglecting broader measures of robustness and adaptability.

To address these gaps, this study proposes an enhanced DRL framework incorporating a data augmentation strategy to handle dynamic food items. By simulating diverse environmental conditions during training, the proposed approach aims to improve model robustness and adaptability, enabling reliable performance in previously unseen environments. Additionally, it has been defined comprehensive evaluation metrics, including grasp success rate, average grasp time, and generalization score, to assess the framework's effectiveness in dynamic scenarios.

3. Methodology

This section explains the methodology of this study, starting from the system overview, deep reinforcement learning framework, reward function design, data augmentation pipeline, and training environment setup.

3.1. *System Overview*

The robotic system in this study is based on the setup described in our previous work, which includes a robotic arm equipped with a custom gripper, an RGB-D camera for visual observations, and a digital scale for measuring grasped weights. The system was designed to grasp and manipulate spaghetti from a stationary tray, to achieve a target weight. The training and evaluation were conducted in a controlled environment, leveraging Deep Reinforcement Learning (DRL) for task optimization. The system as shown in Fig. 1 utilized a Yaskawa SIA5F robotic arm for its seven-axis versatility, coupled with a custom 3D-printed gripper. An Azure Kinect RGB-D camera provided both color and depth information, while the digital scale recorded the weight of the grasped spaghetti in real-time.

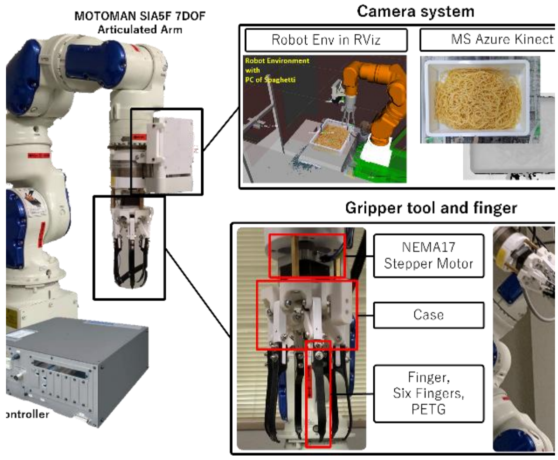


Fig.1 The robotic system in the study mainly includes an arm manipulator, gripper, and camera.

The system's overall architecture allowed precise grasping and manipulation of deformable food items like spaghetti. The PC configuration we utilized for this study is as follows:

- Ubuntu Linux 20.04 LTS
- 12th Gen Intel® Core™ i9-12900K×24
- NVIDIA GeForce RTX 3090Ti 24GB
- 64GB of RAM
- ROS Noetic
- Stable Baselines3 1.8.0 and OpenAI Gym 0.21.0

3.2. *Deep Reinforcement Learning Framework*

The Deep Reinforcement Learning (DRL) framework employed in this study builds upon the approach outlined in our prior work. The Soft Actor-Critic (SAC) [8] algorithm was utilized, an off-policy method suitable for continuous action spaces, enabling efficient training through exploration and exploitation. The policy network

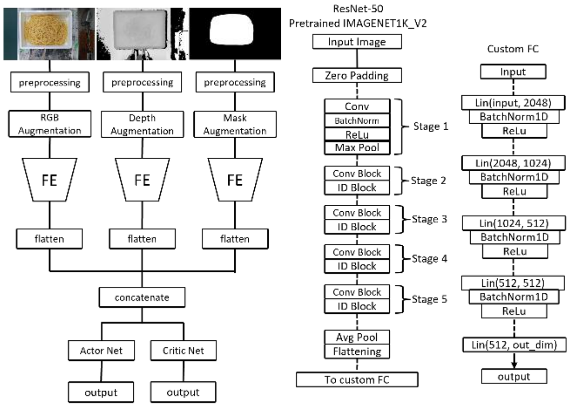


Fig.2 Net architecture for the feature extractor built on ResNet-50 with customized FC after the FE

architecture integrated a pretrained ResNet50 [11] model for feature extraction from RGB-D inputs, followed by a fully connected neural network for action generation as shown in Fig. 2. The action and observation spaces Tables 1 and 2 have been improved compared to the previous study. After obtaining the output action from the policy, we map it to the range required by the real hardware. For the gripper width, a simple mapping is applied from the original range to the operational range of 0–50 mm. To identify the grasping point, the action parameters corresponding to pixels i and j are first mapped to the pixel range. Using the depth information of the specified pixel (i, j), the distance is calculated. The pinhole camera model is then utilized to determine the grasping point in the robot's coordinate system. For the insert depth action, this value is mapped from the predefined lowest point to the depth of the pixel (i, j).

Table 1. The observation space of this study

No.	Mass matrix	Min	Max	unit
1	RGB			pixels
2	Depth	0	255	pixels
3	Mask			pixels
4	Actions	Refer to Table 2		
5	Noodle grasp weight	0.0	200.0	grams
6	Tray weight	0.0	1200.	grams

Table 2. The action space of this study

No.	Mass matrix	Min	Max	unit
1	Pixel i			-
2	Pixel j	-1.0	1.0	-
3	Insert depth			-
4	Gripper width			-

3.3. *Reward Function Design*

The reward function utilized in this study was derived from our prior research and optimized for weight-sensitive grasping tasks. It was based on a Probability Density Function (PDF), which penalized deviations from the target weight and rewarded as in Eq. (1).

$$r(x) = a \cdot \left[\frac{1}{\sigma\sqrt{2\pi}}\right] \cdot e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]}, -a \le r(x) \le a \quad (1)$$

Here, x represents the grasped weight, μ is the target weight, and σ is the acceptable deviation. The scaling factor a ensures the reward remains within a predefined range. Additional penalties are applied for collisions or failed grasps, encouraging efficient and precise actions. The function incentivized actions resulting in weights within a predefined range, penalizing grasps that exceeded or fell short of the target weight.

3.4. *Data Augmentation Pipeline*

To improve the robustness and generalization of the robotic manipulation model, a data augmentation pipeline is implemented, focusing on RGB images, depth maps, object masks, and tray weight. For RGB augmentation, variations in brightness, contrast, saturation, and hue are applied to simulate diverse lighting conditions and environmental changes during robotic operation. Depth augmentation introduces noise and small perturbations to emulate sensor inaccuracies and real-world inconsistencies in depth perception. Object mask augmentation involves adding noise to the edges, resizing, or slightly shifting the masks to account for inaccuracies in object detection, ensuring the model can handle imperfect segmentation outputs. Tray weight augmentation introduces variations to simulate different loading scenarios, teaching the model to adapt to varying weight distributions during grasping and placement tasks.

Each augmented sample is carefully synchronized across RGB, depth, and mask channels to maintain data consistency. These augmentations not only mimic real-world variabilities but also enhance the model's ability to learn robust grasping strategies under diverse scenarios. By applying this pipeline during training, the system gains improved resilience to sensor noise, lighting changes, and variations in object properties, ultimately enabling more reliable performance in real-world manipulation tasks. This is depicted in Fig. 3.

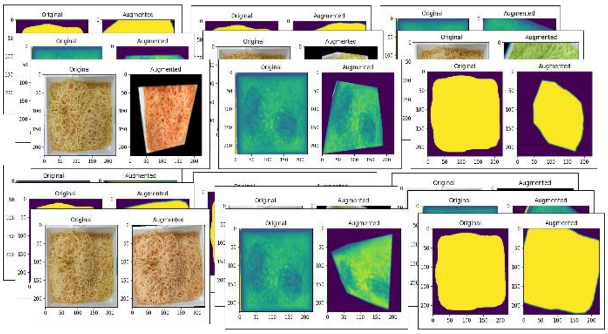


Fig.3 Data augmentation for the RGB, depth, and mask image for the observation

3.5. *Training Environment Setup*

The environment integrates RGB and depth data captured from a calibrated camera setup, along with object masks generated through segmentation techniques. These inputs are used to construct a 3D representation of the workspace, enabling precise action planning. The simulated workspace includes a tray placed within the robot's reachable area, where the robot performs grasping and placement tasks. The tray is equipped with a digital scale to measure the accumulated weight of the grasped spaghetti, providing a critical feedback signal for reinforcement learning. The robotic system operates within predefined constraints, including the gripper's range of motion, width, and insertion depth. These constraints are reflected in the action space, which maps policy outputs to real-world parameters using transformations based on the robot's kinematics and the pinhole camera model. This is depicted in Fig. 4.

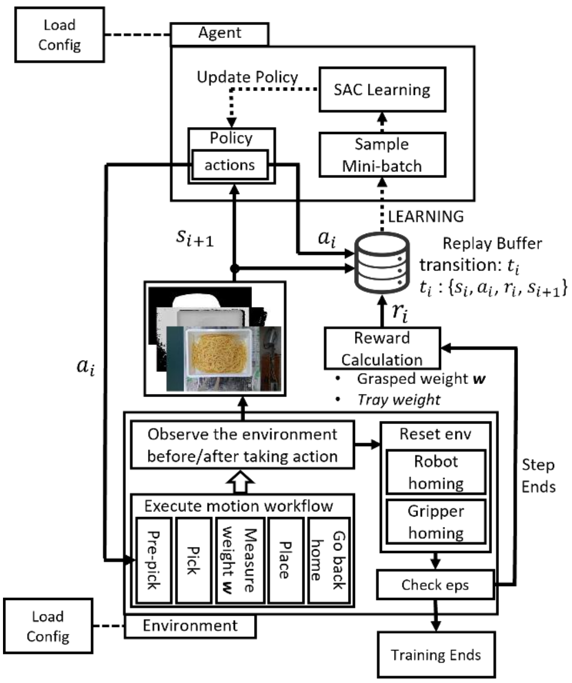


Fig.4 Diagram of how the system trains under the MDP-based method

The training environment also incorporates domain randomization to enhance generalization, varying properties such as lighting conditions, object appearance, and noise in the depth data. This ensures that the model learns robust policies that can handle real-world variabilities. The reward function is carefully crafted to balance weight accuracy, grasp stability, and execution efficiency, with penalties for dropping objects or exceeding weight limits. The environment supports episodic learning, resetting the state after each completed task or failure, and logging detailed metrics to monitor the model's performance. By combining realistic sensor inputs, diverse variations, and task-specific constraints, the training environment provides a robust foundation for learning effective policies for spaghetti manipulation in real-world scenarios.

4. Results and Discussion

4.1. *Experiment Result*

The experimental results demonstrate the effectiveness of the proposed system for autonomous robotic manipulation of deformable objects. The trained model achieved a grasp success rate (GSR) of 82%, indicating a high level of accuracy in successfully executing grasping tasks. The average grasp time (AGT) was measured at 16.99 seconds, reflecting the time efficiency of the robotic system in completing each grasp. The task efficiency (TE) was calculated as 0.0411 tasks per second, showcasing the throughput of the system. Additionally, the model was evaluated for its generalization capabilities, achieving a generalization score (GS) of 75% on new, unseen test data. This score suggests that the system can effectively adapt to new scenarios and conditions, further highlighting its robustness in real-world applications involving deformable objects like spaghetti.

4.2. *Analysis of Limitations*

Despite the promising performance of the proposed system, certain limitations were observed during the experiments, particularly when dealing with highly tangled spaghetti. In such cases, while the system is sometimes able to grasp the correct weight of the spaghetti, the tangling of the strands often interferes with the robot's ability to execute an accurate grasp. The tangles can cause the gripper to miss the target or improperly grasp multiple strands, reducing the overall grasp success rate. Additionally, the model's performance may be affected by the complexity of the entanglement, as the presence of overlapping strands introduces uncertainties that are not always accounted for in the training data. These challenges highlight the need for further improvements in the model's ability to handle highly deformable and tangled objects, which is crucial for ensuring consistent performance in real-world scenarios.

5. Conclusion

This study presents an autonomous robotic system for manipulating deformable objects, focusing on spaghetti using deep reinforcement learning (DRL). The model achieved a grasp success rate (GSR) of 82%, an average grasp time (AGT) of 16.99 seconds, and a task efficiency (TE) of 0.0411 tasks per second, demonstrating efficient performance. A generalization score of 75% indicates its adaptability to unseen scenarios. However, limitations arose when dealing with tangled spaghetti. While the system sometimes grasped the correct weight, entanglement often led to misgrasping or difficulty handling multiple strands. This issue highlights the challenge of working with tangled deformable objects.

Future work should focus on improving entanglement handling through better vision systems, real-time feedback, and expanded training data. In conclusion, while the system shows promise, overcoming tangling challenges is crucial for real-world deployment.

References

1. Gamolped, P., Koomklang, N., & Hayashi, E. (2024, February). Autonomous Spaghetti Grasping for Bento Box Assembly: A Deep Reinforcement Learning Approach. In 2024 1st International Conference on Robotics, Engineering, Science, and Technology (RESTCON) (pp. 181-186). IEEE.
2. Hu, J., Jiang, Y., & Weng, P. (2024). Revisiting Data Augmentation in Deep Reinforcement Learning. arXiv preprint arXiv:2402.12181.
3. Ma, G., Wang, Z., Yuan, Z., Wang, X., Yuan, B., & Tao, D. (2022). A comprehensive survey of data augmentation in visual reinforcement learning. arXiv preprint arXiv:2210.04561.
4. Endo, G., & Otomo, N. Development of a Food Handling Gripper Considering an Appetizing. In Book name IEEE International Conference on Robotics and Automation (ICRA).
5. F. Liu, F. Sun, B. Fang, X. Li, S. Sun, and H. Liu, Hybrid robotic grasping with a soft multimodal gripper and a deep multistage learning scheme', IEEE Transactions on Robotics, 2023.
6. A. Rajeswaran et al., Learning complex dexterous manipulation with deep reinforcement learning and demonstrations', arXiv preprint arXiv:1709. 10087, 2017.
7. H.-G. Cao, W. Zeng, and I.-C. Wu, 'Reinforcement learning for picking cluttered general objects with dense object descriptors', in 2022 International Conference on Robotics and Automation (ICRA), 2022, pp. 6358–6364.
8. Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018, July). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In the International conference on machine learning (pp. 1861-1870). PMLR.
9. Fan, J., Wang, Z., Xie, Y., & Yang, Z. (2020, July). A theoretical analysis of deep Q-learning. In Learning for dynamics and control (pp. 486-489). PMLR.
10. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.
11. Koonce, B., & Koonce, B. (2021). ResNet 50. Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization, 63-72.

Authors Introduction

Prem Gamolped

He received a bachelor's degree in engineering in 2021 from Robotics and Automation Engineering, King Mongkut's University of Technology North Bangkok in Thailand, and received a master's in 2022 from Kyushu Institute of Technology at Hayashi Laboratory. He is currently a PhD student at Kyushu Institute of Technology at Hayashi Laboratory and researches Robotics and AI.

Yon Pang Ja Sin

She obtained her bachelor's degree in mechanical engineering in 2021 from the Faculty of Engineering and Technology, Multimedia University, Malaysia. Currently, she is pursuing her master's degree at the Kyushu Institute of Technology, Japan, and is conducting research at the Hayashi Laboratory.

Vjosa Bytyqi

She received her bachelor's degree in computer engineering in 2023 from the University of Prishtina, Department of Electric and Computer Engineering. She is currently a research student at Kyushu Institute of Technology and conducts research at Hayashi Laboratory.

Prof. Eiji Hayashi

Prof. Eiji Hayashi is a professor in the Department of Intelligent and Control Systems at Kyushu Institute of Technology. He received the Ph.D. (Dr. Eng.) degree from Waseda University in 1996. His research interests include Intelligent mechanics, Mechanical systems, and Perceptual information processing. He is a member of The Institute of Electrical and Electronics Engineers (IEEE) and The Japan Society of Mechanical Engineers (JSME).

LiDAR-Enhanced Real-Time Tree Position Mapping for Forestry Robots

M.A Munjer *

Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-City, Fukuoka, 820-8502, Japan

Tan Chi Jie

Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-City, Fukuoka, 820-8502, Japan

Eiji Hayashi

Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-City, Fukuoka, 820-8502, Japan

Email: munjer.abul-md690@mail.kyutech.jp, tan.jie-chi339@mail.kyutech.jp, haya@ics.kyutech.ac.jp

Abstract

This article evaluates the effectiveness of an autonomous robot in creating a tree pose map using both simulated and experimental environments, validating its potential for forestry applications. This work also demonstrates the implementation of FastSLAM on a four-wheeled differential-drive robot, integrating real-time tree detection and tracking through LiDAR-based point cloud data. An algorithm is proposed to generate a map showing both the robot's path and detected tree positions during movement. Performance Metrics Analysis (PMA) revealed a high True Acceptance (TA) rate, confirming accurate tree position estimation. Experimental results validated the algorithm's reliability, showcasing strong distance accuracy with minimal discrepancies between actual and estimated positions. These findings highlight the system's potential for advancing forestry management through precise robotic navigation and mapping.

Keywords: LiDAR , DBSCAN, FastSLAM, Eccentricity

1. Introduction

Recent studies indicate, due to lack of labor force [1], mobile robots have the potential to substitute human labor in various sectors as forestry tasks. Among various Simultaneous Localization and Mapping (SLAM) methods, the particle filter-based FastSLAM approach has been demonstrated to accurately estimate a robot's trajectory while concurrently constructing a map of its surroundings, as described in study [2]. This method is particularly advantageous in scenarios involving dynamic or partially observable environments, where traditional SLAM techniques may struggle with computational efficiency or data association challenges. In the context of forestry robotics, study [3] on Online SLAM for Forestry Robots assumed the central positions of trees to correspond to the centers of clusters derived from LiDAR point-cloud data. However, due to the inherent limitations of LiDAR sensors, it was observed that trees are only partially visible within the limited scope of a single LiDAR scan, often leading to incomplete data capture for individual tree trunks. Additionally, study [4] noted that when robots operate near trees, only a smaller portion of the tree trunks is detected by the LiDAR sensor. This discrepancy causes significant deviations between the calculated cluster centers and the actual tree trunk centers, impacting the accuracy of SLAM and overall mapping fidelity. To address these challenges, it has been identified that a more precise and reliable method for estimating tree positions is essential for the successful application of FastSLAM in autonomous forestry robots. This research suggests improving tree position estimation by detecting tree trunks up to a specific height after separating other non-tree objects from the 3D point-cloud data generated by the LiDAR sensor. Such segmentation not only enhances the accuracy of the detected tree locations but also mitigates the effects of occlusions and noise in complex forestry environments. Furthermore, this study introduces a robust and real-time methodology for tree detection and tracking, leveraging point-cloud data from LiDAR sensors. The proposed approach integrates the capabilities of the Robot Operating System (ROS2) [5] for system management, Open3D for efficient point-cloud processing [6], and advanced filtering and clustering techniques to ensure the accurate identification of trees in diverse scenarios. By focusing on real-time processing, the method is designed for static forest environments, where tree positions remain relatively stable, allowing accurate mapping and tree identification based on point cloud data. This research not only contributes to improved tree mapping but also lays the groundwork for enhancing the navigation and task-planning capabilities of autonomous field robots in forestry applications.

2. System overview

The proposed forestry mapping system integrates multiple sensors and computational components to enable autonomous navigation and real-time mapping. Fig. 1 represents the system overview of this research where it relies on a differential-drive robot equipped with a Velodyne LiDAR sensor, an Intel RealSense Depth Camera, and an XSense IMU sensor. The Velodyne LiDAR captures high-density 3D point cloud data, crucial for detecting tree clusters and estimating their positions. Complementing this, the Intel RealSense camera provides RGB and depth data to enhance environmental perception, while the IMU sensor supplies orientation, angular velocity, and acceleration data to support accurate localization and robot stability. All sensor data is processed by a ROS2-based computational unit, which serves as the control hub for real-time operations.

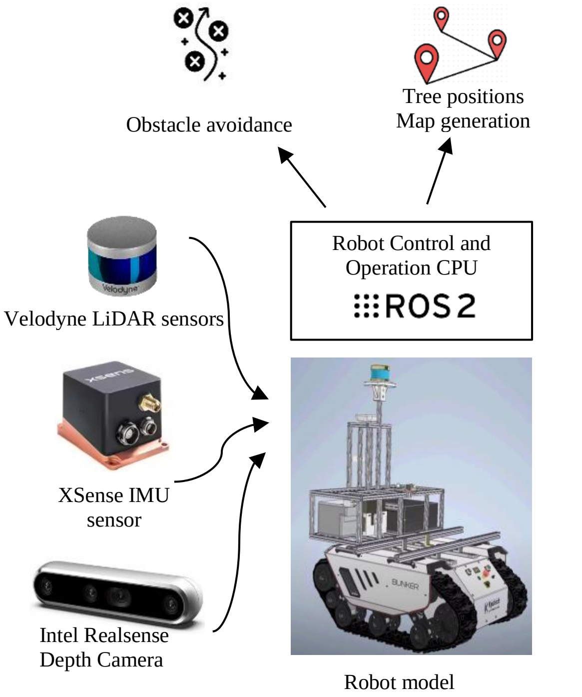


Fig. 1. Robot architecture and system overview for forestry mapping.

The system leverages LiDAR and camera inputs for tree detection and tracking while utilizing IMU data to stabilize navigation in uneven terrains. Key functionalities include obstacle avoidance, which ensures safe traversal through the forest, and tree position mapping, where the detected tree locations are transformed into a global frame for map generation. This architecture is designed for scalability and robustness, allowing the robot to operate in challenging forestry environments. The generated maps, displaying both the robot's path and tree locations, provide valuable insights into the environment, contributing to efficient forestry management and autonomous operations.

3. Methodology

3.1. Process Flow for Real-Time Tree Mapping

The system follows a structured pipeline for tree detection and mapping which have been shown in fig. 2. It begins with LiDAR and camera data collection, followed by preprocessing steps like noise filtering and down-sampling. Clustering is performed using DBSCAN to identify initial tree groups. These clusters are refined by calculating eccentricity through PCA. Tree positions are then mapped to a global frame, with Kalman Filter smoothing applied for accuracy.

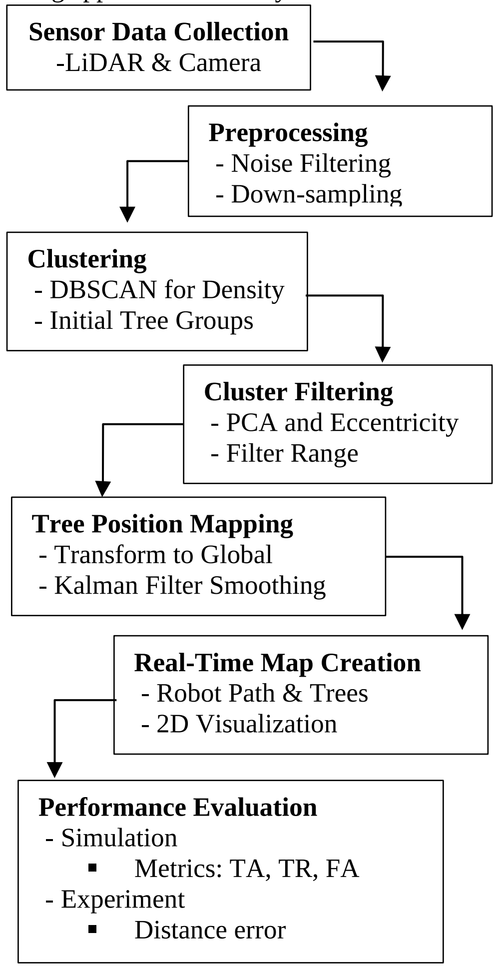


Fig. 2. Process flow diagram of the tree pose mapping.

The final output is a real-time 2D map combining the robot's path and tree locations. The system's performance is evaluated using metrics like TA, TR, FA, and FR under the Performance Metrics Analysis (PMA) framework. The details of the methodology can be described in the next sections.

3.2. Point Cloud Acquisition and Preprocessing

The initial stage of the methodology involves acquiring a 3D point cloud using a Velodyne LiDAR sensor mounted on mobile robot. This sensor generates a high-density array of points in 3D space by emitting laser pulses and capturing their reflections off surrounding objects [7], resulting in a set P of individual points. This point cloud data can be expressed as in Eq. (1)

$$P = \{p_i \mid p_i = (x_i, y_i, z_i) \in \mathbb{R}^3, i=1...,N\} \quad (1)$$

where each point p_i is represented by coordinates (x_i, y_i, z_i), with N denoting the total number of points. However, raw point clouds typically contain noise due to sensor inaccuracies or interference from environmental elements like foliage or small debris. To address this, a series of preprocessing steps have been applied to remove irrelevant data and reduce computational complexity. The first step in preprocessing is applying a pass-through filter along each axis, which removes points outside a specified spatial range. For instance, to filter along the x-axis, points is retained only $p_i \in P$ if $0 < x_i < \beta_x, -\beta_y < y_i < \beta_y, -\beta_z < z_i < \beta_z$ where β_x β_y and β_z are the distance threshold along the x,y and z-axis respectively in the LiDAR's frame. By selectively filtering points, the algorithm reduces data size and removes noise outside our region of interest. Following this, a voxel grid filter is applied to down-sample the point cloud further. The voxel grid divides the 3D space into small cubes (voxels) of size, averaging the points within each voxel into a single representative point. This creates a reduced set of down-sampled points, P′, which can be mathematically represented as in Eq. (2)

$$P' = \left\{ \frac{1}{|V|} \sum_{p_i \epsilon V} p_i \,|\, V \subset \mathrm{P},\ |\mathrm{V}| > 0 \right\} \qquad (2)$$

where V is a voxel containing multiple points, and |V| is the number of points within that voxel. This step not only reduces data volume but also preserves key structural features, allowing for efficient processing in later stages.

3.3. Tree Detection and Clustering

With a cleaned and down-sampled point cloud, the next stage involves identifying clusters within the data that likely represent tree trunks. To achieve this, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm has been applied [8], which is effective for clustering data with arbitrary shapes and noise. DBSCAN identifies clusters based on two main parameters: ϵ, the maximum distance between two points to consider them as neighbors, and minPts, the minimum number of points required to form a cluster. The clustering can be defined mathematically as $\mathrm{DBSCAN}(P', \epsilon, \mathrm{minPts}) \rightarrow C = \{C_1, C_2, \ldots, C_k\}$ where C_k represents a cluster of points satisfying DBSCAN's density-based criteria. These clusters are candidates for tree trunks, but additional filtering is required to confirm their shape. To illustrate the application of the mathematical principles of voxelization and DBSCAN clustering, Fig. 3 demonstrates the sequential processing of a noisy point cloud captured by a LiDAR sensor. As shown in Fig. 3(a), the raw point cloud data contains noise and extraneous points due to sensor inaccuracies and environmental interference. To reduce the computational complexity, the point cloud is voxelized, as shown in Fig. 3(b). The voxelization process divides the point cloud into uniform 3D grid cells, retaining a representative point for each occupied cell. This step significantly reduces the number of points while preserving the overall structure of the data. After voxelization, DBSCAN clustering is applied to identify dense regions and separate noise points.

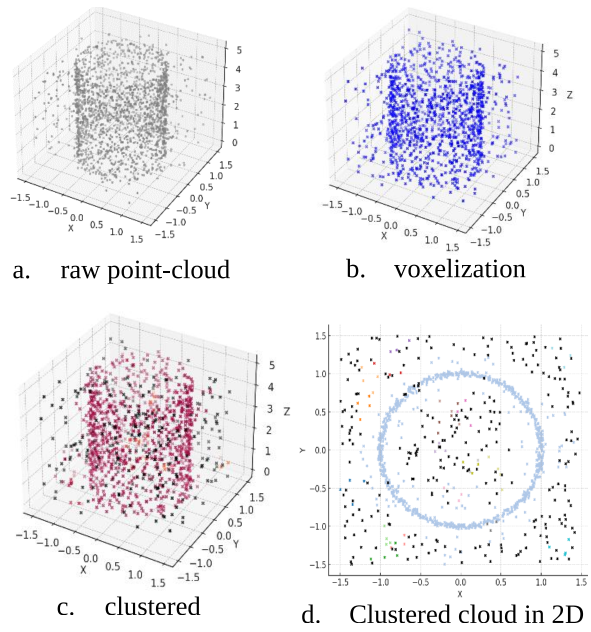


Fig. 3. Point-cloud processing for the cauterization.

Fig. 3(c) shows the output of DBSCAN, where each cluster is marked with a distinct color, representing isolated groups such as tree trunks. Noise points, which fail to meet the density criteria, are excluded. To simplify interpretation, the clustered point cloud is projected onto a 2D plane, as depicted in Fig. 3(d). This 2D visualization provides a clear representation of the spatial distribution of clusters, such as tree trunks, and aids in analyzing their positions relative to the environment. Once the clusters are identified, their shapes are analyzed to distinguish tree trunks from other objects. Eccentricity, a measure of how much a shape deviates from being circular, is calculated for each cluster. To assess the shape of each cluster and distinguish trees from other objects, eccentricity is being calculated using Principal Component Analysis (PCA) [9]. For each cluster C_k, the covariance matrix Σ has been computed, which quantifies how points in the cluster vary along each axis.

The covariance matrix is represented by Eq. (3)

$$\Sigma = \frac{1}{|C_K|} \sum_{p_i \epsilon C_k} (p_i - \bar{p})(p_i - \bar{p})^T \qquad (3)$$

where $\bar{p}$ represents the centroid of the cluster. An eigenvalue decomposition of Σ yields two principal eigenvalues, λ_1 and λ_2, with $\lambda_1 > \lambda_2$. Eccentricity e can be calculated as in equation Eq. (4).

$$e = \frac{\lambda_1}{\lambda_2} \qquad (4)$$

This ratio indicates how elongated or cylindrical a cluster is; higher values suggest a more tree-like, hyperbolic-cylindrical shape, while lower values imply a more rounded or irregular shape. Clusters with an eccentricity above a specific threshold are classified as trees, while others are discarded. In this research, it was observed that eccentricity values between 10 and 20 effectively classify tree-like hyperbolic clusters. This threshold aligns with the inherent geometry of partial tree stems, distinguishing them from noise or non-tree objects, which either exhibit lower eccentricity (more circular shapes) or significantly higher eccentricity (extremely elongated shapes). By exploiting the hyperbolic geometry of tree clusters, the proposed method ensures tree like shape detection.

3.4. SLAM and Trajectory Estimation

To localize the robot while mapping tree positions, FastSLAM has been implemented, an algorithm that combines particle filtering with landmark-based mapping[10]. FastSLAM maintains multiple particles, each representing a hypothesis of the robot's state and map. Each particle, denoted $S_t^{(k)}$, consists of a state hypothesis (the robot's estimated position and orientation) and a map hypothesis (detected tree positions), formulated as Eq. (5)

$$S_t^{(k)} = \left\{state_t^{(k)}, map_t^{(k)}\right\} \quad (5)$$

where k indexes the particle. The robot's state x_t at time t is updated based on control inputs (e.g., wheel velocities) and a process noise term ω_t, using the motion model can be express as in Eq. (6).

$$x_t = f(x_{t-1}, u_t) + \omega_t \quad (6)$$

Here, f(x,u) is a function describing how control inputs u_t affect the robot's position, and ω_t accounts for uncertainties in movement. Upon detecting a tree, each particle's map hypothesis is updated based on the observed tree position, with an observation model given by Eq. (7)

$$z_t = h(x_t, t_i) + v_t \quad (7)$$

where $h(x_t, t_i)$ is a function mapping the robot's state and the tree's local position to the predicted measurement, and v_t is measurement noise. Each particle is weighted based on how well its predicted map aligns with observed data, favoring particles that are consistent with detected trees.

3.5. Transformation of Tree Coordinates to Global Frame

Since detected tree positions are initially represented in the LiDAR's local frame, we need to transform them to the robot's global frame for consistency. Each detected tree position is part of a set $T_{velodyne}$ that can be represented as in Eq. (8).

$$T_{velodyne} = \{t_i \mid t_i = (x_i, y_i, z_i) \in \mathbb{R}^3, i=1\ldots,M\} \quad (8)$$

To represent these coordinates in the global odometry frame, we apply a series of transformations. The global transformation matrix T_{global_odom} is computed as following Eq. (9).

$$T_{global\_odom} = T_{velodyne} \cdot T_{base\_link} \cdot T_{odom} \quad (9)$$

where $T_{velodyne}$ represents the LiDAR data in its own frame, T_{base_link} maps the LiDAR frame to the robot's base, and T_{odom} maps the base frame to the global odometry frame. Each tree coordinate $t_i = (x_i, y_i, z_i) \in$ is transformed to $t_i' = (x_i', y_i', z_i')$ in the global frame, ensuring alignment with the robot's path and other map elements. By applying these transformations, each tree's position is now fixed on the global map. However, these global coordinates may still contain slight inaccuracies due to sensor noise or minor fluctuations in the robot's position.

3.6. Kalman Filter for Tree Position Estimation

Once the tree coordinates are transformed to the global frame, the Kalman Filter refines these positions by filtering out noise and providing a more stable estimate over time [11]. This step is essential because even with the global transformation, the tree positions can vary slightly due to measurement noise or sensor drift. The Kalman Filter treats each tree's global position as a state that is updated continuously, using a combination of prediction and measurement updates. In the Kalman Filter, the state transition matrix predicts each tree's future position based on the previous estimate, in each *dt* time step. This matrix relates the robot's previous state to its current state. Using the Kalman Gain the predicted state is adjusted according to measurement *z*, providing a more stable and accurate tree position estimate. The Kalman Gain *K* is a dynamic matrix that determines how much weight the filter should place on the new measurement relative to the prediction where *K* is calculated as in Eq. (10).

$$K = PH^T(HPH^T + R)^{-1} \quad (10)$$

The measurement matrix *H* maps the predicted state to the actual measurement space, enabling the filter to compare the predicted position with the observed position and **c**ovariance matrix *P* in a Kalman Filter represents the uncertainty in the state estimate and is typically initialized based on the expected initial uncertainty. Since our measurements only provide the position (not the velocity), H extracts the positional elements from the state whereas measurement noise covariance matrix R represents the uncertainty in the measurement process.

3.7. Real-Time Map Generation

With the tree positions refined in the global frame, the final step is to integrate these positions into a real-time map that also includes the robot's trajectory. The robot's path R_t, which is a sequence of positions over time based on odometry data, and the refined global tree positions T_{global} are now combined to create a cohesive, accurate map both of them can be represented as in Eq. (11) and Eq. (12).

$$R_t = \{r_j | r_j = (x_j, y_j) \epsilon R^2, j = 1, \dots, T\} \quad (11)$$
$$T_{global} = \{t_i^{'} | t_i^{'} = (x_i^{'}, y_i^{'}) \epsilon, i = 1, \dots, M\} \quad (12)$$

The combination of these elements results in a dynamic, real-time 2D map. This map continuously updates as the robot moves, reflecting both the path the robot has traveled, and the positions of trees detected in the environment.

4. Simulation and Experimental Results

The proposed algorithm was tested in a simulated environment using Gazebo with real-world scenarios. Key performance metrics, such as True Accepts (TA), True Rejects (TR), False Accepts (FA), and False Rejects (FR), were calculated to evaluate the system's effectiveness according to the analysis in study [12]. In this case TA and TR signifies number of trees correctly identified and amount of non-trees object correctly identified respectively. On the other hand, FA and FR signify the amount of non-tree objects mistakenly identified as trees and trees that were missed are estimated.

Table 1. Tuned Parameters and Threshold Values.

Parameter	Value(s)
number of particles for fastSLAM (unit)	100
eccentricity_threshold (unitless)	20
voxel_size (meter)	0.01
eps (DBSCAN) (meter)	0.25
min_points (DBSCAN) (number of points)	20
proximity_radius (meter)	1.0
cylindrical_stddev_thresh (meter)	0.1

Table 1 outlines various tuned parameters used in the algorithm, including a voxel size of 0.01m for down-sampling the point cloud and an eps value of 0.25m for DBSCAN clustering to define the neighborhood distance where voxel size of 0.01 means each voxel in the grid has a dimension of 0.01 units, reducing the point cloud density. Additionally, the pass-through filters have specific ranges for x (0 to 10), y (-10 to 10), and z (-0.5 to 0) axes, ensuring relevant points are retained and an eccentricity threshold of 20 are also specified to optimize the detection and tracking process.

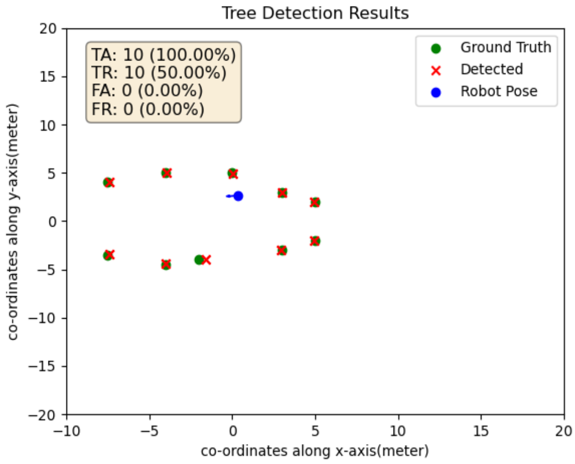


Fig. 4. Simulation result of tree detection accuracy.

The X and Y axes are represented in Fig. 4, the global coordinate system used to plot the positions of the robot and the trees, indicating the area covered by the robot's exploration. The results in Fig. 2 indicate that the system has a high rate of correctly identifying trees (TA: 100%) and non-tree objects (TR: 50%). However, the False Accepts rate (FA: 0%) is relatively low, suggesting that the system didn't rarely mistake non-tree objects for trees. The False Rejects rate (FR: 0%) indicates that no actual trees were missed, indicating good sensitivity. The percentage rate of TR could be due to the presence of objects in the environment that have similar shapes or features to trees, such as human or other vertical structures. On the other hand, the rate of TR as such might be attributed to occlusion, where trees are blocked by other trees or objects, or due to the limitations of the clustering and filtering algorithms used.

Fig. 5 represents the positions and trajectories of the robot navigating through a forest-like environment in Kyutech, Iizuka campus playground area. The red triangle marks the current position and orientation of the robot detected, dynamically tracked as it navigates the environment. The grey line traces the path that the robot has traveled, providing a visual representation of the robot's movement over time. Different colored circles represent confirmed objects in the environment, identified as trees. Table 2

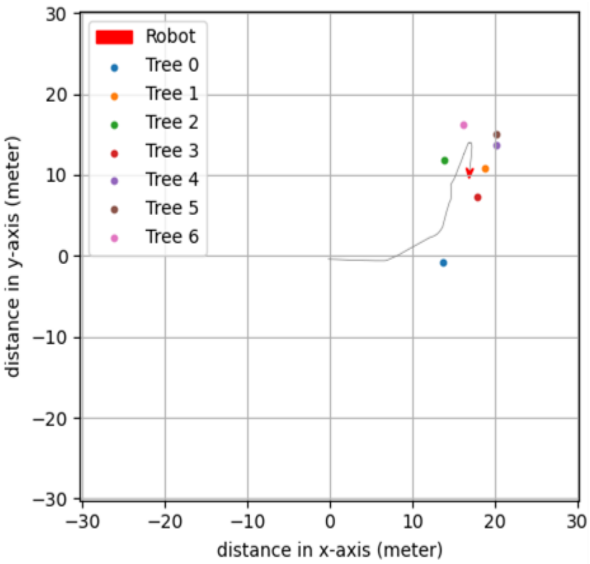


Fig. 5. Real time estimated map through experiment.

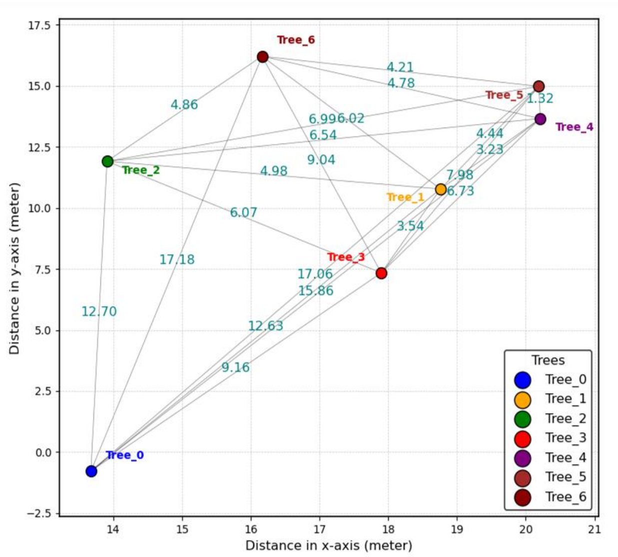


Fig. 6. Distance plot of detected tree pairs.

represents the coordinates and respective eccentricity of detected trees by the robot. Each tree is defined by its X and Y coordinates, indicating its location in the 2D mapping space, and an eccentricity value that characterizes the shape of the detected cluster associated with each tree.

Table 2. Tree position and their eccentricity values.

Tree Number	X	Y	Eccentricity
0	13.599	-0.782	11.539
1	18.74	10.778	14.58
2	13.91	11.912	10.003
3	17.897	7.7.342	16.457
4	20.201	13.664	11.066
5	20.177	14.983	10.765
6	16.02	16.214	12.993

The x and y coordinates reveal a varied spatial distribution, with positions ranging from approximately (13.599, -0.804) for "Tree 0" to (20.177, 14.963) for "Tree 5," suggesting a spread across a specific area. The eccentricity values, derived from Principal Component Analysis (PCA), quantify the elongation of each tree cluster, where higher values represent more elongated, cylindrical shapes indicative of tree trunks. For instance, "Tree 3" has an eccentricity of 16.457, reflecting a highly elongated shape, while "Tree 2" has a lower eccentricity of 10.003, indicating a less pronounced elongation. This data is critical for distinguishing tree-like structures from other objects in the environment, supporting accurate mapping and analysis in forestry applications by identifying clusters with shapes typical of tree trunks. From Fig. 6, it can be observed a visual representation of the spatial distribution and distances between selected trees within the experimental environment. Each tree is labeled from Tree 0 to Tree 6, with unique colors for easy identification. The graph illustrates both the x-axis and y-axis distances, with each inter-tree distance labeled, providing insight into the relative proximity of each tree. The distances between pairs are annotated, showing varying values based on spatial positioning. The lines connecting the trees indicate the paths used for calculating the distances, which are essential for mapping and navigation purposes. This visualization highlights the relative proximity of trees in the environment and provides a quantitative way to validate the accuracy of tree position detection. By comparing these distances, we can assess whether the detected tree positions align with the expected real-world configuration, which is crucial for verifying the algorithm's performance in tree detection and mapping. In Fig. 7, the matrix plot illustrates the percentage errors in distance measurements between various tree pairs. The average error of 7.13% highlights the effectiveness of the tree detection and mapping algorithm where most tree pairs exhibit minimal

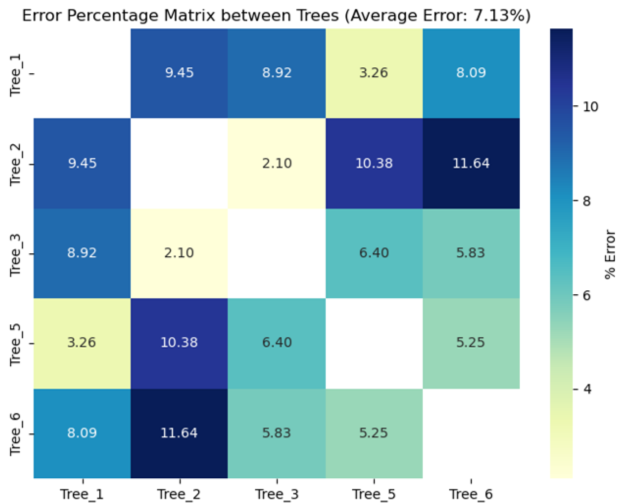


Fig. 7. Matrix plot of errors between measured and estimated distance.

error, indicating high accuracy and consistent detection. revealing higher errors in pairs such as "Tree 2 to Tree 6" (11.64%) and "Tree 2 to Tree 5" (10.38%). Several factors likely contribute to these discrepancies in a real-time forestry mapping context. First, sensor noise and interference from environmental elements like dense foliage and branches can scatter or absorb LiDAR beams, leading to inaccuracies, particularly over greater distances or when foliage obstructs the line of sight.

Additionally, tree occlusion and overlapping canopies pose challenges; trees may partially block each other from the sensor's view, which can distort perceived distances and increase errors for certain pairs. Real-time constraints add to this challenge, as rapid processing is required to maintain continuous updates, often involving approximations that prioritize speed over precision. Furthermore, terrain variability and sensor angles can introduce alignment issues; as the robot moves over uneven ground, slight tilts can misalign point cloud data, affecting distance estimates. In a forestry environment, these levels of error are generally acceptable given the inherent complexity and natural variability. For this navigation task, low rate of error suggest that the robot can effectively perceive its environment and navigate safely around obstacles without significant impact on overall performance. Overall, for real-time forestry mapping, the observed errors may be tolerable for general navigation. If higher precision is needed, though, further refinement in the algorithms or sensor calibration would be advisable to reduce the impact of environmental factors and improve accuracy.

5. Conclusion

The system shows strong performance in correct detections; however, challenges remain, such as distinguishing tree trunks from other cylindrical objects like poles or humans, which can lead to false positives. Environmental factors, including low-hanging branches or overlapping canopies, also introduce noise and hinder accurate detection. Despite these challenges, the low average error for experimental output demonstrates the algorithm's robustness and reliability in estimating tree positions. Future work could focus on integrating camera-based tree detection to complement LiDAR data, enabling more precise tree identification and classification. This multimodal approach could address current limitations, such as false positives from cylindrical objects, by leveraging advanced image-based classification techniques. Additionally, improvements in filtering techniques, clustering algorithms, and adaptive thresholding could enhance performance in complex forest environments. The generated maps remain a valuable tool for visualizing the robot's navigation and tree pose estimation capabilities, contributing significantly to the development of autonomous robots for forestry applications.

References

1. C. J. Tan, S. Ogawa, T. Hayashi, T. Janthori, A. Tominaga, and E. Hayashi, "3D Semantic Mapping based on RGB-D Camera and LiDAR Sensor in Beach Environment," *2024 1st International Conference on Robotics, Engineering, Science, and Technology, RESTCON 2024*, pp. 21–26, 2024.
2. Sebastian Thrun, Michael Montemerlo, Daphne Koller, Ben Wegbreit, Juan Nieto, and Eduardo Nebot, "FastSLAM: An Efficient Solution to the Simultaneous Localization And Mapping Problem with Unknown Data," *Journal of Machine Learning Research*, pp. 1–48, 2004.
3. Sylvain Geiser, Sakmongkon Chumkamon, Ayumu Tominaga, Takumi Tomokawa, Eiji Hayashi, Online SLAM for Forestry Robot, Journal of Robotics, Networking and Artificial Life, vol-9-2, pp-177-182, 2022-2023.
4. Geiser Sylvain, Chumkamon Sakmongkon, Tominaga Ayumu, Tomokawa Takumi, Jie Tan Chi, and Hayashi Eiji, "Practical Implementation of FastSLAM for Forestry Robot,Proceedings of International Conference on Artificial Life and Robotics," Feb. 2023, pp. 318–322.
5. S. Macenski, T. Foote, B. Gerkey, C. Lalancette, and W. Woodall, "Robot Operating System 2: Design, architecture, and uses in the wild," *Sci Robot*, vol. 7, no. 66, May 2022.
6. Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A Modern Library for 3D Data Processing," *arXiv:1801.09847*, pp. 1–6, Jan. 2018.
7. Tan Chi Jie, "The BCRobo dataset for Robotic Vision and Autonomous Path Planning in Outdoor Beach Environment," *ICAROB2023*, pp. 327–332, Feb. 2022.
8. D. Deng, "DBSCAN Clustering Algorithm Based on Density," in *2020 7th International Forum on Electrical Engineering and Automation (IFEEA)*, Hefei, China, 2020, pp. 949–953.
9. A. Maćkiewicz and W. Ratajczak, "Principal components analysis (PCA)," *Comput Geosci*, vol. 19, no. 3, pp. 303–342, Mar. 1993.
10. M. W. M. Gamini Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba, "FastSLAM: a factored solution to the simultaneous localization and mapping problem," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 3, pp. 229–241, Jun. 2001.
11. R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Fluids Engineering, Transactions of the ASME*, vol. 82, no. 1, pp. 35–45, 1960.
12. Alireza Baratloo, Mostafa Hosseini, Ahmed Negida, and Gehad El Ashal, "Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity," *Emerg (Tehran)*, vol. 3(2), p. 48, 2015.

Authors Introduction

M.A Munjer

He received his Masters of Science in Engineering from the Department of Electrical and Electronic Engineering, RUET, Bangladesh in 2019. He is currently a Doctoral student at Kyushu Institute of Technology and conducts research at Hayashi Laboratory.

Tan Chi Jie

He received his Masters of Science in Engineering from the Department of Computer Science and Systems Engineering, Kyushu Institute of Technology, Japan in 2023. He is currently a Doctoral student at Kyushu Institute of Technology and conducts research at Hayashi Laboratory.

Prof. Eiji Hayashi

Prof. Eiji Hayashi is a professor in the Department of Intelligent and Control Systems at Kyushu Institute of Technology. He received the Ph.D. (Dr. Eng.) degree from Waseda University in 1996. His research interests include Intelligent mechanics, Mechanical systems and Perceptual information processing. He is a member of The Institute of Electrical and Electronics Engineers (IEEE) and The Japan Society of Mechanical Engineers (JSME).

Kalman-YOLO Improving YOLO Tracking Performance through the Integration of a Kalman Filter for a Beach Cleaning Robot

Rut Yatigul
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan

Tan Chi Jie, Prem Gamolped, Eiji Hayashi
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan
Email: yatigul.rut663@mail.kyutech.jp, haya@mse.kyutech.ac.jp

Abstract

Ocean waste poses a significant threat to both human and marine life as industries and individuals continue to dump garbage into the ocean. Sea creatures are poisoned by materials such as plastics and chemicals, which in turn contaminate humans who consume them. This paper introduces an innovative approach using Image Instance Segmentation with YOLOv8 to segment and track beach garbage. However, YOLOv8's object tracking struggles in dynamic environments with challenges like occlusion, shadows, and perspective changes in RGB frames. To address this, the author presents Kalman-YOLO, combining the Kalman Filter with YOLO for improved performance. Results show notable performance improvement, especially in tracking garbage for the Beach Cleaning Robot.

Keywords: Computer Vision, Instance Segmentation, Kalman Filter, Multi Object Tracking, Data Association

1. Introduction

In the current era, object detection research has gained significant popularity, with tracking-by-detection becoming the leading trend in multiple object tracking [1]. Multi-object tracking is one of the fundamental tasks in computer vision, with a wide range of applications in fields such as intelligent security, autonomous driving, pedestrian tracking, and surveillance [2]. Object trajectories can be discovered as part of a global optimization problem that processes entire video batches at once, using methods such as probabilistic graphical models [3], [4], [5],[6] or flow network formulations [7], [8], [9]. The Kalman Filter, a linear Gaussian estimator, is optimal for certain applications. It was first introduced by Rudolf E. Kalman in his seminal 1960 paper titled *A New Approach to Linear Filtering and Prediction Problems* [10]. The Kalman Filter has been applied across diverse fields, including aeronautics, signal processing, and futures trading. The exponential increase in plastic production, coupled with varying efficiencies in global waste management systems, has led to large amounts of plastic waste entering the ocean annually [11]. In our research area, Munakata City, Fukuoka, Japan, beach litter ranges from tiny objects like microplastics to larger items such as fishing equipment [12]. Given the predictive capabilities of the Kalman Filter, the authors found it intriguing apply its theory to improve YOLO's object detection performance by integrating tracking capabilities. The cost matrix and Hungarian algorithm were used as matching algorithms to facilitate object re-identification during tracking.

With this in mind, the authors propose an innovative method combining the Kalman Filter, cost matrix, and Hungarian algorithm [13] to perform multi-object tracking and object re-identification. This method was tested and evaluated using the Hayashi Laboratory Beach Cleaning Robot and a custom garbage dataset collected from Munakata City, Fukuoka Prefecture, Japan.

2. Methodology

2.1 YOLOv8

YOLO (You Only Look Once) is a single-stage object detection algorithm initially introduced by Redmon and Farhadi in 2017 [14]. The operation process of YOLO divides the input image into grid cells and predicts a predefined number of bounding boxes and confidence scores for each cell. Additionally, YOLO predicts the class probabilities for each bounding box and combines them with the confidence scores to produce the final detection results.

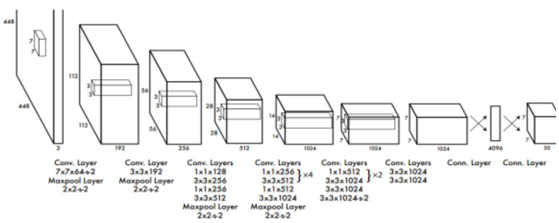


Fig.1. YOLO architecture [15]

Recently, YOLO has been adapted for instance segmentation, enabling precise boundary delineation for detected objects. Fig.1. illustrates the architecture of YOLO, beginning with a 448x448x3 input image. The image is processed through convolutional and max pooling layers, reducing the feature map to 7x7x1024. This is followed by two fully connected layers, resulting in a 7x7x30 output tensor that represents bounding boxes, objectness scores, and class probabilities. Garbage detection example shown in Fig.2.

Fig.2. Image segmentation based on garabge dataset

2.2 Track Handling and State Estimation

The Kalman Filter is currently used for handling track estimation. In this context, we assume a general tracking scenario that can be expanded as follows: the camera used for tracking is uncalibrated, and ego-motion data is unavailable. Such a situation presents significant challenges for the Kalman Filter prediction framework, yet it is a common scenario in recent object tracking benchmarks [15]. The prediction equation of the Kalman Filter is defined as follows:

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1} \quad (1)$$

$$\hat{P}_k^- = AP_{k-1}A^T + Q \quad (2)$$

Update equations are defined as follows:

$$K_k = \frac{P_k^- C^T}{CP_k^- C^T + R} \quad (3)$$

$$\hat{x}_k = \hat{x}_k^- + K_k(y_k - C\hat{x}_k^-) \quad (4)$$

$$P_k = (I - K_k C)P_k^- \quad (5)$$

Where K_k, $\hat{x}_k$, P_k, I is the Kalman gain matrix, the optimum filter value, To delve deeper into our tracking algorithm, we define a four-dimensional state space consisting of (x, y, v_x, v_y) t Each component is broken down as follows: (x, y) represent the position of the object in the x and y planes of the image frame, while (v_x, v_y) denote the velocity of the object along the x and y axes. The Kalman Filter we use is a standard implementation with a constant velocity motion model, and the observation model is also linear. Expanding on our observation model, (x, y) positional information is extracted from the centroid coordinates of the detected object provided by YOLO. This positional data is then fed into the filter as a measurement update.

2.3 Data Assignment Algorithms.

When dealing with a multi-object tracking problem, data association is crucial. This is because we have multiple detected object position data and multiple Kalman Filter-predicted object positions. To ensure continuous prediction and updating of newly incoming positional data, we need to match these two sets of information. To address this challenge, we introduce a cost matrix and the Hungarian algorithm, which work together to calculate matching pairs between predicted object positions and newly detected positions.

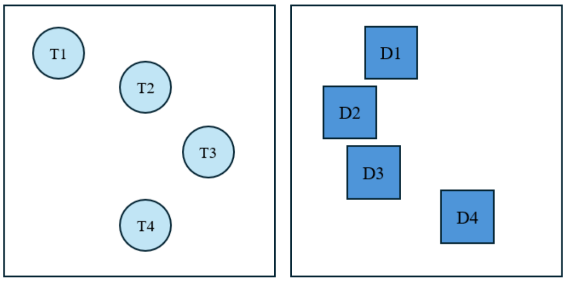
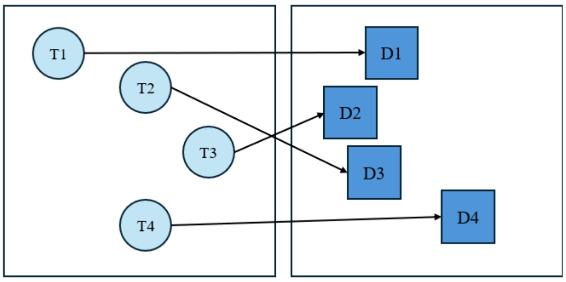


Fig.3. Diagram example of associating track and detections. (a) Tracks at time t-1, (b) Detection at time t

To determine the optimal assignment between tracks and detections, a cost matrix is crucial for representing the cost of each possible association between track positions and detections [16]. Various metrics can be used to calculate these costs; however, in our experiment, we opted to use the Euclidean distance. The Hungarian algorithm processes the cost matrix as input, performing a series of operations to simplify the assignment process [16]. Table.1. illustrates the optimal assignment, where cells with 1s indicate the mapping of tracks to detections. Fig.3. and Fig.4. demonstrates the association of objects across consecutive frames. For coding convenience, we chose to use the SciPy library, which provides the linear_sum_assignment() method. This method implements the Hungarian algorithm and returns the row and column indices corresponding to the optimal assignment [16].

	D1	D2	D3	D4
T1	1	0	0	0
T2	0	0	1	0
T3	0	1	0	0
T4	0	0	0	1

Table.1. Optimal assignment for the cost matrix



(b) (b)

Fig.4. Final mapping example of track and detections. (a) Tracks at time t-1, (b) Detection at time t

3. Results and Discussion

3.1 Dataset

The dataset used in this study was created by combining the publicly available TACO dataset [17], which focuses on general trash, with images captured by us in Kanezaki, Munakata, Fukuoka, Japan. TACO is a comprehensive dataset containing diverse examples of trash in natural environments, featuring various locations beyond beaches, with a total of 1,500 images. Additionally, 97 images were specifically taken at cleanup sites to supplement the dataset. To improve its effectiveness in dynamic environments, we enhanced the dataset with various annotation techniques. This was particularly important given the challenges of working in uncontrolled outdoor conditions, such as varying lighting throughout the day and motion blur caused by the beach cleaning robot's movement. The final dataset consists of 3,845 images.

3.2 Object tracking and Identity switch

The primary goal of our object tracking program was to reduce identity switches caused by detection loss in the YOLOv8 detection algorithm. Detection loss is a common issue in object detection programs that rely on RGB image processing. One major cause of this problem is occlusion, along with changes in illumination and object size between frames due to different camera perspectives. Since we are working with 2D images, the information captured by the camera is inherently limited compared to 3D imaging, as it only provides a 2D perspective. This limitation introduces several points of potential error in object classification and detection. For instance, in one frame at time t-1, the algorithm might detect a PET bottle, but at time t, the image of the bottle could be distorted (e.g., sheared), causing the detection algorithm to fail to recognize it. At time step t+1, the algorithm might detect the same bottle again, but without a robust tracking program, it would treat the bottle as a new instance, leading to false object counts. This issue is referred to as an identity switch. Since the objective of our research is to develop a program for garbage tracking and counting, addressing this issue is our top priority.

Garbage Type	Detection loss	ID switch
PET bottle	18	0
Styrofoam	21	1
Wood	7	1
Shoe	10	0

Table.2. Results of ID switch quantities compared to detection loss across multiple frames

As shown in Table 2, after implementing the tracking algorithm, the number of identity switches was significantly reduced compared to detection losses across multiple frames. The test was performed on the researchers' laboratory laptop, equipped with an Intel Core i9-12900H processor running at 2.5 GHz, 16 GB of RAM, and an NVIDIA GeForce RTX 3070 Ti Laptop GPU. The test conducted by slowing down the object detection program to 3 frames per second and observing each frame to determine how often the algorithm failed to continuously detect the same object as it moved across the frame. It was observed that the detection algorithm experienced frequent losses. Despite this, the Kalman Filter effectively used its motion model to predict the possible positions of objects accurately. Once a prediction was made, the cost matrix calculated the association cost between each track and detected object using Euclidean distance. Finally, the Hungarian algorithm was applied in the last step to match each track with a detected instance. Overall, the performance of the tracking algorithm met our requirements for the garbage tracking and counting application.

4. Conclusion

This research aims to develop a garbage tracking and counting algorithm using a Kalman Filter for tracking and predicting objects' future positions in cases of detection loss across multiple video frames. The algorithm incorporates a cost matrix and the Hungarian algorithm for data association. The YOLOv8 detection algorithm served as the foundation for the tracking program, and the detection model was trained on our custom dataset. The final results demonstrated satisfactory performance, with a small number of identity switches. For future work, we aim to implement fuzzy logic to make the maximum Euclidean distance threshold adaptive. This improvement would account for uncertainties in the detected position and the reappearance probability of each object in the multi-object tracking algorithm.

References

1. Wojke, N., Bewley, A., & Paulus, D. (2017, September). Simple online and realtime tracking with a deep association metric. In *2017 IEEE international conference on image processing (ICIP)* (pp. 3645-3649). IEEE.
2. Ciaparrone, G., Sánchez, F. L., Tabik, S., Troiano, L., Tagliaferri, R., & Herrera, F. (2020). Deep learning in video multi-object tracking: A survey. *Neurocomputing*, *381*, 61-88.
3. B. Yang and R. Nevatia, "An online learned CRF model for multi-target tracking," in CVPR, 2012, pp. 2034 2041.
4. B. Yang and R. Nevatia, "Multi-target tracking by on line learning of non-linear motion patterns and robust appearance models," in CVPR, 2012, pp. 1918–1925.
5. A. Andriyenko, K. Schindler, and S. Roth, "Discrete continuous optimization for multi-target tracking," in CVPR, 2012, pp. 1926–1933.
6. A. Milan, K. Schindler, and S. Roth, "Detection- and trajectory-level exclusion in multiple object tracking," in CVPR, 2013, pp. 3682–3689.
7. L. Zhang, Y. Li, and R. Nevatia, "Global data associa tion for multi-object tracking using network flows," in CVPR, 2008, pp. 1–8.
8. H. Pirsiavash, D. Ramanan, and C. C. Fowlkes, "Globally-optimal greedy algorithms for tracking a vari able number of objects," in CVPR, 2011, pp. 1201 1208.
9. J. Berclaz, F. Fleuret, E. T¨ uretken, and P. Fua, "Multi ple object tracking using k-shortest paths optimization," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 9, pp. 1806–1819, 2011.
10. Kalman, R. E. (1960). A new approach to linear filtering and prediction problems.
11. Zhu, X., Rochman, C. M., Hardesty, B. D., & Wilcox, C. (2024). Plastics in the deep sea–A global estimate of the ocean floor reservoir. *Deep Sea Research Part I: Oceanographic Research Papers*, *206*, 104266.
12. Ogawa, S., Jie, T. C., Tomokawa, T., Geiser, S., Chumkamon, S., Tominaga, A., & Hayashi, E. (2023). Development of Beach Litter Detection System using Deep Learning on Beach Clean-up. *Journal of Advances in Artificial Life Robotics*, *3*(4), 238-241.
13. Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval research logistics quarterly*, *2*(1-2), 83-97.
14. Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7263-7271).
15. A. Milan, L. Leal-Taix´ e, I. Reid, S. Roth, and K. Schindler, "Mot16: A benchmark for multi-object tracking," arXiv preprint arXiv:1603.00831, 2016.
16. R. Sadli, Beginner's Guide to Multi-Object Tracking with Kalman Filter . 2023.
17. "TACO Dataset ", http://tacodataset.org/

Authors Introduction

Mr. Rut Yatigul

He received his Bachelor's degree in Engineering in 2023 from Robotics and Automation Engineering, King Mongkut's University of Technolgoy North Bangkok in Thailand. He is currently a Master student in Kyushu Institute of Technology at Hayashi Laboratory and conducts research about Robotics and AI.

Mr. Tan Chi Jie

He received his Bachelor of Engineering Electronics Majoring in Robotics and Automation from the Faculty of Engineering, Multimedia University, Malaysia in 2020 and received masters in 2023 from Kyushu Institute of Technology at Hayashi Laboratory. He is currently a PhD student at Kyushu Institute of Technology and conducts research at Hayashi Laboratory.

Mr. Prem Gamolped

He received bachelor degree in Engineering in 2021 from Robotics and Automation Engineering, King Mongkut's University of Technology North Bangkok in Thailand and received masters in 2022 from Kyushu Institute of Technology at Hayashi Laboratory. He is currently a PhD student in Kyushu Institute of Technology at Hayashi Laboratory and conducts research about Robotics and AI.

Prof. Eiji Hayashi

Prof. Eiji Hayashi is a professor in the Department of Intelligent and Control Systems at Kyushu Institute of Technology. He received the Ph.D. (Dr. Eng.) degree from Waseda University in 1996. His research interests include Intelligent mechanics, Mechanical systems and Perceptual information processing. He is a member of The Institute of Electrical and Electronics Engineers (IEEE) and The Japan Society of Mechanical Engineers (JSME).

The research of AR System for introducing Industrial Robots

Takuma Aiko
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-city, Fukuoka, 820-8502, Japan

Eiji Hayashi, Prem Gamolped
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-city, Fukuoka, 820-8502, Japan

Email: aiko.takuma723@mail.kyutech.jp

haya@ics.kuytech.ac.jp,

gamolped.prem129@mail.kyutech.jp

Abstract

In recent years, Japan has been suffering from a labor shortage in all industries. By introducing robots, it is possible to reduce manpower, and it is expected to contribute to resolving labor shortages. However, the introduction of industrial robots is not easy due to the high cost of equipment and system integration. Therefore, we are developing an Augmented Reality (AR) application for the purpose of introducing robots. In this study, we developed a mobile AR system that can check the movement path of a robot when it is introduced without using the actual robot and confirmed its operation.

Keywords: AR, Factory Automation Robots, Unity, ROS

1. Introduction

In recent years, Japan has been suffering from a labor shortage in all industries, not only in the manufacturing industries. The main cause of the labor shortage is said to be the decrease in the working population. By introducing robots to production lines instead of people, it is possible to reduce manpower and improve productivity, and it is expected to contribute to resolving labor shortages. On the other hand, the introduction of robots into a company requires a huge amount of money for system development, installation, and maintenance, as well as for the robots themselves and related equipment. Therefore, the enormous cost (expense and time) is a major barrier to the introduction of robots.

In order to compensate for the shortage of manpower, we develop an autonomous lunchbox serving system using reinforcement learning (Fig. 1). To reduce the enormous cost, we also develop an AR system using 3D data of a robot arm.

AR is a technology that makes it appear as if virtual information that does not exist in the real world is actually present in the real world by overlaying it onto the real world through a mobile device camera or other device. This time we will introduce a 3D model of a robot arm and then move the model. By running this system in a factory, it is expected that it will be possible to easily check traffic flow on-site. In this study, we developed a mobile AR system that can check the movement path of a robot when it is introduced without using the actual robot and confirmed its operation.

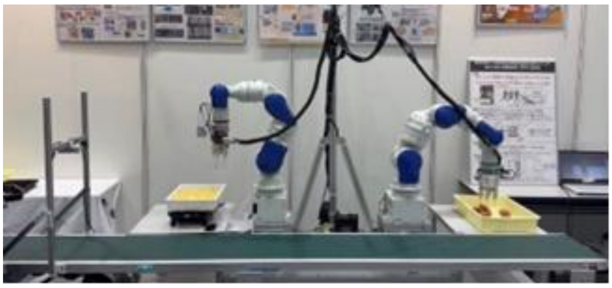
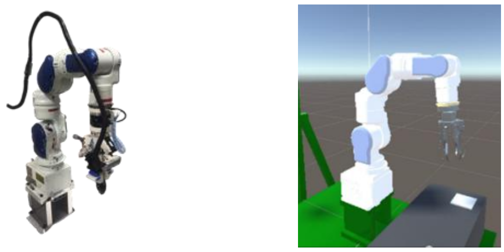


Fig. 1. A self-learning robot for the food industry

2. Methodology

2.1. *Robot Arm*

Yaskawa Electric's MOTOMAN-SIA5F (Fig. 2), we also introduced a 3D model of the same robot arm to the mobile AR system because we are developing a lunch-serving system using same robots.



a. MOTOMAN-SIA5F b. 3D model

Fig. 2. Robot model used in the study

2.2. *Development environment and framework*

The system consists of two main platforms, Unity and ROS (Robot Operating System)[1][2]. Since Unity provides AR Foundation as a framework for AR development, we used it for AR development. ROS provides a variety of functions, one of which is MoveIt, which plans the motion of the robotic arm's hand position and finds the optimal path to reach the target coordinates and orientation. We applied this functionality to a 3D model of a robotic arm in an application [3].

2.3. *Communication*

Unity and ROS use TCP communication to exchange messages. On the ROS side, ROS-TCP-Endpoint, a package for creating an endpoint to exchange messages from Unity, was installed. On the Unity side, ROS-TCP-Connector, a package for creating a connector to exchange messages from ROS.[3]

2.4. *System Overview*

An overview of the system is shown in Fig. 3. The mobile terminal is equipped with a project developed in Unity and converted into a file executable on the mobile terminal.

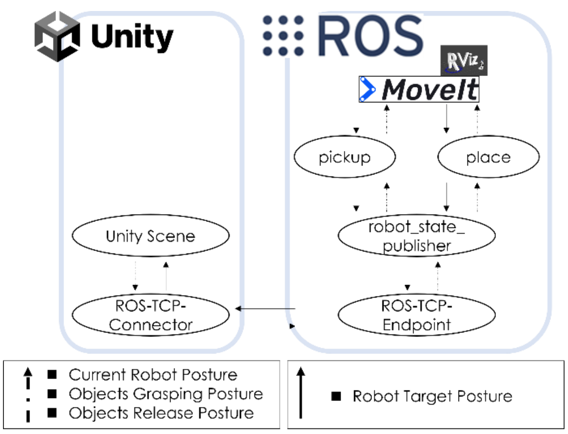


Fig. 3. System Overview

3. Experiment

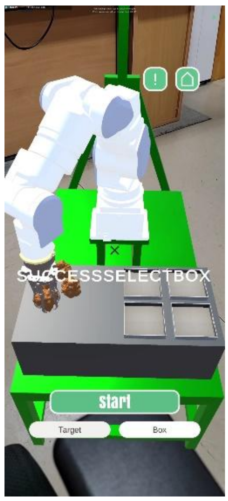
3.1. *Operation Check*

We checked that the mobile AR system we created was working properly. First, the flow of the created mobile AR system is shown below. When the application is launched, the home screen is first displayed. When the screen moves to the next screen, the camera of the mobile terminal is activated and the robot arm, girders, grippers, and objects to be grasped (hereinafter referred to as "objects") appear on the screen. When the "Start Simulation" button is pressed, a simulation using a 3D model of the robot arm starts after exchanging messages with the ROS.

The result of the operation after pressing the start button on the mobile terminal is shown in the following Fig. 4.

a. Iitial Sate

b. State of grasping an object



c. State in which an object is placed

d. State of operation completed

Fig. 4. Mobile AR system screen

In this case, the application flow is as follows: a, b, c and d.

a. Initial Cost
   The camera on the mobile device is activated and the simulation can begin.

b. State of Grasping an Object
   The robot moves to the position coordinates where the object is located and grasps it with the gripper.

c. State in which an Object is Placed
   The robot arm moves the object to the target coordinates and releases the gripper.

d. State of operation Completed
   The operation is completed, and the robot arm returns to the initial state.

4. Conclusion

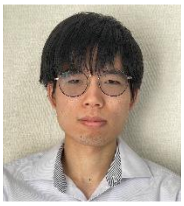
Through this experiment, we were able to confirm that the robot arm was able to perform the desired operation. Therefore, it was found that the trajectory of the robot arm's hand position up to grasping the object was calculated and simulated more appropriately than with MoveIt cargo. On the other hand, in this mobile AR system, the robot appears in front of the camera when it is activated, so once it appears, the position cannot be adjusted. Solutions include the creation of a UI that allows the robot's position to be adjusted and the introduction of AR markers.

References

1. ROS, Open Robotics, https://www.ros.org/
2. Unity, Unity Technologies, https://unity.com/
3. ROS-TCP-Endpoint, Unity-Technologies, https://github.com/Unity-Technologies/ROS-TCP-Endpoint

Authors Introduction

Mr. Takuma Aiko



He received his Bachelor's degree in Engineering in 2024 from the Faculty of Engineering, Kyushu Institute of technology in Japan. He is currently a master student in Kyushu Institute of Technology, Japan

Mr. Gamolped Prem

He received bachelor degree in Engineering in 2021 from Robotics and Automation Engineering, King Mongkut's University of Technolgoy North Bangkok in Thailand and received masters in 2022 from Kyushu Institute of Technology at Hayashi Laboratory. He is currently a PhD student in Kyushu Institute of Technology at Hayashi Laboratory and conducts research about Robotics and AI.

Prof. Eiji Hayashi

Prof. Eiji Hayashi is a professor in the Department of Intelligent and Control Systems at Kyushu Institute of Technology. He received the Ph.D. (Dr. Eng.) degree from Waseda University in 1996. His research interests include Intelligent mechanics, Mechanical systems and Perceptual information processing. He is a member of The Institute of Electrical and Electronics Engineers (IEEE) and The Japan Society of Mechanical Engineers (JSME).

Research on performance information editing support system for automatic piano
- Development of a network model for improved dynamics accuracy-

Taiyo Goto
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan

Yoshiki Hori, Eiji Hayashi
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan
Email: gotou.taiyou711@mail.kyutech.jp, haya@mse.kyutech.ac.jp

Abstract

The automatic piano player, which was previously developed in this laboratory, is attached to the keys and pedals of a grand piano, and enables accurate keystrokes and pedal operation with appropriate control from a computer. To control the device, music data is required, but if music score data is simply input into the device, the performance will be flat, and will not sound like a human being, which is the goal. This is because pianists play with their own intonation when they play. Previous research has developed a system that uses deep learning to predict performance information, but the accuracy of predicting sound volume (Velo) was not good. This research aims to enhance the accuracy of Velo in automatic piano performance. A new deep learning system combining two networks was developed to address limitations in existing methods.

Keywords: Deep Learning, Automatic Piano, Drop out, Skip connection

1. Introduction

In order for the automatic piano player [1] developed in this laboratory to perform with human-like intonation, MIDI (Musical Instrument Digital Interface) standard data with intonation for each note is required. Therefore, inputting score data as it is into the device as-is does not result in human-like performance, which is the objective. Therefore, previous research has developed an inference system using deep learning to represent MIDI-standard data with intonation. However, it was difficult to reproduce the intensity of the sound. To solve this problem, we developed a deep learning model that combines two networks. This paper describes the inference system developed and the inference results.

2. Data used

The training data used was performance data recorded in accordance with the MIDI standard. This data includes parameters of performance information that represent the pianist's performance expression, as well as musical notation information such as notes and dynamics. However, it is difficult to use the MIDI standard data as-is in machine learning, so we extracted the performance information and score information to be used from the performance data and set the parameters so that they can be used in machine learning.

2.1. Performance information

Among the performance data, information that is expressed differently by different pianists is defined as performance information, and four parameters are set. Table 1 shows the performance information.

Table 1 Performance information

Parameters	Detail	Unit
Velo	Sound intensity	-
Gate	Sound length	ms
Step	Interval until next note	ms
Time	Sound time	ms

2.2. *Score information*

We defined score information as information in the performance data that is uniquely determined regardless of the pianist, such as notes and dynamics, and set five parameters.
Table 2 shows the score information.

Table 2 Score information

Parameters	Detail	Unit
Tgate	Sound length on the score	ms
Tstep	Interval until next note on the score	ms
Key	Sound height	-
Bar	Bar number	-
Dyn	Dynamics mark	-

3. Inference System

Fig.1 shows the inference system created for this project. This system was created by combining two networks. Both Network 1 and Network 2 have three intermediate layers. The flow of inference results until they are output is as follows: First, music information is input to Network 1, and then the music information and Velo output by Network 1 are input to Network 2. The final Velo value is obtained from these two networks. Considering the possibility of overlearning, skip connections and dropouts were applied to each intermediate layer.

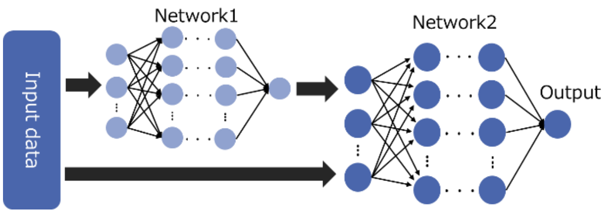


Fig.1 Network layout diagram

4. Experiments

4.1. *Experimental details*

Eight pieces performed by the pianist Vladimir Davidovich Ashkenazy were prepared, one of which was treated as test data, and the remaining seven pieces were used to infer the test data. The actual test data and the inference results are then compared. In this case, Prelude Op.28-7 was used as the test data.

4.2. *Data partitioning for cross-validation*

The hyperparameters of the model were determined by increasing the data set using K-fold cross-validation. By using this method, the evaluation of the model does not depend on the division of the data, and more reliable results can be obtained. Specifically, the eighth song was fixed as the test data, and the other seven songs were divided into five training data and two validation data, for a total of 21 different data sets. The divided data is shown in Table 3.

Table 3 Pattern of data

Pattern number	Train data	Evaluate data	Test data
1	1,2,3,4,5	6,7	8
2	1,2,3,4,6	5,7	8
3	1,2,3,4,7	5,6	8
4	1,2,3,5,6	4,7	8
5	1,2,3,5,7	4,6	8
		⋮	
21	3,4,5,6,7	1,2	8

4.3. *Experimental Results*

Fig.2 shows a graph of Velo output by the inference system and Velo in the actual performance data. The graph of Velo output by Network 1 is also shown in Fig.3.

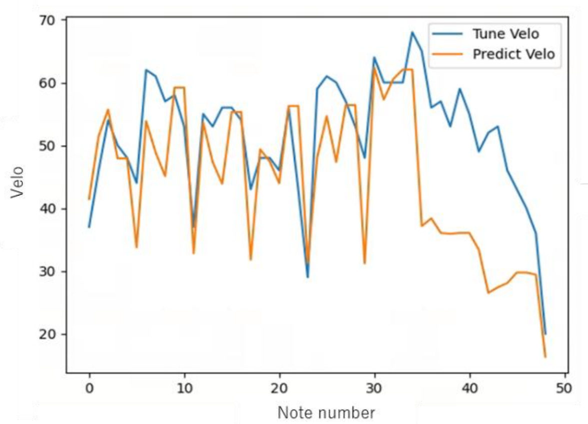


Fig.2 Inference results in the two-stage system

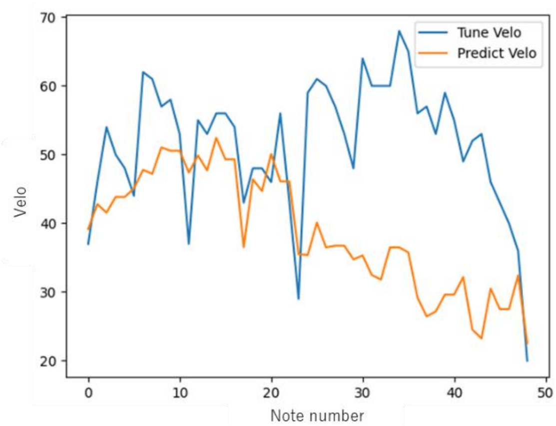


Fig.3 Inference results in the first stage system

5. Consideration

Fig.4 shows the inference results of the previous study. In the previous study, a two-layer inference system was used for the middle layer.

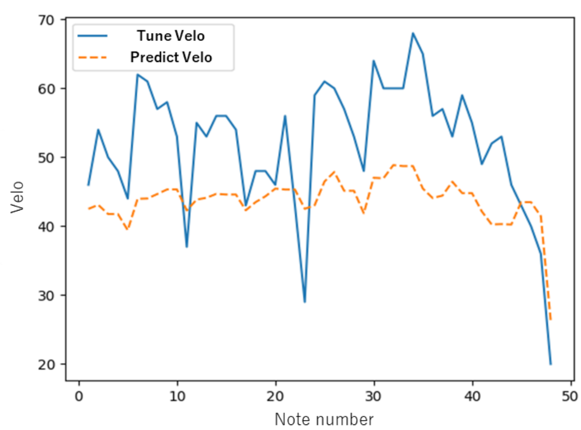


Fig.4 Inference results in previous research

The results of inference using the system combining the two networks in Fig.2 and 4 show that the waveforms of the graphs up to around musical score number 35 are more similar than when inference was performed using the system in the previous study. This is because the features obtained in Network 1 contributed to the output of Network 2, amplifying the change in the waveform, although the change in the graph of Network 1 in Fig.3 is low. However, the waveforms of the graphs after score number 35 are not similar. In particular, the waveforms of Network 1 are less similar to those of the previous study. Network 2 has more correspondence than Network 1, but it does not correspond to the part where the sound suddenly becomes quieter at the end. This may be because the parameters set for the large fluctuation at the end were not appropriate.

6. Conclusion

This study attempted to improve the accuracy of inference of sound intensity by using a model that combines two networks. Experimental results showed that the two network systems produced graphs with waveforms similar to those of the performance data, compared to previous studies. The future prospect is to develop an inference system that can handle large fluctuations. For this purpose, we will verify the improvement of inference accuracy by directly acquiring strength and weakness symbols on the score, which are not included in the MIDI standard data, using image recognition, and inputting them as new parameters into the system.

References

1. E. Hayashi, M. Yamane, H. Mori, Development of a moving coil actuator for an automatic piano, Int. J. Japan Soc. Prec. Eng. 28 (1994), 164–169.

Authors Introduction

Mr. Taiyo Goto

He received bachelor degree in Engineering in 2024 from mechanical system engineering, Kyushu Institute of Technology in Japan. He is currently a Master student at Kyushu Institute of Technology and conducts research at Hayashi Laboratory.

Mr. Yoshiki Hori

He received bachelor degree in Engineering in 2023 from mechanical system engineering, Kyushu Institute of Technology in Japan. He is currently a Master student at Kyushu Institute of Technology and conducts research at Hayashi Laboratory.

Prof. Eiji Hayashi

Prof. Eiji Hayashi is a professor in the Department of Intelligent and Control Systems at Kyushu Institute of Technology. He received the Ph.D. (Dr. Eng.) degree from Waseda University in 1996. His research interests include Intelligent mechanics, Mechanical systems and Perceptual information processing. He is a member of The Institute of Electrical and Electronics Engineers (IEEE) and The Japan Society of Mechanical Engineers (JSME).

Research on Tactile-Gripping for Difficult-to-Grasp Objects

Yoshitaka Sakata
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan

Gamolped Prem, Eiji Hayashi
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan

Email: sakata.yoshitaka414@mail.kyutech.jp, gamolped.prem129@mail.kyutech.jp, haya@ics.kyutech.ac.jp

Abstract

This study focuses on the automation of food preparation and boxing in the food manufacturing industry. An important point of food grasping by robots is that the shape of the food should not be damaged. However, it is difficult for a conventional robot hand to perform this task perfectly. Therefore, an end-effector equipped with a camera-based tactile sensor has been developed to perform this task in previous studies. However, the performance of this end-effector depends on the reflectance of the target object, since it estimates contact based on the reflectance of light. We have developed a camera-based tactile sensor and contact estimation system to solve this problem. In addition, we have developed a pickup motion combined with object detection.

Keywords: Soft Robotics, Tactile Sensors, Deep Learning

1. Introduction

This study focuses on the automation of food preparation and boxing in the food manufacturing industry. In recent years, the need for robot automation has been increasing due to the expected worsening of labor shortages caused by the decline in the working-age population. Serving and boxing soft foods is a challenging task for robots. In a previous study, soft and deformable objects were defined as “difficult-to-grasp objects,” and an end-effector equipped with a vision-based tactile sensor was developed to perform these tasks. Although this end-effector has excellent ability to grasp difficult-to-grasp objects, it has a problem that accurate estimation of the grasping state becomes difficult when the light reflectance of the object is low. We developed a vision-based tactile sensor and grasp state estimation system to improve this problem. In addition, we have developed a pickup motion combined with object detection.

2. Robot Composition

The robot consists of a vertically articulated 6-axis arm, an RGB-D camera, and a hand tip equipped with a tactile sensor as shown in Figure 1.

The camera-based tactile sensor consists of a finger with Fin Ray structure, rubber-coated elastic skin, LED strip, and RGB camera as shown in Figure 2. Since it is necessary to grasp the object without crushing it, we designed it based on the structure proposed in [1]. For the placement of the elastic skin, LEDs, and cameras, we refer to [2] and [3]. The RGB camera takes pictures of the elastic skin and captures the deformation of the skin due to contact with an object. The contact is shown in Figure 3.

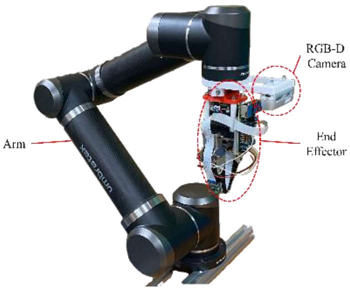


Fig.1 Robot Overview

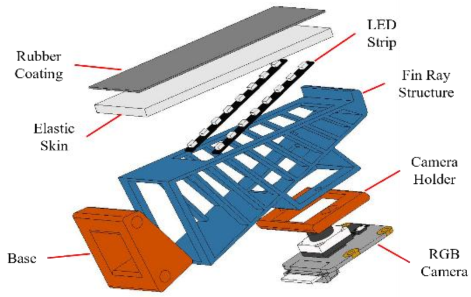


Fig.2 Composition of camera-based tactile sensor

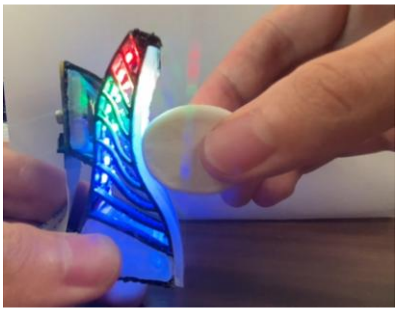


Fig.3 Deformation of tactile sensor due to object contact

3. System configuration

Figure 4 shows the system configuration. The system consists of object detection, grasp state estimation, motion planning, and paw control, and uses ROS (Robot Operating System) topic communication for inter-process communication.

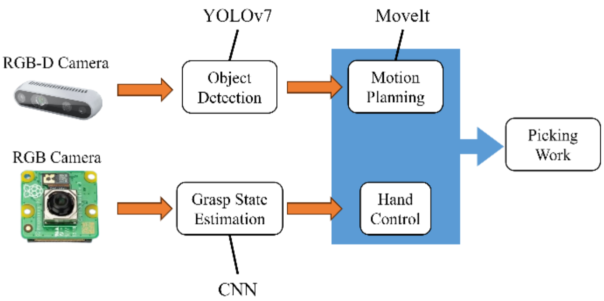


Fig.4 Pickup System Configuration

3.1. Object Detection

The algorithm uses YOLOv7, a real-time object detector. One hundred marshmallow images were used as the dataset, and 300 training runs were conducted. The algorithm was trained 300 times on a dataset of 100 marshmallow images. Object detection was performed using the trained model, and the 3D coordinates of the object were calculated by combining the center pixel of the detected object (Figure 5), depth information, and internal camera parameters.

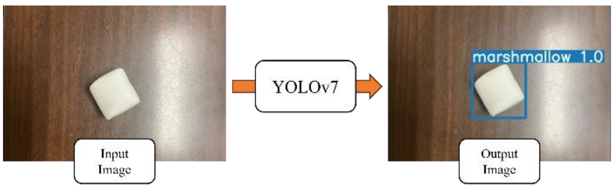


Fig.5 Object detection results

3.2. Motion Planning

The coordinates calculated by object detection are converted to a coordinate system based on the robot. MoveIt (motion path planning library) is used to plan the motion using this as the target position of the robot's hand.

3.3. Contact State Estimation Model

The grasping state estimation model is shown in Figure 6, where a convolutional neural network consisting of a convolutional layer, an all-joint layer, and an output layer was created using VGG-16 with RGB images as input. Here, the target of grasping is a marshmallow, and 214 images in the non-contacted state and 250 images in the contacted state were used for fine tuning.

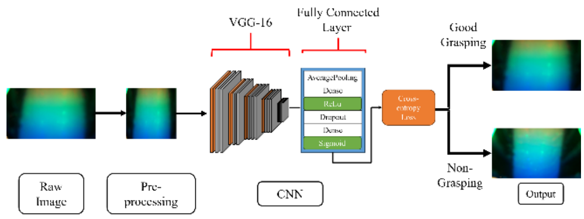


Fig.6 Contact State Estimation Model

3.4. End-Effector Control

The hand control using the estimation model is shown in Figure 7, where the video from the RGB camera is used as input for estimation. The model closes the fingers if they are not grasped and maintains the position of the fingers if they are in contact.

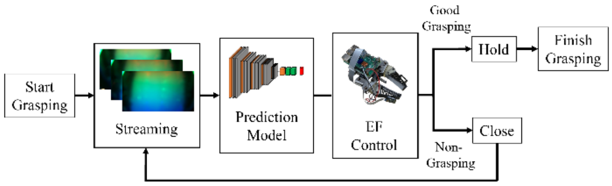


Fig.7 End-effector control using an estimation model

4. Experiment

Experiments were conducted to measure the performance of tactile sensors on cylindrical and ball-shaped marshmallows. Figure 8 shows an overview of the grasping experiment when the marshmallow was in contact with elastic skin areas A, B, and C, each about 11 mm in diameter. The size of the marshmallow before grasping and the distance between the fingers after grasping were measured 10 times each, and the amount of deformation of the marshmallow before and after grasping was calculated.

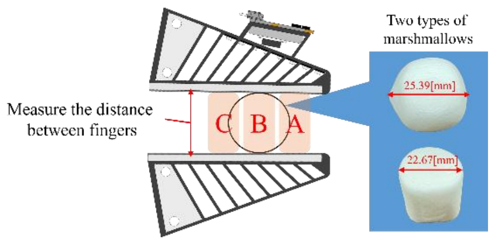


Fig.8 Experiment Overview

5. Results

Table 1 shows the average deformation and the ratio of deformation to the size of the marshmallow before grasping. It can be confirmed that the deformation of the ball-shaped marshmallow is larger than that of the cylinder-shaped marshmallow in the grasping of the ball-shaped marshmallow. The smallest deformation was observed in region C for both shapes. In the case of the cylindrical marshmallow, the deformations in regions A and B are almost the same, and in the case of the ball-shaped marshmallow, the difference between the deformations in regions A and B is large.

Table 1 Experimental Results

Contact Surface	Deformation (Cylinder)	Deformation Rate (Cylinder)	Deformation (Ball)	Deformation Rate (ball)
A	0.648	2.815	0.999	3.928
B	0.639	2.825	0.879	3.366
C	0.447	1.9997	0.827	3.359

6. Conclusion

We developed a vision-based tactile sensor and grasp state estimation system for grasping difficult-to-grasp objects. This system was then applied to the control of an end-effector to verify the performance of the tactile sensor on two different shapes of marshmallows.

The performance of the tactile sensor was verified for two shapes of marshmallows, and its effectiveness in grasping difficult-to-grasp objects was confirmed. Future work will include verification of the Fin Ray structure and the modification of the elastic skin and RGB camera positions. In addition, we aim to develop a deep learning system with multiple data inputs, such as a fingertip image from another camera in addition to the RGB camera image.

References

1. Emerson, J., Elgeneidy, K., (2020). Optimising Soft Fin Ray Robotic Fingers using Finite Element Analysis to Reduce Object Slippage. UKRAS20 Conference: "Robots into the real world" Proceedings, 43-45.
2. S. Q. Liu and E. H. Adelson, "GelSight Fin Ray: Incorporating Tactile Sensing into a Soft Compliant Robotic Gripper," *2022 IEEE 5th International Conference on Soft Robotics (RoboSoft)*, Edinburgh, United Kingdom, 2022, pp. 925-931.
3. S. Q. Liu, Y. Ma and E. H. Adelson, "GelSight Baby Fin Ray: A Compact, Compliant, Flexible Finger with High-Resolution Tactile Sensing," 2023 IEEE International Conference on Soft Robotics (RoboSoft), Singapore, Singapore, 2023, pp. 1-8.

Authors Introduction

Mr. Yoshitaka Sakata

He received bachelor's degree in Engineering in 2024 from Intelligent and Control Systems, Kyushu Institute of Technology in Japan. He is currently a Master student at Kyushu Institute of Technology and conducts research at Hayashi Laboratory.

Mr. Gamolped Prem

He received bachelor degree in Engineering in 2021 from Robotics and Automation Engineering, King Mongkut's University of Technolgoy North Bangkok in Thailand and received masters in 2022 from Kyushu Institute of Technology at Hayashi Laboratory. He is currently a PhD student in Kyushu Institute of Technology at Hayashi Laboratory and conducts research about Robotics and AI.

Prof. Eiji Hayashi

Prof. Eiji Hayashi is a professor in the Department of Intelligent and Control Systems at Kyushu Institute of Technology. He received the Ph.D. (Dr. Eng.) degree from Waseda University in 1996. His research interests include Intelligent mechanics, Mechanical systems and Perceptual information processing. He is a member of The Institute of Electrical and Electronics Engineers (IEEE) and The Japan Society of Mechanical Engineers (JSME).

Development of a drone obstacle avoidance system based on depth estimation

Sora Takahashi
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan

Eiji Hayashi
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan
Email: takahashi.sora614@mail.kyutech.jp, haya@mse.kyutech.ac.jp

Abstract

This study developed an obstacle avoidance system for drones using depth estimation from RGB cameras, aiming to reduce reliance on expensive sensors like RGB-D cameras or LiDAR. The system employs the deep learning model ZoeDepth for depth estimation and integrates it with ROS and Gazebo for simulation. Two autonomous systems were evaluated: one using RGB-D cameras and the other using depth estimation with RGB cameras. Experimental results show that while the RGB-D camera system outperformed in accuracy, the depth estimation-based system provided cost-effective and reasonable performance, especially in complex environments. The research concludes with plans to improve the system for denser obstacle environments and conduct real-world experiments.

Keywords: Deep Learning, Autonomous vehicle, Presumption of Depth,

1. Introduction

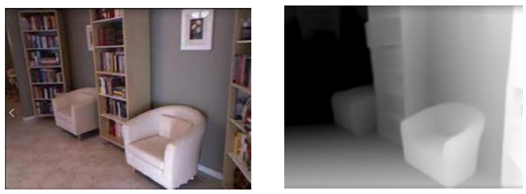
In recent years, drone surveys have been conducted at disaster sites and on remote islands. In such locations, it is difficult for a person to operate a drone in close proximity due to safety, labor, and financial costs. Although there are advanced technologies that allow drones to follow predetermined routes remotely, there is the problem of collisions with drones when there are unexpected obstacles along the route. In addition, there are few drones equipped with sensors that can measure a wide range of distances, such as RGB-D cameras and LiDAR, which are necessary for obstacle detection. In this study, I developed and evaluated an autonomous system that can avoid obstacles using an RGB camera installed on a common drone.

2. System Overview

In this study, I develop two autonomous moving systems with different sensor inputs: the first is an autonomous moving system with RGB-D camera sensor input, and the second is an autonomous moving system with RGB camera sensor input. The first is to evaluate the system proposed in this study. The distance information necessary for obstacle detection is obtained from the RGB cameras by depth estimation.

2.1. Depth Estimation Model

The depth information was estimated using ZoeDepth, a depth-learning model that estimates depth from a single image [1]. The model is a metric depth estimation model, which is built by fine-tuning a model trained on a relative depth dataset with a metric depth dataset. Fig1 shows the RGB image and (b) shows the depth image estimated from the image in (a).

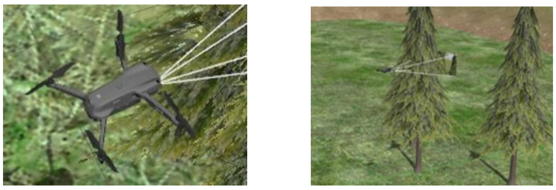


(a)RGB image (left) (b) Depth portrait (right)

Fig 1Example of depth presumption

2.2. Building a Simulation Environment

The simulation environment was built on the assumption that DJI's DJI Air 2S is handled as actual equipment. The simulation environment was built on the assumption that DJI's DJI Air 2S would be used as the actual device. Gazebo, ROS (Robot Operating System), and ArduPilot were used to construct the simulation environment. Gazebo provides a 3D robot simulation environment that uses a physics engine to simulate the environment and the robot's movements in real time, while ROS handles communication and data processing between devices and provides system flexibility and scalability. Fig2 shows the drone and environment projected on Gazebo.



(a) Drone (left) (b) Environment (right)

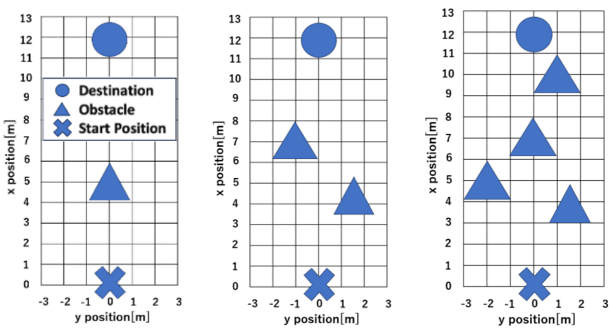
Fig 2 Simulation Environment

2.3. Autonomous Mobility Systems

For autonomous movement and obstacle avoidance, I used the ROS Navigation Stack, a collection of frameworks and tools that provide a set of functions necessary for autonomous robot navigation through the environment. The Navigation Stack is a collection of tools and frameworks that allow the robot to move autonomously through its environment. RGB-D images are used as input to RTAB-Map [2]. The user then sets the destination in the autonomous system, and the drone moves toward the destination. The generated map of the surrounding environment is updated as needed, and obstacles are detected in real time and appropriate route corrections are made.

3. Experiment

Experiments were conducted in a simulation environment to see if the two developed obstacle avoidance autonomous systems could avoid obstacles and reach the set destination. As shown in Fig3, the robot's initial position (x,y,z)[m],(0,0,3) was set as the origin, and the destination (12,0,3) was set. In each of the three obstacle situations, the robot made three autonomous movements to the destination in two patterns, and evaluated whether it could avoid the obstacles in the two patterns, and the error between the destination and the actual location.



(a) Situation 1 (b) Situation 2 (c) Situation 3

Fig 3 Obstacle situation

4. Experimental Results

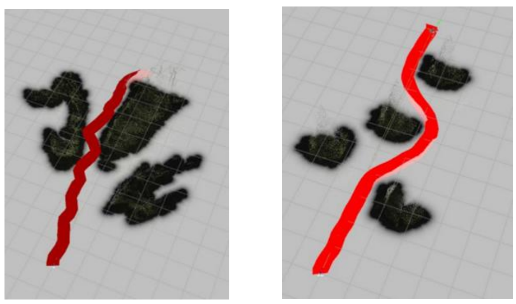
table1 and table2 summarize the experimental results of the autonomous moving system using depth estimation and the autonomous moving system using the RGB-D camera, respectively. table1 and table2 show that the system was able to avoid obstacles in all obstacle situations. In all obstacle situations, the autonomous system with the RGB-D camera had a smaller error with the destination. The autonomous system using depth estimation showed smaller errors in relation to the destination as the number of obstacles increased. Fig4 shows the path taken by the drone and the results of the drone's recognition of obstacles in obstacle situation 3 for each of the autonomous systems tested. Fig4 shows that the autonomous moving system using depth estimation has a larger area recognized as the same obstacle than the autonomous moving system using the RGB-D camera.

table 1 Autonomous mobility system using depth estimation

Situation	Avoiding Obstacle	Destination Error(Average)[m]
Obstacle Status1	◯	6.2334[m]
Obstacle Status2	◯	3.9669[m]
Obstacle Status3	◯	2.2863[m]

table 2 Autonomous moving system using RGB-D camera

Situation	Avoiding Obstacle	Destination Error(Average)[m]
Obstacle Status1	◯	0.4951[m]
Obstacle Status2	◯	0.5881[m]
Obstacle Status3	◯	0.3669[m]



(a) Depth estimation of Drone trajectory (left)

(b) Drone trajectory by RGB-D camera (right)

Fig 4 Drone trajectory and obstacle recognition results

5. Consideration

The accuracy of the estimated depth was lower than the depth obtained from the RGB-D camera. It is thought that fluctuations occurred in the distance from obstacles based on depth estimation, resulting in a wider area recognized as an obstacle, as shown in Fig4 (a). In addition, the autonomous moving system used visual information from the camera and depth information to estimate the drone's self-position. The autonomous movement system uses visual information from the camera and depth information to estimate the drone's self-position. In this case, I believe that errors in depth information accumulated as the drone moved, resulting in errors between the destination and the location where the drone arrived. table1 shows that the more obstacles there are, the smaller the error between the destination and the autonomous system using depth estimation. This is because the number of feature points such as edges and corners in the image increases as the number of obstacles increases, and the accuracy of self-position estimation increases as a result. To improve the accuracy of self-position estimation, I believe that the drone's self-position estimation can be achieved by using sensors such as GPS and IMU and integrating these sensors, instead of using only visual and depth information.

6. Conclusion

In this study, a simulation environment was constructed. I also developed and evaluated an obstacle avoidance system for drones based on depth estimation. As a result, the system was able to recognize and avoid obstacles, although the accuracy of self-position estimation was less than that of the RGB-D camera system. In the future, I will develop an obstacle avoidance system for environments with more obstacles and conduct experiments in real environments.

References

1. Bhat, S. F., Birkl, R., Wofk, D., Wonka, P., & M ¨ uller, M. , ZoeDepth: Zero-shot Transfer by Com-bining Relative and Metric Depth , arXiv preprint arXiv:2302.12288, 2023.
2. b´e, M., & Michaud, F. RTAB-Map as an Open-Source Lidar and Visual SLAM Library for Large-Scale and Long-Term Online Operation. Journal ofField Robotics, 36(2), 416-446. 2019

Authors Introduction

Mr. Sora Takahashi

He received bachelor degree in Engineering in 2024 from mechanical system engineering, Kyushu Institute of Technology in Japan. He is currently a Master student at Kyushu Institute of Technology and conducts research at Hayashi Laboratory.

Prof. Eiji Hayashi

Prof. Eiji Hayashi is a professor in the Department of Intelligent and Control Systems at Kyushu Institute of Technology. He received the Ph.D. (Dr. Eng.) degree from Waseda University in 1996. His research interests include Intelligent mechanics, Mechanical systems and Perceptual information processing. He is a member of The Institute of Electrical and Electronics Engineers (IEEE) and The Japan Society of Mechanical Engineers (JSME).

Practical Exercise on An Autonomous Driving System Using Mobile Devices and IoT Devices for An Agricultural Tractor

Daigo Katayama
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Yuto Nakazuru
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hikaru Sato
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Shoun Masuda
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Yuya Nishida
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Shinsuke Yasukawa
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Kazuo Ishii
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Email: katayama@brain.kyutech.ac.jp, nakazuru.yuto211@mail.kyutech.jp, sato.hikaru292@mail.kyutech.jp, masuda.shoun528@mail.kyutech.jp, y-nishida@brain.kyutech.ac.jp, s-yasukawa@brain.kyutech.ac.jp, ishii@brain.kyutech.ac.jp
https://www.kyutech.ac.jp

Abstract

The current method of agriculture is expected to make sustainable production difficult due to the effects of a declining and aging workforce. To solve these issues, research and development of smart agriculture technologies, including automated tractor operation, have been underway. We have developed an autonomous driving system for a commercially mini-tractor using mobile and IoT devices to more facilitate the introduction of automated driving technology for tractors. In addition, the exercise was conducted for students to implement and operate this system with an aim of education for robotics engineers. This exercise consists of lectures and development exercises for the system. The demonstration results showed that the students' difficulty level was able to set appropriate difficulty. The results of the rubric also showed that the students improved their proficiency in the algorithms and programming of applications that are fundamental to autonomous driving systems.

Keywords: Smart agriculture, Autonomous driving, Tractor, Mechatronics exercise

1. Introduction

According to the survey by the Ministry of Agriculture, Forestry and Fisheries in Japan, the number of core agricultural workers is expected to decrease to 1/4 of the current level (from 1.16 million to 0.30 million) over the next 20 years [1]. Therefore, improving the efficiency of agricultural work and reducing the manpower for agricultural work has been required. To solve this problem, a production method called "smart agriculture" that applies robotics and Internet-of-Things (IoT) technology has been proposed, and a lot of research and development has been conducted. We have focused on an autonomous driving system for tractor, one of contents of smart agriculture. Tractors can perform many agricultural tasks. Therefore, we think that an autonomous operating system makes it possible to reduce time required for agricultural work and perform multiple tasks simultaneously. A tractor with an autonomous driving system has already been put into practical use and commercialized [2], [3], [4]. Also, installing automatic steering devices as an automation method for common tractors [5] has been promoted. We also think that exercises for training engineers who can develop autonomous driving systems that can be introduced by agricultural workers are needed for promotion of smart agriculture. Interdisciplinary education for human resources development for an autonomous driving system is required [6]. The exercise for an autonomous driving system is one example of small-scale RC cars being used as targets for implementation [7], but there are few cases of tractors being used as targets. Therefore, we have designed an exercise based on the development of an autonomous driving system for a tractor. This paper reports on the outline, the equipment used, and the results of the exercise.

2. Exercise

2.1. *Outline*

This exercise has been conducted as part of the exercises for the "Joint Graduate School Intelligent car, Robotics & AI" [8]. This joint graduate school has been the collaboration between the University of Kitakyushu, Waseda University, and our university since 2019 (The previous joint graduate school began in 2009). This exercise is a short, intensive two-week course, and students taking part include not only students from joint universities but also students from National College of Technology and bachelor students from other universities. This year period is the first year that this exercise has been conducted.

In this exercise, we hold lectures and practices on system development to develop an autonomous driving system for a tractor equipped with an autonomous driving unit, as shown in Fig. 1 and Fig. 2. The hardware required for the system configuration, the mini-tractor, tablet, and autonomous driving unit, are provided. Students try to learn their understanding of mechatronics, semiconductor devices, and communications through lectures and acquire the method for developing applications, sensing, and motion control to realize an autonomous driving system through practice.

2.2. *Mechatronics exercise*

In the mechatronics exercises, students practiced turning on LEDs using digital I/O and PWM, acquiring sensor values using A/D conversion and UART communication, and controlling servo motors by circuits with Arduino microcomputers, which are also used in the autonomous driving unit. We also hold practice analog circuits using operational amplifiers and full-bridge motor drive circuits using FETs to understand the semiconductor devices used in the exercise. Additionally, we hold lectures and practice on network communications used in the exercise.

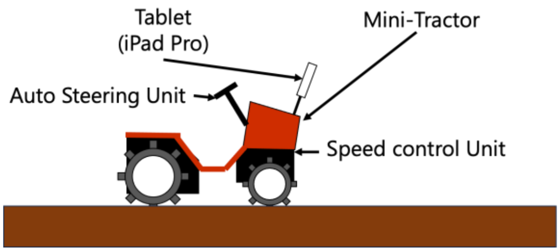


Fig. 1 Overview of autonomous driving system

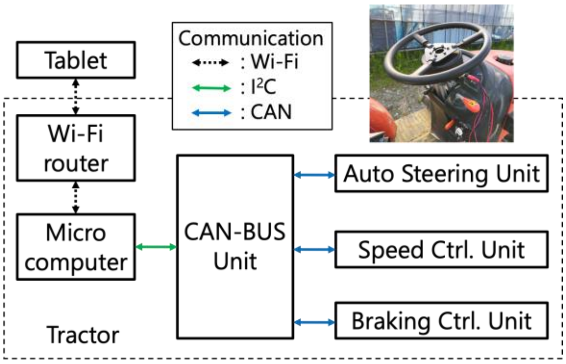


Fig. 2 Hardware configuration of the system

2.3. *System development exercise*

In the developing system, we use Unity, which is used to develop Augmented Reality (AR) applications. In this environment, we write code by C# as the programming language, so we carry out some basic exercises using this language. The exercises on developing an autonomous driving system are conducted in three parts: exercises on the programming language used in the development environment, exercises on handling sensor information on tablet devices, and exercises on implementing autonomous steering and obstacle detection. As the next step, we carry out an exercise to acquire sensor information on the application: the tablet's self-position and posture and depth images obtained from the LiDAR sensor installed on the tablet. The acquired information is displayed on the application screen, as shown in Fig. 3, and an exercise is carried out to use it on the application. As the final step, we carry out exercises on the algorithms for autonomous steering and obstacle detection that are necessary to realize an autonomous driving system. We explain and practice the Pure pursuit method [9], which is widely used in the field of autonomous driving for the autonomous steering algorithm and a nearest point search using a depth image for the obstacle detection algorithm.

2.4. *Final exercise and demonstration*

As the final practice of this exercise, the students are divided into some groups to develop an autonomous driving system for a tractor and then to demonstrate and present their development. The course for the demonstration is shown in Fig. 4. In this course, the following three tasks are set on the course so that the content of the previous exercises can be applied.

1. Autonomous steering near target points
2. Autonomous obstacle avoidance
3. Autonomous stopping around starting position

In task 1, the tractor is required to pass through the position outside of the two specified positions. In task 2,

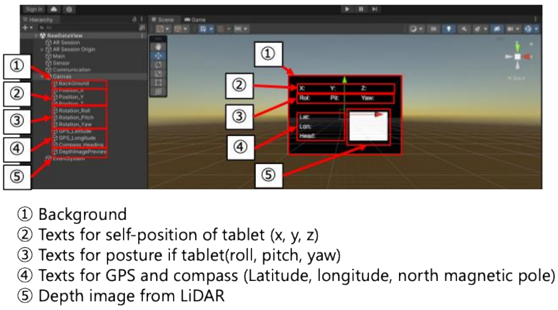


Fig. 3 Example of application GUI

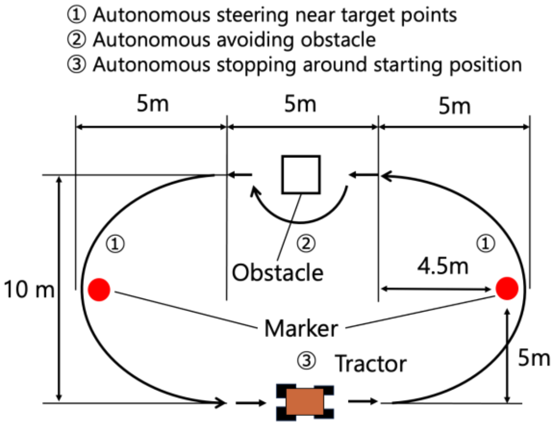


Fig. 4 Overview of course for demonstration

the tractor is required to move while avoiding obstacles that are randomly placed within a certain range. In task 3, after completing tasks 1 and 2, the robot is required to return to the starting point and stop. The evaluation criteria and scores for the demonstration are shown in Table 1. The maximum total score of tasks is 50. The score of task 1 is evaluated for each target point. In this exercise, students tested the system they had developed in their groups on the course set up outdoors, as shown in Fig. 5.

Table 1 The evaluation criteria and scores

Task	Criteria	Score
1	Passing outside the target position	+15
	Passing inside the target position	+10
	Going off the course	0
2	Success of obstacle avoidance	+10
	Success of stopping	+5
	No avoiding or stopping	0
3	Stopping inside of the area	+10
	Stopping outside of the area	+5
	No stopping	0

Fig. 5 Demonstration practice on the course

3. Results

3.1. *Demonstration results*

Table 2 shows the demonstration scores. The number of groups in this exercise was three, and each team had between three and four members. As a summary of the demonstration results, Group 3 completed all of the tasks. The other groups found it difficult to complete task 1-1. In addition, one of the tasks in this exercise was creating the application's GUI. The students used the development methods they had learned in the exercise to devise and implement the functions necessary for the autonomous driving system, such as displaying the tractor's current status and setting parameters for autonomous driving.

3.2. *Proficiency evaluation by rubric*

In this exercise, the proficiency of the students is evaluated using a rubric. The survey using the rubric is conducted before and after the exercise. The students evaluate their own proficiency on each item of the rubric on a four-point scale: Novice (N), Intermediate (I), Proficient (P), and Distinguished (D). The items of the rubric for this exercise are listed below. The result of the rubric is also shown in Table 3.

1. Proficiency of embedded microcontrollers and surrounding electronic circuits
2. Proficiency of sensing technology
3. Proficiency of programming for developing an application
4. Proficiency of technology related to autonomous driving for tractor

Table 2 The score results of the demonstration

Group	Score				
	1-1	1-2	2	3	Total
1	0	0	0	0	0
2	10	0	0	0	10
3	15	10	10	5	40

Table 3 The results of the rubric survey

Contents	Proficiency [%]			
	N	I	P	D
1-before	82	0	9	9
1-after	0	73	9	18
2-before	73	18	0	9
2-after	0	73	18	9
3-before	100	0	0	0
3-after	0	18	36	46
4-before	91	9	0	0
4-after	9	9	27	55

4. Discussion

The result of the demonstration showed that Group 3 had successfully completed all of the tasks. Group 2 was unable to complete all of the tasks in the demonstration due to communication problems, but they completed all of the tasks in the practice for the demonstration. We think that the difficulty level of the tasks on the demonstration was appropriate because the result is not that not all of the teams could complete all the tasks or all the tasks easily. We think the tasks will be necessary to set something closer to tractor-based farming tasks while maintaining this difficulty level.

The results of the rubric showed that the progress of the students' proficiency through the exercises resulted in a significant increase in the items related to programming for tablet applications (Item 3) and technology related to autonomous driving systems (Item 4). In this exercise, these two items are particularly required for developing an application for the demonstration. Therefore, we think that the students' proficiency increased as they progressed with the development. The level of proficiency in embed computer and electronic circuits (item 1) and sensing (item 2) increased less than the two items above. We think this is because we provided packaged versions of the autonomous driving unit and tablet sensor acquisition program. Therefore, there were not many opportunities to apply the knowledge in application development. In order to improve these skills, teaching materials and tasks that can be applied to all contents. However, an autonomous driving system requires many underlying technologies, and time available for exercise is limited. For this reason, we think discussing the packaging of teaching materials and assignment settings is necessary based on the above contents.

5. Conclusion

In this paper, we have designed and carried out exercises on the subject of an autonomous driving system for a tractor and evaluated the results, with the aim of developing human resources who can develop autonomous tractor driving systems that are easy for agricultural workers to install. The demonstration results showed that the students' difficulty level was able to set appropriate difficulty. The results of the rubric also showed that the students improved their proficiency in the algorithms and programming of applications that are fundamental to autonomous driving systems. The future works are considering the demonstration tasks closer to farming work using a tractor and the scope of packaging teaching materials for the underlying technology involved in an autonomous driving system.

Acknowledgements

This exercise have been supported by the Subsidy for Advanced Human Resources Development Projects in the Field of Car Electronics, etc. by City of Kitakyushu and Kitakyushu Foundation for the Advancement of Industry, Science and Technology (FAIS).

References

1. Ministry of Agriculture, Forestry and Fisheries in Japan, “About the situation surrounding smart agriculture,” https://www.maff.go.jp/j/kanbo/smart/attach/pdf/index-240.pdf [Accessed 2024-12-16], in Japanese.
2. Yanmar Holdings Co., Ltd., “Robot / Auto Tractor YT488R ・ YT498 ・ YT4104R ・ YT5114R,” https://www.yanmar.com/jp/agri/products/tractor/yt488r_yt498r_yt4104r_yt5114r_ra/robot.html [Accessed 2024-12-16], in Japanese.
3. Kubota Corp. “Agri Robo,” https://agriculture.kubota.co.jp/product/tractor/MR-1000-AH/ [Accessed 2024-12-16], in Japanese.
4. ISEKI & Co., Ltd., “Robot tractor TJV5 Series, ” https://products.iseki.co.jp/tractor/trac-robot/ [Accessed 2024-12-16], in Japanese.
5. Yanmar Holdings Co., Ltd., “GNSS guidance System・Autonomous Driving System, ” https://www.yanmar.com/jp/agri/products/self_driving/guidance/ [Accessed 2024-12-16], in Japanese.
6. S. Liu, J. Gaudiot, and H. Kasahara, “Engineering education in the age of autonomous machines,” Computer (Long Beach Calif.), vol. 54, 2021, pp. 66–69.
7. A. Agnihotri, M. O’Kelly, R. Mangharam, and H. Abbas, “Teaching autonomous systems at 1/10th-scale: Design of the F1/10 racecar, simulators and curriculum,” in Proceedings of the 51st ACM Technical Symposium on Computer Science Education, 2020.
8. Joint Graduate School Intelligent Car, Robotics & AI, “Joint Graduate School Intelligent Car, Robotics & AI,” https://jgs.kyutech.ac.jp/english/ [Accessed 2024-12-17].
9. R. Craig Coulter, “Implementation of the Pure Pursuit Path Tracking Algorithm, ” Technical Report CMU-RI-TR-92-01, Robotics Institute, Pittsburgh, PA, January 1992.

Authors Introduction

Dr. Daigo Katayama

He received his and Ph. D. degrees from Kyushu Institute of Technology, Japan, in 2023. He is an Assistant Professor at Center for Future Society Implementation for Robotics, Kyushu Institute of Technology, Japan.

Mr. Yuto Nakazuru

He graduated in 2023 from Kyushu Polytechnic Collage, Japan. He is a master student in Kyushu Institute of Technology, Japan.

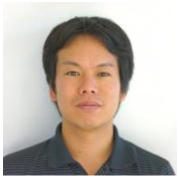
Mr. Hikaru Sato

He received his Bachelor's degree in Engineering in 2023 from National Institute of Technology, Turuoka College, Japan. He is a master student in Kyushu Institute of Technology, Japan.

Mr. Shoun Masuda

He received his Bachelor's degree in Engineering in 2024 from Hiroshima Institute of Technology, Japan. He is a master student in Kyushu Institute of Technology, Japan.

Dr. Yuya Nishida



He received his Ph.D. degrees from Kyushu Institute of Technology, Japan, in 2011. He is an Associate Professor at Department of Human Intelligence Systems, Kyushu Institute of Technology, Japan.

Dr. Shinsuke Yasukawa

He received his Ph.D. degrees from Osaka University, Japan, in 2017. He is an Associate Professor at Department of Human Intelligence Systems, Kyushu Institute of Technology, Japan.

Prof. Kazuo Ishii

He received his Ph.D. degrees from University of Tokyo, Japan, in 1996. He is an Professor at Department of Human Intelligence Systems, Kyushu Institute of Technology, Japan.

Estimation of Image-Based End-Effector Approach Angles for Tomato Harvesting Robots

Kizuna Yoshinaga
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hikaru Sato
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Kazuo Ishii
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Shinsuke Yasukawa
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Email: kizuna.yoshinaga187@mail.kyutech.jp, sato.hikaru292@mail.kyutech.jp, ishii@brain.kyutech.ac.jp, s-yasukawa@brain.kyutech.ac.jp

Abstract

We propose a method to estimate a suitable approach angle for the end-effector of a tomato harvesting robot based on image data. Agricultural harvesting robots often face obstacles such as other fruits or stems around the target crop. Additionally, it is important to approach the target from a direction appropriate for harvesting, considering the shape of the end-effector. The proposed method uses a deep learning-based instance segmentation model to extract regions of fruits and stems, and estimates the suitable approach angle based on their positional relationships. We demonstrated the usefulness of the proposed method using an image dataset acquired in an actual tomato greenhouse.

Keywords: Image processing, Robot vision, Agricultural robot, Harvesting robot

1. Introduction

In recent years, aging workers and lack of successors have become major issues in the agricultural sector due to the declining birthrate and aging population. To solve these problems, research and development of smart agriculture using AI and robots has been actively conducted since around the 1970s. However, there are few practical examples of agricultural robots in the field. Two factors are considered to be responsible for this: technological factors due to differences in the complexity and diversity of crops, tasks, and environments, and economic factors such as high cost [1].

There have been many research and development efforts on robots for the selective harvesting of high-value crops, such as strawberries, tomatoes, and bell peppers, which ripen heterogeneously and require careful handling to harvest only the ripe produce while avoiding damage to surrounding plants and unripe fruits[2], [3], [4].

,In our research group, Fujinaga et al. have developed a tomato harvesting robot with a 3-axis orthogonal arm [5], [6]. This robot is characterized by its simple structure. However, this simple structure has a disadvantage that it cannot harvest tomatoes by avoiding unripe fruits and stems in front of the harvesting target. Therefore, our group has developed a new robot with a horizontally articulated arm to avoid obstacles and to adapt the harvesting approach to the target.

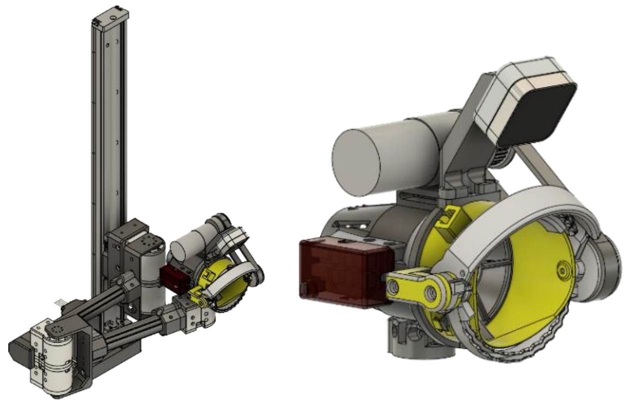
For harvesting that takes advantage of the high degree of freedom of the arm, a method for determining the target posture of the robot's end-effector to the posture of the fruit is required. To determine the target posture of the robot end-effector, it is necessary to extract 3D posture information of the tomato from the image information. The following is an example of a previous study on estimating the 3D posture of a harvesting target. Zhang et al. used a method called tomato pose method (TPM) to estimate the 3D position of a tomato bunch [7]. TPM is based on object detection of a tomato bunch and detection of 11 key points defined as the main stem, stem, and fruit in the bunch. Xiaoqiang et al. proposed a method for detecting tomato posture called the TPD algorithm [8]. The TPD algorithm uses the YOLO-lmk model, which is a modification of YOLOv5s, point cloud segmentation, and a point cloud processing module to achieve fruit and key point detection.

In this paper, we propose and evaluate an image processing method for estimating the direction of tip entry, which is necessary for a robot with a horizontally articulated arm.

2. Proposed Method

2.1. *Harvest strategy*

We describe the configuration and specifications of our group's articulated robot to explain the target environment and requirements of the proposed method. Fig. 1(a) shows the 3D CAD model of the robotic arm used in this study. The horizontally articulated arm is mounted on a linear actuator that moves in the direction

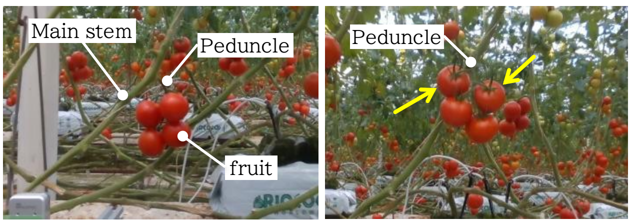


(a) Appearance of the robot arm (b) Appearance of the end-effector

Fig.1 3D CAD model

of gravity. This arm allows the tomatoes to be harvested not only from the front, but also from the side. Fig. 1(b) shows an end-effector with a harvesting mechanism. The suction-cut harvesting mechanism is designed to harvest one fruit per operation. A fan mounted behind the end-effector rotates to draw the fruit in, while blades cut the stems. The proposed image processing method estimates the optimal entry posture for the end-effector.

Fig. 2 shows an example of the appearance of a tomato that is the target of the proposed image processing method. The figure also labels the parts of the tomato as defined in this study. In this study, the thick stem of a tomato cluster is called the "main stem" and the thin stem extending from the main stem to the fruit is called the "peduncle". This study focuses on tomatoes located near the robot and without occlusion, as shown in the figure. In the case of the tomato shown in Fig. 2(a), it can be imagined that any tomato can be harvested because the fruit can be sucked by a frontal approach. However, in the case of the tomato indicated by the arrow in Fig. 2(b), the end-effector interferes with the peduncle when approaching from the front, and the fruit cannot be sucked properly, resulting in a failure of harvest. In this case, the robot can harvest the tomatoes by approaching from the right or left to avoid the peduncles. The proposed method aims to estimate the appropriate approach posture of the end-effector according to the orientation of the fruit by image processing.



(a) Example of a preferred frontal harvest approach (b) Example of a preferred harvesting approach from the left or right

Fig. 2 Example images of camera view the tomatoes that are the subject of the proposed method.

2.2. *Image Processing Algorithm*

After an overview of the proposed method, each process is described in detail. Fig. 3 shows the processing flow of the proposed method.

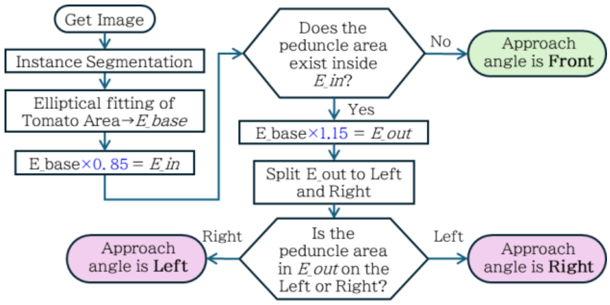
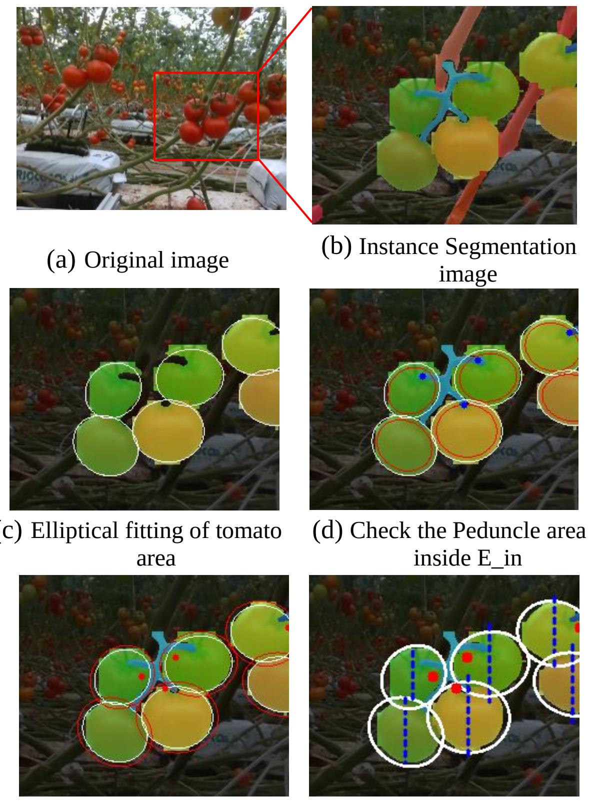


Fig. 3 Process flow of the proposed method.



(a) Original image (b) Instance Segmentation image

(c) Elliptical fitting of tomato area (d) Check the Peduncle area inside E_in

(e) Check the Peduncle area inside E_out (f) Which the Peduncle area in E_out is on the right or left

Fig. 4 Output of each processing step.

Fig. 4 shows the output of each processing step. First, an RGB-D camera captures an image of a tomato, and the RGB component (Fig. 4(c)) is segmented into the main stem, peduncle, and fruit regions (Fig. 4(b)). Hereafter, the analysis focuses on the tomato bunches within the square box in Fig. 4(a). Next, the fruit region is approximated by an ellipse by performing ellipse fitting on the fruit region (Fig. 4(c)). The size of the ellipses is changed as needed (Fig. 4(c)-(e)), and the positional relationship between the fruit and the peduncle is obtained based on the criteria of whether and where the peduncle region exists inside these ellipses(Fig. 4(f)).

Based on the acquired positional relationships, the appropriate end-effector entry posture is estimated.

Instance segmentation (IS) is used to segment the main stem, peduncle, and fruit. IS is an image processing task that uses deep learning to predict pixel-by-pixel class labels for objects in an image, and then identify them individually. We adopted IS for the proposed method for the following three reasons. First, it can flexibly adapt to changes in visibility and illumination, Second, the applied images are suitable for analyzing thin objects, and third, it is necessary to recognize individual fruits for harvesting. In this study, the YOLOv8 segmentation model was used as the training model for IS [9].

After segmentation, consider the ellipse E_in (Fig. 4(d)), which is a 15% reduction of the ellipse generated by ellipse fitting (Fig. 4(c)) for the tomato fruit region. If the peduncle region does not exist inside this *E_in*, the peduncle is judged not to be in the front, and the harvest approach is determined to be front. The reason for reducing the original ellipse (=*E_base*) is as follows. Since the generated ellipse is only an approximation, the ellipse often contains areas near the edges that are not tomato areas. This region may include the surrounding peduncle region, which is not directly related to the target tomato, and using the ellipse *E_base* as it is in the processing may affect the estimation results.

If the peduncle region is included in E_in, this means that the peduncle is located in front of the tomato fruit. Therefore, an ellipse E_out (Fig. 4(d)) is defined by expanding E_base by 15%. By calculating the intersection region between E_out and the peduncle region (Fig. 4(e)) within E_out, we can determine the position of the peduncle around the fruit in relation to the fruit (Fig. 4(f)). Finally, the direction in which the intersection region does not exist is determined to be the end-effector's entry direction.

3. Experiments and Results

To demonstrate the effectiveness of the proposed method, verification is performed using images from the image dataset used during IS training. The evaluation is based on accuracy. The true value was set by the author's own visual inspection using the following concept. Among the tomatoes in the image used, the tomato to be evaluated is determined, and the approach direction of the end effector that the author himself judges to be appropriate is chosen from the three directions: right, front, and left as true value. The accuracy is calculated by comparing this true value with the estimated value by the proposed method. In addition, the tomatoes to be evaluated were classified into three categories according to the growth direction of the stalk, as shown in Table 1, and 15 tomatoes were selected for each category for evaluation and verification.

Table 1. Categorization of tomatoes to be evaluated.

Ⅰ	When the peduncles extend toward the front of the image	Fruit head and peduncle are visible.
Ⅱ	When the peduncles extend sideways of the image	Fruit head and peduncle may or may not be visible in some cases
Ⅲ	When the peduncles extend toward the back of the image	Fruit head and peduncle are not visible.

Table 2 shows the accuracy and average accuracy for categories *Ⅰ* to *Ⅲ*. When viewed by category, the highest values were *Ⅲ*, followed by *Ⅰ* and *Ⅱ*.

Table 2. Accuracy by Category and Overall Average.

Ⅰ	0.800
Ⅱ	0.600
Ⅲ	0.933
Ave.	0.777

The results are analyzed separately for Categories *I*, *II*, and *III*. The estimation accuracy of the proposed method depends heavily on the recognition accuracy of IS. First, In Category *I*, the peduncle is fully visible, enabling accurate recognition, while in Category *III*, the tomato fruit is easily recognized.

Category *III* achieves the highest accuracy because the recognition of tomato fruits, which have simple shapes, is easier than that of peduncles. In Category *II*, the peduncle often extends sideways, and occlusion by the fruit frequently makes parts of the peduncle invisible. This occlusion reduces the recognition accuracy of the peduncle, which subsequently decreases the estimation accuracy. Factors such as the shooting angle, which affects the visibility of peduncles, and the resolution and shooting distance of images, which influence feature extraction, significantly affect recognition accuracy.

4. Conclusion

We proposed an image processing method for estimating the direction of entry of the end-effector based on tomato recognition and posture estimation of tomato fruit for implementation in a tomato harvesting robot with a horizontally articulated arm. We confirmed that the proposed method can estimate the direction of entry of the end-effector with an average accuracy of 0.777. In this method, it was analyzed that the visibility of the fruit peduncle, depending on the orientation of the tomato, significantly affects the performance of the pose estimation. To solve this problem, we plan to examine strategies for shooting at close range and re-photographing from different viewpoints, and to develop an IS model that can handle these strategies. Demonstrating the effectiveness of this method through harvesting trials using actual equipment remains a future challenge.

5. Acknowledgements

This work was supported by JSPS KAKENHI (grant number: 23H03476).

References

1. Zhou, H., Wang, X., Au, W., Kang, H., & Chen, C., “Intelligent robots for fruit harvesting: Recent developments and future challenges”, Precision Agriculture, Vol. 41, No. 5, 1856-1907, 2022.
2. Rajendran, V., Debnath, B., Mghames, S., Mandil, W., Parsa, S., Parsons, S., & Ghalamzan-E, A., "Towards autonomous selective harvesting: A review of robot perception, robot design, motion planning and control," Journal of Field Robotics, Vol. 41, No. 7, pp. 2247-2279, 2024.
3. Arad, B., Balendonck, J., Barth, R., Ben-Shahar, O., Edan, Y., Hellström, T., Hemming J., Kurtser, P., Ringdahl O., Tielen, T., & van Tuijl, B., "Development of a sweet pepper harvesting robot," Journal of Field Robotics, Vol. 37, No. 6, pp. 1027-1039, 2020.
4. Xiong, Y., Ge, Y., Grimstad, L., & From, P. J., "An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation," Journal of Field Robotics, Vol. 37, No. 2, pp. 202-224, 2020.
5. Fujinaga, T., Yasukawa, S., & Ishii, K., "Development and evaluation of a tomato fruit suction cutting Device," 2021 IEEE/SICE International Symposium on System Integration (SII), 628-633, 2021.
6. Fujinaga, T., Yasukawa, S., Li, B., & Ishii, K., " Image Mosaicing Using Multi-Modal Images for Generation of Tomato Growth State Map," Journal of Robotics and Mechatronics, Vol. 30, No. 2, 187-197, 2018.
7. Zhang, F., Gao, J., Zhou, H., Zhang, J., Zou, K., & Yuan, T., "Three-dimensional pose detection method based on keypoints detection network for tomato bunch," Computers and Electronics in Agriculture, Vol. 195, 106824, 2022.
8. Du, X., Meng Z., Ma, Z., Lu, W., Cheng, H., "Tomato 3D pose detection algorithm based on keypoint detection and point cloud processing," Computers and Electronics in Agriculture, Vol. 212, 108056. 2023.
9. Jocher, G., Chaurasia, A., & Qiu, J., "Yolo by ultralytics," Code Repository, 2023.

Authors Introduction

Mr. Kizuna Yoshinaga

He received his Bachelor of Engineering degree in 2023 from National Institute of Technology, Sasebo College, Advanced Engineering Course Department of Complex Engineering. Currently enrolled in the Master's Program of the Graduate School of Kyushu Institute of Technology.

Mr. Hikaru Sato

He received his Bachelor of Engineering degree in 2023 from National Institute of Technology, Tsuruoka College, Advanced Engineering Course Department of Production Systems Engineering. Currently enrolled in the Master's Program of the Graduate School of Kyushu Institute of Technology.

Dr. Kazuo Ishii

He is currently a Professor at the Department of Human Intelligence Systems of Kyushu Institute of Technology, Japan. He obtained his M. S. degree in 1993 and his D. Eng. degree in 1996 at The University of Tokyo. His research interests are in the fields of Underwater Robotics, Field Robotics, Neural Networks and Intelligent Systems.

Dr. Shinsuke Yasukawa

He received his Ph.D. degree in electrical, electronic, and information engineering from Osaka University, Japan, in January 2017. Currently, he is an Associate Professor in the Department of Human Intelligence Systems, Kyushu Institute of Technology, Japan. His research interests include information processing in biological systems and their applications in robotics.

Visual-Based System for Fish Detection and Velocity Estimation in Marine Aquaculture

Raji Alahmad*
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
**Email: raji@brain.kyutech.ac.jp*

Dominic Solpico[1], Shoun Masuda[2], Takahito Ishizuka[3], Kenta Naramura[4], Zhangchi Dong[5], Zongru Li[6], Yuya Nishida[7], Kazuo Ishii[8]
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
[1]Email: dominic@brain.kyutech.ac.jp, [2]Email: masuda.shoun528@mail.kyutech.jp,
[3]Email: ishizuka.takahito370@mail.kyutech.jp, [4]Email: naramura.kenta804@mail.kyutech.jp,
[5]Email: dong.zhangchi671@mail.kyutech.jp, [6]Email: li.zongru477@mail.kyutech.jp,
[7]Email: y-nishida@brain.kyutech.ac.jp, [8]Email: ishii@brain.kyutech.ac.jp

Abstract

As global aquaculture continues to expand to meet the rising seafood demand, optimization of feeding remains a crucial issue for the industry to address to achieve sustainable development. This study proposed a visual-based system for estimating fish velocity, which is to be integrated into a farmer's feeding operation to determine the optimal feed amount. The YOLOv8 algorithm was utilized to detect fish in underwater videos, enabling precise monitoring of fish behavior. The results indicate a successful fish detection with an accuracy of 85%. The fish velocity estimation approach demonstrated the difference between the hungry fish and the normal fish behavior. The findings suggested that integrating fish velocity data into feeding operations can significantly enhance feed efficiency, reduce waste, and promote sustainable aquaculture practices, ensuring optimal fish growth while minimizing environmental and economic impacts.

Keywords: YOLOv8, Aquaculture, Fish Detection, Velocity Estimation.

1. Introduction

With the increasing demand for seafood and declining marine fishery stocks, aquaculture has continued to expand into becoming the major contributor to seafood production worldwide. In 2022, aquaculture production has surpassed the production by capture fisheries, accounting for 51% of aquatic animal production [1]. However, the expansion of the industry also came with sustainability issues, most brought on by poor management practices. If not properly addressed, these issues could get out of control and cause harm not only to the aquatic environment but also to aquaculture production itself [2]. One of the most important issues in the industry is the optimization of feeding. In many aquaculture operations, feeds make up one of the largest, if not the largest portion of the production costs [3]. To maximize the fish growth, farmers estimate the amount of feed supplied and the timing of the feeding, mostly by observing the fish feeding behavior. However, the estimation is based on the farmers' own intuition and their subjective experiences. This results in a notable difference in quality between the fish fed by veteran farmers and those fed by novice farmers. The inferior quality of novice-fed fish could be attributed to the farmers' inefficiency in supplying feed. While underfeeding results in slower growth, overfeeding results in uneaten feed decomposing in water, leading to water pollution affecting the health and growth of fish, resulting in poor quality of fish [4]. In addition, they incur more feeding expenses without producing a substantial increase in the weight and quality of the fish. In the end, the farmers would make less profits or even incur losses from their operation. By applying digital transformation (DX) on fish farms, farmers can get a more accurate estimation of fish feeding behavior and optimize their feeding regime. This in turn can help improve the growth of the fish, potentially resulting in increased margins. With the ubiquity of digital cameras, computer vision (CV) has become one of the most widely used technologies in aquaculture [5]. With the recent emergence of machine learning and artificial intelligence, it has been used for intelligent recognition of fish activity [6]. However, application in most research works is focused on feeding behavior recognition in environments where lighting conditions and water clarity are controlled. Varying conditions in marine cages pose many difficulties in estimating fish feeding behavior. On top of visibility challenges underwater, non-fish objects also hamper the ability of vision systems to recognize fish behavior. This paper presents an underwater imaging system for recognizing fish-feeding behavior in marine fish cage environments. Here we present our approach to enhancing videos of fish activity captured underwater and then tracking their movement. This paper then discusses the relation of the estimated fish motion to the farmer's feeding regime to determine the optimal amount of feed to be supplied.

2. Methodology

Measuring the fish velocity is the key to understanding the fish behavior toward the effective management and optimization of fishery operations. The biggest challenges in underwater video analysis are poor visibility, different lighting conditions, noise, and color distortion due to wavelength-selective water absorption. These challenges result in difficulty in fish detection and tracking. The research aims to track the fish from the video from the fish cage and estimate their velocity. This is critical for maintaining sustainable and productive fishery practices. The study is divided into three mains objectives shown in Figure 1.

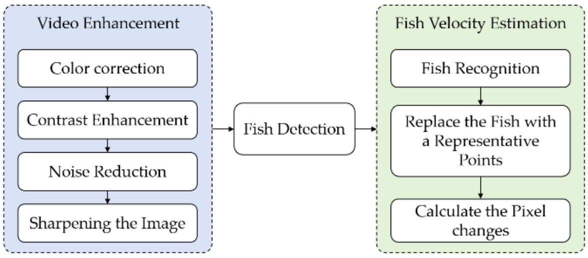


Figure 1. The flowchart of visual-based fish motion tracking

2.1. *Video Enhancement*

Enhancing underwater video is challenging due to the unique condition of the water. The wavelength-dependent absorption result is a color distortion [7]. The blue color travels longest in the water while the red color faces higher attenuation over short distances. Furthermore, image degrading is strongly affected by light attenuation and scattering [8]. Addressing these issues is an essential process for improving video quality, it is also promising to facilitate more accurate fish detection and tracking. Several steps were applied to enhance the underwater video: color correction, contrast enhancement, noise reduction, and sharpening. Figure 2 shows the snapshot of the video enhancement.

Figure 2. Image enhancement result

2.2. *Fish Detection*

Fish detection is critical for monitoring and analyzing fish behavior. You Only Look Once (YOLO) algorithm is a well-known object detection algorithm, especially in complex environments. Its versions are widely used for different applications such as face recognition, small object detection, and underwater object detection [9]. In this research, YOLOv8 was applied to detect the fish in the video by breaking the video into individual frames, then, applying the detection algorithm, and finally recombining the frames to make the video with the detection information.

The datasets were collected from the video in an underwater real environment for Yellow tail fish aqua-cage. The camera was set at depths of 3.4m to continuously capture a video before, during, and after feeding processing. Afterward, the video frames were extracted to make training and validation dataset images. The dataset consists of 150 augmented images, divided into 71% for training, 20% for validation, and 10% for testing. The training iterations were set to 150 epochs, at each epoch, a careful analysis of the model performance such as precision and recall was applied. Using hyperparameters such as learning rate and batch size, the model has been adjusted to optimize detection accuracy. The model was fine-tuned to balance sensitivity with specificity to minimize false positives which is particularly challenging in a complex environment like underwater.

2.3. *Fish Velocity Estimation*

Several methods, such as optical flow, Kalman filter, and DeepSort, can be applied to estimate the fish velocity in the aqua-cage. However, in complex environments such as underwater, where lighting significantly affects pixel brightness, traditional tracking techniques become more complicated due to difficulties in extracting features from the object to be tracked. In this research, we proposed a novel method to simplify the issue of motion tracking and velocity calculation.

Tracking the fish motion in the sequencing frames involves tracking the whole fish by extracting fish features and calculating the pixel displacements among every two frames. Nevertheless, the changes in the light condition of the sequencing images, and unexpected fish directions and poses make it more difficult to apply the tracking methods to the fish directly. Therefore, replacing the fish with representative points is an essential step

3. Results and discussion

With the enhanced video, the YOLOv8 was trained using 500 images, the trained model then was applied on separate images to evaluate its performance. To evaluate the level of video enhancement, the trained model was applied on two test sets; 50 images before enhancement and the same 50 images after enhancement. Table 1 illustrates the dataset partitioning for the training, validation and testing, the comparison of the mean average perdition (mAP-50), and the system accuracy between the video before and after enhancement.

Table 1. The dataset partitioning for training YOLOv8 and model mAP-50

	Before enhancement	After enhancement
Dataset	150	
Training set	106	
Validation set	30	
Test set	14	
mAP-50 best	0.874	0.864
Test accuracy	70.3%	73.2%

The table showed a slight decrease in the mAP, but increasing in the test accuracy, indicating the need for further analysis of the images. The testing set was carefully checked for better understanding. Table 2 summarizes the comparison results of the deep analysis of the testing dataset before and after enhancement.

Table 2. Comparison results of the deep analysis of the testing dataset before and after enhancement.

parameter	value
# of testing images	14
# of images with the same detection rate	3
The model detected more fish before enhancement than after	5 images
The model detected more fish after enhancement than before	6 images
Total number of fish in the testing images	104
# of detected fishes in the images before enhancement, (detection rate)	77, (74%)
# of detected fishes in the images before enhancement, (detection rate)	80, (76.9%)

The table shows a slight improvement in detection. The detection rate of the images before enhancement was 74%, while after enhancement, it was 76.9%.

To track the fish's motion, the trained model of YOLOv8 was first applied to detect the fish and then replaced the detected fish with representative dots. Figure 3 illustrates a snapshot from a video demonstrating velocity estimation of fish movement.

Figure 3. Estimated fish velocity

The calculation is conducted based on the distance between corresponding pixels in consecutive frames, frame (i) and frame (i+1) successively, for each detected fish. This method allows tracking of the movement of individual fish. For each frame, the approach calculates and provides the average velocity of all the fish present within the frame, offering an overall measure of their movement.

The proposed method is employed in two videos, the case of normal fish movement, and the second the case during fish feeding. Choosing two different videos has two purposes:

1- To determine the difference in the detection of YOLOv8 since the fish's random and high-velocity movement might affect the detection due to the quality of the video.

Figure 4 illustrates the number of detected fish in the two described cases.

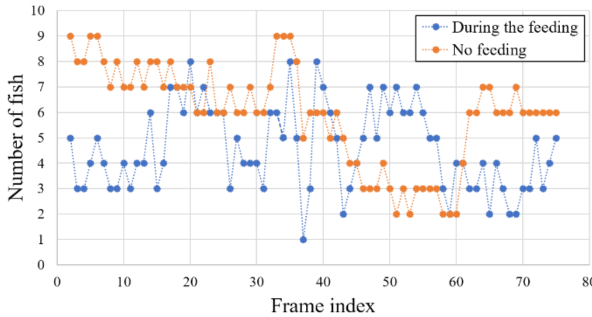


Figure 4. Number of detected fish in two cases (during the feeding presses and without feeding)

From the figure, the number of detected fish during the feeding is higher than in the case of no feeding. The reason is the direction of the fish. During the feeding, the fish is hungry, thus they move in random directions, also the direction camera view axis, in this case, the detection algorithm failed to detect these fish. In addition, the high speed of the fish decreased the video quality resulting in frailer detection.

2- To evaluate the velocity estimation approach with different fish velocities. Figure 5 illustrates the fish velocity estimation for the two videos: During the feeding and no feeding.

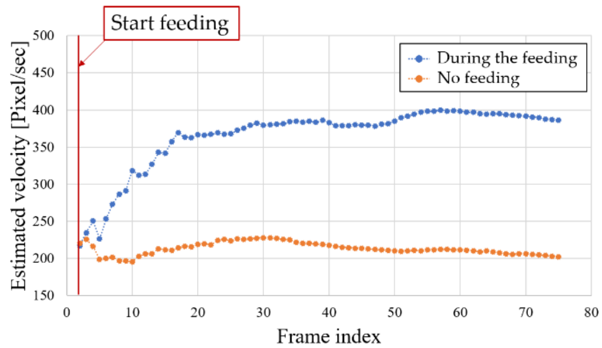


Figure 5. Estimated fish velocity

The orange line in Figure 5 represents the case when the fish are behaving normally without feeding. The blue line represents the case when the feeding process started. The estimated fish velocity when feeding is going is higher than the normal case. This estimated velocity was calculated by pixel per second

4. Conclusion

This study proposed a visual-based system for optimizing feeding strategies in aquaculture by underwater imaging, fish detection, and velocity estimation. The video enhancement process addressed several challenges such as color distortion, poor visibility, and noise, resulting in improved fish detection accuracy from 70.3% to 73.2%. YOLOv8 was utilized for fish detection, trained with 106 underwater images, and tested on 14 images. Despite a slight decrease in mAP-50, the in-depth analysis of the test images demonstrates that detection was better after enhancement by 2.9%. Track fish velocity was conducted by pixel displacement which is a simplified optical flow technique. The system proved that fish velocity significantly increased during feeding, offering a potential approach for optimizing feeding schedules.

This research highlights the potential for integrating image enhancement, machine learning, and velocity tracking to promote sustainable aquaculture. It offers a pathway to reduce feed waste, lower production costs, and improve profitability while supporting sustainable and efficient fish farming practices.

Future work will expand the dataset and refine detection models, for accuracy improvement. Developing an algorithm that analyzes both videos before and after enhancements will enable dynamic adjustments of detection performance.

Acknowledgements

This research was supported by the research program on development of innovative technology grants from the Project of the Bio-oriented Technology Research Advancement Institution (BRAIN) (No. JPJ007097). We would like to acknowledge Dr. Hiroyasu Mizoguchi, Dr. Akihiko Matsuda, and Dr. Yuichiro Fujinami of the Japan Fisheries Research and Education Agency, and Mr. Tokuo Suetsugu and Yoshinori Yatsunami of Belltechne Co., Ltd., for their collaboration in this research.

References

1. Food and Agriculture Org., "The State of World Fisheries and Aquaculture 2024. Blue Transformation in action," Food and Agriculture Org., Rome, Italy, 2024.
2. M. Martinez-Porchas and L. R. Martinez-Cordova, "World Aquaculture: Environmental Impacts and Troubleshooting Alternatives," The Scientific World Journal, vol. 2012, 2012.
3. W. R. Rola and M. R. Hasan, "Economics of aquaculture feeding practices: a synthesis of case studies undertaken in six Asian countries," in Economics of aquaculture feeding practices in selected Asian countries, M. R. Hasan, Ed., Rome, Food and Agri. Org., 2007, pp. 1-31.
4. M. Sun, S. G. Hassan and D. Li, "Models for estimating feed intake in aquaculture: A review," Computers and Electronics in Agriculture, vol. 127, pp. 425-438, 2016.
5. M. Saberioon, A. Gholizadeh, P. Cisar, A. Pautsina and J. Urban, "Application of machine vision systems in aquaculture with emphasis on fish: state-of-the-art and key issues," Reviews in Aquaculture, vol. 9, no. 4, pp. 369-387, 2017.
6. C. Zhou, K. Lin, D. Xu, L. Chen, Q. Guo, C. Sun and X. Yang, "Near infrared computer vision and neuro-fuzzy model-based feeding decision system for fish in aquaculture," Computers and Electronics in Agriculture, vol. 146, pp. 114-124, 2018.
7. C. Li et al., "An Underwater Image Enhancement Benchmark Dataset and Beyond," in IEEE Transactions on Image Processing, vol. 29, pp. 4376-4389, 2020, doi: 10.1109/TIP.2019.2955241.
8. Sahu, P., Gupta, N., & Sharma, N. (2014). "A survey on underwater image enhancement techniques." International Journal of Computer Applications, 87(13), 19-23.
9. Li, Zongru, Hussam Alraie, Dominic Solpico, Yuya Nishida, Kazuo Ishii, Yoshinori Yatsunami, Masanari Fuchigami, and Tokuo Suetsugu. "Recognition of Fish in Aqua Cage by Machine Learning with Image Enhancement." In 2024 IEEE/SICE International Symposium on System Integration (SII), pp. 637-643. IEEE, 2024.

Authors Introduction

Dr. Raji Alahmad

He is currently a researcher at the Department of Human Intelligence Systems of Kyushu Institute of Technology, Japan. He obtained his MEng degree in 2018 and his DEng in 2021 at Kyushu Institute of Technology. His research interests are AI algorithms, Field Robotics, Underwater communication, Image processing, and Neural networks.

Dr. Dominic Solpico

He obtained his M.S. in Electronics Engineering at the Ateneo de Manila University, Philippines in 2015. He obtained his Ph.D. at Kyushu Institute of Technology, Japan in 2022, and currently works as a researcher at the Department of Human Intelligence Systems of the same university. His research interests are in the fields of Intelligent Aquaculture, Wireless Sensor Networks (WSNs), and Field Robotics.

Mr. Shoun Masuda

He received his Bachelor's degree in 2024 from Hiroshima Institute of Technology in Japan. He is currently a master's student at Kyushu Institute of Technology

Mr. Takahito Ishizuka

He received his Bachelor's Degree in Engineering from Fukui University of Technology in 2024. He is currently a master student in Kyushu Institute of Technology, Japan

Mr. Kenta Naramura

He received his Bachelor's degree in Engineering in 2024 from National Institute of Technology, Tsuyama College in Japan. He is currently a master student in Kyushu Institute of Technology, Japan.

Mr. Zhangchi Dong

He received his bachelor’s degree from Nanjing University of Aeronautics and Astronautics, China, in 2021. In 2024, he started his master’s studies at Kyushu Institute of Technology, majoring in Human Intelligence Systems. His research interests include computational fluid dynamics simulation and robotics.

Mr. Zongru Li

He earned a bachelor's degree from Zhengzhou University, china in 2021 and is currently pursuing a master's degree at Kyushu Institute of Technology in Japan.

Dr. Yuya Nishida

He is an Associate Professor at the Graduate School of Life Science and System Engineering, Kyushu Institute of Technology, Japan. His research area is in the field of robotics, its application, and data processing.

Prof. Kazuo Ishii

He is currently a Professor at the Department of Human Intelligence Systems of Kyushu Institute of Technology, Japan. He obtained his M. S. degree in 1993 and his D. Eng. degree in 1996 at The University of Tokyo. His research interests are in the fields of Underwater Robotics, Field Robotics, Neural Networks, and Intelligent Systems.

Evaluating of Tree Branch Recognition Algorithm in Pruning Robots under Augmented Environmental Conditions.

Mohammad Albaroudi
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
albaroudi.mohammad344@mail.kyutech.jp

Raji Alahmad
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
raji@brain.kyutech.ac.jp

Abdullah Alraee
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
alraee.abdullah-abdul648@mail.kyutech.jp

Kazuo Ishii
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
ishii@brain.kyutech.ac.jp

Abstract

Integrating service robots has revolutionized several sectors, by enhancing accuracy, efficiency, and scalability. Those robots are crucial in automating labor-intensive processes such as tree pruning, where accurate branch detection is vital. This research evaluates the performance of YOLOv8-seg model for recognizing tree branches as a step towards fully autonomous pruning. To address the challenges posed by diverse and complex real-world conditions, video sequences are augmented using techniques that simulate environmental variations, such as changes in brightness, contrast, and Gaussian noise. The evaluation metrics including the number of true detections, number of false detections, and precision, demonstrate robust and accurate branch perception under real-world conditions. These results highlight the potential of YOLOv8-seg to improve pruning systems, paving the way for scalable, efficient, and accurate robotic solutions in tree maintenance.

Keywords: Tree Pruning, Automation, Branch, YOLOv8, Recognition.

1. Introduction:

Integrating service robots across various applications has significantly improved operational capabilities, accuracy, efficiency, and scalability [1]. Among these diverse applications, one area that is expected to benefit greatly is the automation of labor-intensive arboricultural operations, such as tree pruning [2].

Tree pruning is a critical maintenance operation that requires precision and accurate branch recognition for pruning to promote healthy growth, and improve overall appearance [3]. Automating this process presents opportunities to boost productivity while reducing the demand for human labor, especially in large-scale operations. Despite its benefits, automated tree pruning remains a challenge due to the need for accuracy, especially for diverse real environments. In this context, advanced machine vision techniques have become essential in enabling robots to accurately identify branches and then perform pruning operations. This research investigates the use of YOLOv8-seg, a state-of-the-art object detection algorithm [4][5], to recognize tree branches as part of a fully autonomous pruning system. Detecting tree branches is difficult due to the complexities of natural environments, which may vary greatly in brightness, contrast, and Gaussian noise conditions, making it necessary to evaluate the performance of the detection model under these dynamic scenarios in the real environment. The findings of this study aim to demonstrate the feasibility of using YOLOv8-seg in practical arboricultural environments, which could lead to significant advances in autonomous tree maintenance.

2. Architecture of YOLOv8 – Segmentation Model:

Segmentation models provide a precise definition of branch boundaries [6], which helps determine their exact shape and contributes to separating them from overlapping branches and performing post-segmentation analysis such as measuring branch thickness and angle, which may facilitate the pruning process. As shown in Figure 1, the segmentation model uses the training image to process it through a convolutional backbone to extract features. These features are fed into two parallel heads: the segmentation head, which predicts segmentation masks among the pixel level, and the detection head, which outputs bounding boxes and confidence scores. This dual output allows for accurate object detection and segmentation, making it particularly effective in applications of overlapping environments such as the recognition of tree branches.

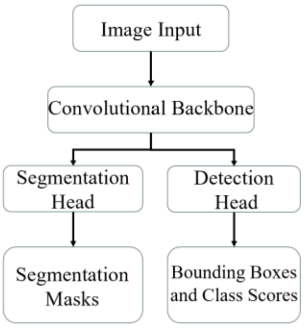


Fig. 1: Simple architecture of the YOLO segmentation.

3. Methodology:

3.1. *Data Collection and Training:*

The YOLOv8-seg model was trained on a dataset of 1973 images taken at different periods to account for seasonal variations in foliage density and scene conditions [7]. Different scenarios were considered, including branches surrounded by dense foliage, varying branch spacing, and diverse environmental backgrounds. This diversity ensures that the model is robust and capable of detecting branches under real-world conditions. Table 1 shows the dataset statistics and training parameters.

Table 1: Dataset statistics and training parameters.

Parameter	Value
Training Images	1,973
# of branches	37,146
Preprocessing	Stretched into 640 × 640 Pixels
Training Model	YOLOv8m-seg
# of Epochs	500
Batch Size	16

Figure 2 shows the decline in the main loss components over 500 epochs. The box loss refers to the error in determining the predicted coordinates of the specified box compared to the training data, while the classification loss measures the error in classifying objects into their correct classes. Finally, the segmentation loss measures the error in predicting the segmentation masks, which estimate the pixels that include the branches. The continuous decline in the loss of components indicates the ability of the model to learn and improve its performance on detection and segmentation tasks.

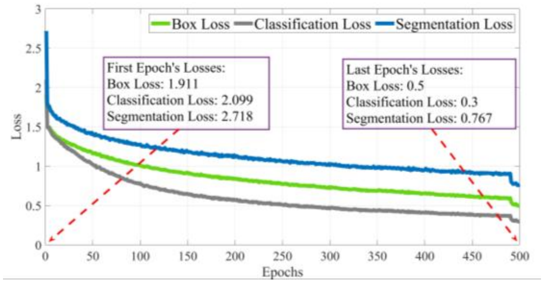


Fig. 2: Training loss reduction across 500 epochs.

The figure shows that the improvement in loss exceeds 1.4 for the box loss, 1.7 for the classification loss, and 1.9 for the segmentation loss. This demonstrates the robustness of the training process and the closeness of the dataset used.

3.2. *Experiment Setup:*

The experiment aims to investigate the detection behavior, for video sequences that initially captured the movement of the pruning robot as it steadily advances towards the branch.
As shown in Figure 3, a mechanical arrangement is utilized to simulate the effect of distance on branch detection, which is interesting to the pruning process. This experiment uses an iPhone mounted on a stick controlled by a DC motor to capture branches of the tree species while changing the distance. The stick advances gradually up to 1 m toward the tree to simulate the effect of distance on branch detection performance.

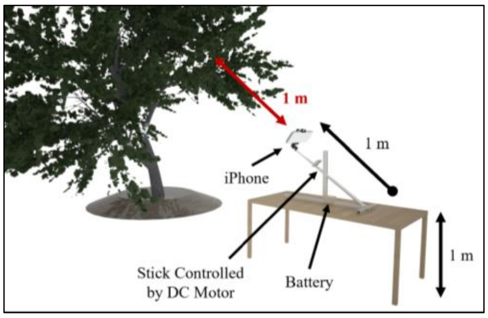


Fig. 3: Experiment setup.

The evaluation data consists of 1352 frames, captured for 3 trees. Figure 4 illustrates a sample of branch recognition for the three captured videos.

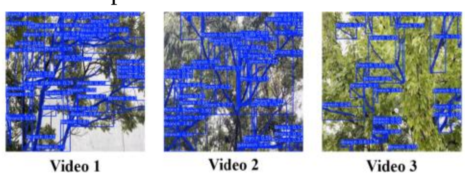


Fig. 4: A sample of branch recognition for evaluation data.

4. Results and Discussion:

4.1. *False Positives and True Positives vs Confidence Threshold:*

As an evaluation of the model, the focus was on the total of true positives (true detections), false positives (false detections), and precision as performance metrics. A true positive (TP) result occurs when the model correctly predicts the positive class. A false positive (FP) result occurs when the model incorrectly predicts the positive class. Precision measures the accuracy of positive detections and is calculated as the ratio of true positives to all predicted positives as the following equation:

$$Precision = TP/TP + FP \qquad (1)$$

Figure 5 shows the variation in the number of TP and FP with different confidence thresholds, providing valuable insight into the model's behavior.

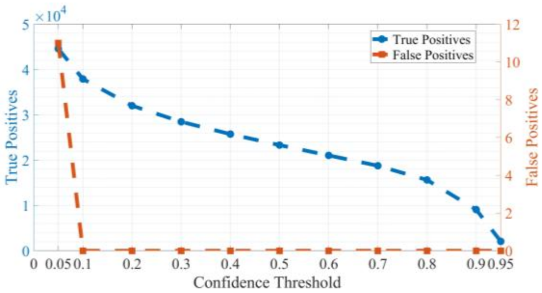


Fig. 5: Impact of confidence threshold on true positives and false positives.

At very low confidence thresholds (0.05), the model detects many cases as TP, at the existence of a few false positives. The precision in this case is 0.999, this value indicates that the model is not lenient in its predictions, even at very low thresholds. As the confidence threshold increases, false positives disappear, indicating a perfect performance in precision; however, this disappearance is accompanied by a decrease in true positives, indicating that the model is becoming more conservative and may miss valid detections. To ensure that there are no false positives, with a satisfied number of true positives, the model can operate with confidence thresholds ranging from 0.1 to 0.3.
Figure 6 shows the true positive detections of tree branches across the video of evaluation data for confidence thresholds 0.1, 0.2, and 0.3.

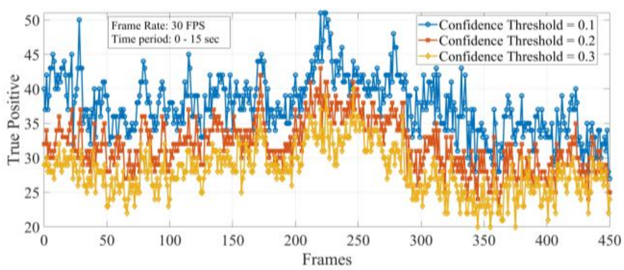


Fig. 6: Analyzing true positives across frames.

The figure shows that the number of TP tends to decrease with time. This decrease may explicate by some branches falling out of the frame range. On the other hand, the fluctuation in the number of TP over close times suggests that scene complexity may affect detection performance.

In the next section, the effect of some real-world environmental conditions like brightness, contrast, and Gaussian noise will be simulated by augmenting the evaluation videos with Python's computer vision libraries, according to the general equation:

$$Output\ (x, y) = \alpha \times Input(x, y) + \beta + N(x, y) \qquad (2)$$

Where; Output (x, y) is the pixel intensity in the output image at position (x, y), Input (x, y) is the pixel intensity in the input image at position (x, y), α is the contrast adjustment factor, β is the brightness adjustment factor, and N(x, y) is the Gaussian noise at position (x, y).
The adjustment parameters α, β, and N values are applied to the RGB domain images.

4.2. *True Positives and Average of Confidence Scores for Simulated Brightness Conditions:*

Evaluating brightness conditions enhances the reliability of the pruning system in dealing with different brightness conditions across the day, season, and different scene conditions. Figure 7 shows the effect of different brightness conditions on TP.

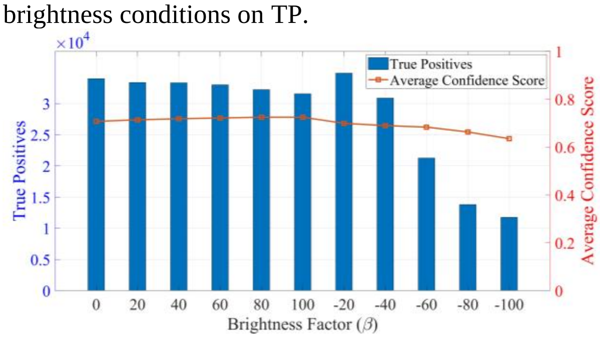


Fig. 7: TP vs. brightness conditions.

The number of TP remains relatively stable across a wide range of brightness levels. This suggests that the model is robust to different brightness conditions. For extremely low lighting levels, significant decreases in TP occur, indicating that very dark conditions negatively affect detection performance. The model generally maintains stable confidence performance across a wide range of brightness conditions.

4.3. *True Positives and Average of Confidence Scores for Simulated Contrast Conditions:*

Contrast refers to the difference in brightness between branches and their background, such as foliage, sky, and other overlapping branches. In our case, shadows and scene conditions can create areas of low contrast, making it difficult to detect branches. Figure 8 shows the detection performance for different contrast levels.

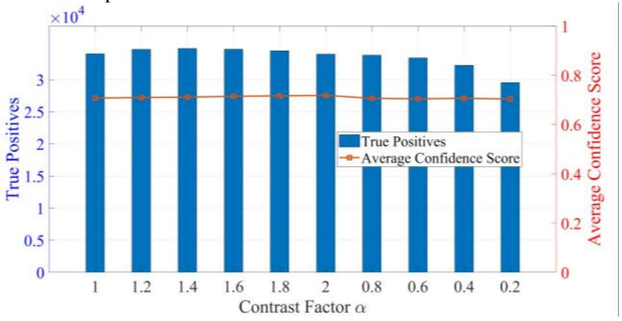


Fig. 8: TP vs. contrast conditions.

From Figure 8, the model demonstrates a high degree of robustness to different contrast levels within a moderate range. The number of TP remains relatively stable across these contrast levels. This indicates that the model is well-trained to handle variations in contrast that might occur due to changes in lighting, shadows, or weather conditions. In addition, the average confidence score remains relatively stable across all contrast levels.

4.4. *True Positives and Average of Confidence Scores for Simulated Gaussian Noise:*

Gaussian noise reflects real-world noise that may be generated by cameras or environmental factors. according to this equation:

$$noise \sim \mathcal{N}(\mu = 0, \sigma^2) \qquad (3)$$

Where: $\mathcal{N}$ is the Normal (gaussian) distribution, μ is the mean, σ is the standard deviation, and σ^2 is the variance.

The noise distribution is symmetric around zero ($\mu = 0$), therefore, there is no systematic increase or decrease in brightness or intensity. Figure 9 illustrates the effect of Gaussian noise on detection performance.

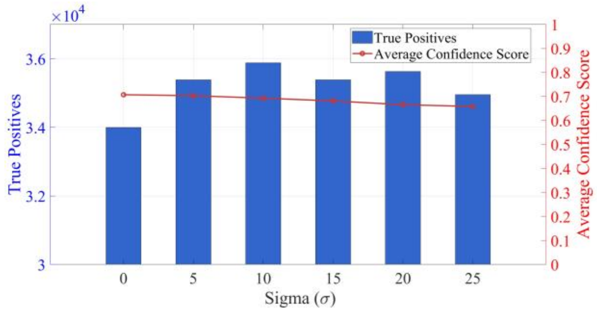


Fig. 9: TP vs. Gaussian noise conditions.

The model maintains a close number of detections under Gaussian noise $\mathcal{N}(0, \sigma^2)$, indicating high robustness, the average confidence scores decrease slightly as noise levels increase. suggesting reduced prediction certainty in noisy conditions.

5. Conclusion:

This research demonstrates the robustness of the Yolov8-seg model in detecting tree branches under very low confidence thresholds with very few false positives and high precision. To investigate the robustness of the model in the real world, different brightness, contrast, and normal Gaussian noise conditions were simulated using computer vision's Python libraries. The results showed that the model maintains a stable number of detections under most conditions. A slight decrease in confidence scores was observed for extreme environmental conditions, indicating a decrease in certainty while maintaining detection ability. This stability indicates that the model is effective and reliable in terms of precision for low thresholds and different environmental conditions.

As a future of this research, we will investigate expanding the evaluation scope to include additional environmental factors, such as blur noise, the effect of the distance between the robot and branches, and the camera angles. The YOLO model could also be integrated into an autonomous real-time pruning system and evaluated in field environments.

References

1. Gonzalez-Aguirre, J. A., Osorio-Oliveros, R., Rodriguez-Hernandez, K. L., Lizárraga-Iturralde, J., Morales Menendez, R., Ramirez-Mendoza, R. A., ... & Lozoya-Santos, J. D. J. (2021). Service robots: Trends and technology. *Applied Sciences*, *11*(22), 10702.
2. Zeng, H., Yang, J., Yang, N., Huang, J., Long, H., & Chen, Y. (2022, October). A review of the research progress of pruning robots. In *2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA)* (pp. 1069-1073). IEEE.
3. Morris, H. (2010). Tree pruning: A brief history.
4. Terven, J., Córdova-Esparza, D. M., & Romero-González, J. A. (2023). A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine Learning and Knowledge Extraction*, *5*(4), 1680-1716.
5. Hussain, M. (2024). Yolov1 to v8: Unveiling each variant–a comprehensive review of yolo. *IEEE Access*, *12*, 42816-42833.
6. Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2021). Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, *44*(7), 3523-3542.
7. *Project1. (2024, November). Branch Segmentation Dataset [Open source dataset]. Roboflow Universe. Roboflow. https://universe.roboflow.com/project1-jo7ku/branch-segmentation-is9he*

Authors Introduction

Mr. Mohammad Albaroudi

He received his M.Sc. degree in Radio and Mobile Telecommunication Systems from Higher Institute of Applied Sciences and Technology, Damascus, Syria in 2021. He is currently pursuing a Ph.D. in the Department of Life Science and System Engineering at Kyushu Institute of Technology, Fukuoka, Japan. His research interests are machine learning algorithms, image processing, and mobile telecommunication.

Dr. Raji Alahmad

He is currently a researcher at the Department of Human Intelligence Systems of Kyushu Institute of Technology, Japan. He obtained his MEng degree in 2018 and his DEng in 2021 at Kyushu Institute of Technology. His research interests are AI algorithms, Logistics, Underwater Robotics, Image processing, and Neural networks.

Mr. Abdullah Alraee

He received his bachelor’s degree in communication engineering from Al-Wataniya Private University, Syria, in 2018, and his master’s degree from Kyushu Institute of Technology, Japan, in 2024. His master’s research focused on Tennis Ball Detection and Position Estimation for Mobile Robot Applications using Artificial Intelligence. Currently, he is pursuing a Ph.D. at Kyushu Institute of Technology, starting in 2024, with research interests in robotics, artificial intelligence, Image processing, and automation.

Prof. Kazuo Ishii

He is currently a Professor at the Department of Human Intelligence Systems of Kyushu Institute of Technology, Japan. He obtained his M. S. degree in 1993 and his D. Eng. degree in 1996 at The University of Tokyo. His research interests are in the fields of Underwater Robotics, Field Robotics, Neural Networks, and Intelligent Systems.

Trajectory Analysis for a Mobile Robot Adapted Three Omni Rollers in Constant Roller's Speed

Kenji Kimura
Department of Control Engineering, National Institute of Technology, Matsue College
14-4 Nishi-ikuma-cho, Matsue-shi, Shimane,690-8518, Japan

Kazuki Nakayama
Department of Control Engineering, National Institute of Technology, Matsue College
14-4 Nishi-ikuma-cho, Matsue-shi, Shimane,690-8518, Japan

Katsuaki Suzuki
Kumamoto Industrial Research Institute
3-11-38 Higashi machi, Higashi-ku, Kumamoto 862-0901, Kumamoto, Japan
E-mail: k-suzuki@kumamoto-iri.jp

Kazuo Ishii
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Email: k-kimura@matsue-ct.acjp, ishii@brain.kyutech.ac.jp

Abstract

The mobile robot is being developed for use in the logistics industry. A mechanism with multiple omni rollers as an omni-directional moving mechanism has been developed and its kinematics analyzed. In this study, the kinematics were verified in a simulation environment as a preliminary step in order to reduce the cost and duration in the verification experiments with good prospects. The trajectory of the robot was derived and analyzed when a constant roller speed was given in the robot kinematics.

Keywords: Trajectory analysis, Mobile robot, Angular velocity

1. Introduction

Recently, efficient mobility has become a requirement for mobile robots in areas like logistics. Thus, omnidirectional movement with either non-holonomic or holonomic characteristics can produce a total of 3 degrees of freedom motion (sum of 2 degrees of freedom translational motion and 1 degree of freedom rotational motion). Robotic vehicle development is getting attention.

Among them, holonomic mobile mechanisms are easy to control and have excellent omni-directional mobility, as the wheels are driven independently. For this reason, mobile robots arranged in equilateral triangles have been developed [1].

In the RoboCup MSL, a mechanism with three omni-rollers is used: the RV-infinity [2], the Musashi150 [3] and the NuBot [4] have three omni-rollers arranged in an equilateral triangle. The kinematics of a sphere with two roller velocities as inputs has been proposed for the RoboCup MSL as a kinematics study of the moving mechanism [5] and validated on a real machine [6][7]. There are studies on the region of existence of the velocity of the mobile robot with respect to the roller velocity [8][9][10] and on the rotational efficiency of the sphere [11][12]. As for the analysis of robot trajectories, [13] has successfully analyzed the trajectory of a sphere robot driven by rollers using a simulator.

To verify whether the derived equations adequately represent the intended conceptual model, as in this study, it is necessary to measure the actual values on the actual machine and check the errors with the theoretical values. However, verification using actual equipment is costly and time-consuming to develop, and feedback is also time-consuming, making it difficult to shorten the speed of development. There is an example of verification of the trajectory of a moving mechanism with three spherical rollers using the simulation function of 3DCAD. This allows simulations to be carried out on a PC at the pre-experimental stage to reduce development costs and shorten the time required.

In this study, a simulation on the trajectory of a mechanism with three constant values as input for a three-wheeled wheeled mobile mechanism is carried out using the mechanism analysis function in the 3DCAD software. The results are then compared with theoretical values using a mathematical model to verify the consistency of the mathematical model. The rest of this study is as follows: Chapter 2 discusses the kinematics and trajectory for mobile robots. Chapter 3 shows simulation resalt. Finally, we present the summary and future tasks.

2. Kinematics and Trajectory for Omnidirectional mobile robot

This section introduces kinematics which the roller placement position can be changed arbitrarily and analysis robot trajectory in constant rollers speed.

2.1. *Kinematics*

Figure 1 shows a top view of the mobile robot that has a common radius of all omni-wheels adapted the *i-th* rollers ($i = 1,2,3$) contact point $\boldsymbol{P_i}$ (contact position angle α_i) on a circle, which has a radius L (the distance from the robot center to the contact points between wheels and floor)(See Table 1).

$\{x_w, y_w\}$ is the global coordinate system (origin $\boldsymbol{O}$) and $\{x_m, y_m\}$ is robot coordinate system (origin $\boldsymbol{\acute{O}}$). And. The robot orientation ϕ is referred as the angle between x_w-axis and x_m-axis.

Thus, $[v_1, v_2, v_3]^{\boldsymbol{T}}$ in robot coordinate is represented as follows.

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} -\sin\alpha_1 & \cos\alpha_1 & 1 \\ -\sin\alpha_2 & \cos\alpha_2 & 1 \\ -\sin\alpha_3 & \cos\alpha_3 & 1 \end{bmatrix} \begin{bmatrix} V_{x_m} \\ V_{y_m} \\ L\dot{\phi} \end{bmatrix} \tag{1}$$

And. $\boldsymbol{V_w}$ denote robot direction vector in robot coordinate system.

Global coordinate system $\{x_w, y_w\}$ is a rotation of the robot coordinate system $\{x_m, y_m\}$ by ϕ. Therefore, following expression can be valid.

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} -\sin(\alpha_1 - \phi) & \cos(\alpha_1 - \phi) & 1 \\ -\sin(\alpha_2 - \phi) & \cos(\alpha_2 - \phi) & 1 \\ -\sin(\alpha_3 - \phi) & \cos(\alpha_3 - \phi) & 1 \end{bmatrix} \begin{bmatrix} V_{x_w} \\ V_{y_w} \\ L\dot{\phi} \end{bmatrix} \tag{2}$$

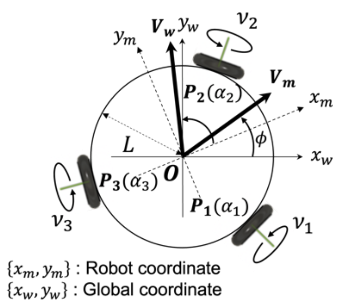


Figure 1 Robot speed vector by robot coordinate system and global coordinate system

2.2 *Kinematics for isosceles triangle three rollers arrangement*

In this section, the trajectory of the center of gravity of the robot is analyzed for a constant roller speed $[v_1, v_2, v_3]^T$ at a constant time t. And as an equilateral triangle roller arrangement we assumed $(\alpha_1, \alpha_2, \alpha_3) = (\pi/6, 5\pi/6, 3\pi/2)$. Here, $v_1 + v_2 + v_3$ is considered by dividing the cases as follows.

(A) Case of $v_1 + v_2 + v_3 \neq 0$

Solving Eq. (1) for, $L\dot{\phi}$ is expressed using the linear expression for v_1, v_2 and v_3 . Therefore, $\dot{\phi}$ is constant. As shown in Figure 2, if the robot moves with an initial speed of $\boldsymbol{V_0}$ and a constant attitude rotation speed of $\dot{\phi}$, the centre $\overrightarrow{OC}$ and radius of the trajectory are expressed as follows

$$\frac{t}{\phi}\left[V_y^o, -V_x^o\right]^T, \quad \frac{\|\boldsymbol{V_0}\|t}{\phi} \tag{3}$$

Thus, the velocity vector $\boldsymbol{V_T}$ after t seconds is the initial velocity $\boldsymbol{V_0}$ rotated by ϕ.

$$\begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \left\{ -\frac{t}{\phi} \begin{bmatrix} V_y^o \\ -V_x^o \end{bmatrix} \right\} + \frac{t}{\phi} \begin{bmatrix} V_y^o \\ -V_x^o \end{bmatrix} \tag{4}$$

$|\boldsymbol{V_w}|^2$ is a constant independent of ϕ and $\phi = \dot{\phi}t$ ($0 \leq t \leq T$) is constant. Therefore, position $[x_T, y_T]^{\boldsymbol{T}}$ is represented as following linear trajectory.

$$\begin{bmatrix} x_T \\ y_T \end{bmatrix} = \frac{t}{\phi} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} V_y^o \\ -V_x^o \end{bmatrix} - \frac{t}{\phi} \begin{bmatrix} V_y^o \\ -V_x^o \end{bmatrix} \tag{5}$$

$$= \frac{1}{\dot{\phi}} \begin{bmatrix} -1 + \cos\dot{\phi}t & -\sin\dot{\phi}t \\ \sin\dot{\phi}t & -1 + \cos\dot{\phi}t \end{bmatrix} \begin{bmatrix} V_y^o \\ -V_x^o \end{bmatrix}$$

where

$$\phi = \dot{\phi}t \tag{6}$$

As shown in Figure 3, Robot trajectories are classified by $v_1 + v_2 + v_3$ and Following property can be valid.

[Property 1]

By the sign of $v_1 + v_2 + v_3$, the following holds.

(i) For $v_1 + v_2 + v_3 > 0$ ⇨ The movement is counter-clockwise rotation.

(ii) For $v_1 + v_2 + v_3 < 0$ ⇨ It moves in a clockwise rotation.

(B) Case 2 of $v_1 + v_2 + v_3 = 0$

In this case, since $\dot{\phi} = 0$ (translational motion only), $[x_T, y_T]^{\boldsymbol{T}}$ is represented as following line trajectory.

$$\begin{bmatrix} x_T \\ y_T \end{bmatrix} = \frac{1}{\dot{\phi}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} V_x^o \\ V_y^o \end{bmatrix} + \frac{1}{\dot{\phi}} \begin{bmatrix} V_y^o \\ -V_x^o \end{bmatrix} \tag{7}$$

$$\begin{bmatrix} x_T \\ y_T \end{bmatrix} = t \begin{bmatrix} V_x^o \\ V_y^o \end{bmatrix} \tag{8}$$

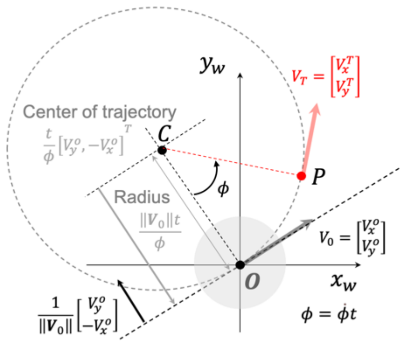


Figure 2 Property of robot trajectory in case of constant roller's speed

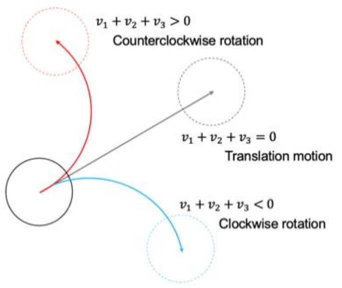


Figure 3 Classification of robot trajectory by input condition

Table 1 Parameter List

Parameter	Name
L	Radius of robot
r	Radius of roller
$v_i = r\omega_i$	Relationship roller speed and rotational speed
ω_i	Rotational speed
$N_i = 60\omega_i/360$	Number of revolutions [RPM]
$N_i = v_i/6r$	Relationship roller speed and rotational speed

3. Simulation

Using the derived Eq. (5) and Eq. (8), simulations on the mobile robot's movement trajectory were carried out. Furthermore, in order to verify whether the derived equations represent the intended conceptual model, simulations on the movement trajectory were conducted using Motion in the 3D CAD software SolidWorks. The results of the two simulations are shown in Figure 4. As shown in Table 2, The three-wheeled assembly of the mechanism with omni-rollers, giving a radius of L for the airframe, r for the omni-rollers and $(\theta_1, \theta_2, \theta_3)$ for the location of the three rollers and the friction is set by giving the contact constraint of the passive longs of the floor and the omni rollers as the analysis conditions.

For the error rate, the following is defined as an evaluation of the closeness of the theoretical value $\mathbb{r}_t(t) = (X_w^m(t), Y_w^m(t))$ to the experimental value $\mathbb{r}_e(t) = (X_w^e(t), Y_w^e(t))$.

Table 2 Parameter adjustment for simulation

Parameter	Value
L	0.108[m]
r	0.023[m]
$(\alpha_1, \alpha_2, \alpha_3)$	$(\pi/6, 5\pi/6, 3\pi/2)$ [rad]

$$e(t) = \frac{1}{T}\int_0^T \frac{\|\mathbb{r}_t(t) - \mathbb{r}_e(t)\|}{\|\mathbb{r}_t(t)\|}dt \qquad (9)$$

The theoretical equation calculated from Eq. (5) and Eq. (8) are shown by the orange line and the experimental values by the blue line.

(A) Simulation 1 (circular motion)

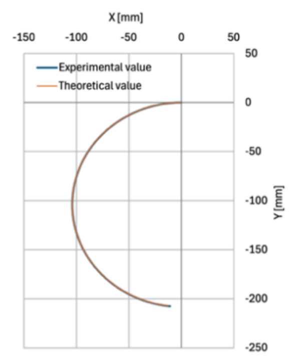
Assume that the origin is the start and the target of the initial velocity vector $\boldsymbol{V}$ such that $\varphi = 180°$ and $\|V\| = 0.022$ [m/s]. In this case, an input $(N_1, N_2, N_3) = (10, 10, 0)$[rpm] is given and time $t = 20$[s].

From $N_1 + N_2 + N_3 = 20 \neq 0$, a circular trajectory is theoretically drawn; in fact, the theoretical line shows a circular orbit close to the experimental line. The error rate was $|1 - L_w^e/L_w^m| = 0.037$.

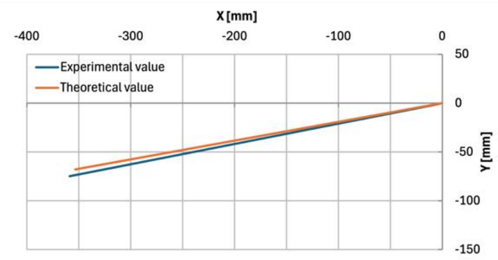
(B) Simulation 1(linear motion)

Assume that the origin is the start and the target of the initial velocity vector $\boldsymbol{V}$ such that $\varphi = 190.9°$ and $\|V\| = 0.10$ [m/s]. In this case, an input $(N_1, N_2, N_3) = (10, 20, -30)$[rpm] is given and time $t = 5$ [s].

From $N_1 + N_2 + N_3 = 0$, the nature of the trajectory implies a translational linear motion, but in fact, the theoretical line is close to the experimental line, indicating a linear trajectory. The error rate was $|1 - L_w^e/L_w^m| = 0.057$. The accuracy of the kinematics is therefore demonstrated.



(a) Case of circle



(b) Case of straight line

Figure 4 Trajectory of mobile robot in case of constant roller's speed

4. Conclusion

In this study, a trajectory equation for a mobile robot for a constant roller speed was derived. The theoretical equation was successfully verified by simulation. The theoretical equation proves that the trajectory of the robot is divided into cases (circles and lines) according to the sum of the input values, which is also confirmed by experiments.

As future work, we would like to study the speed efficiency of the mobile robot in all directions and the mobility efficiency of the mobile robot when the position and direction of the rollers are taken as parameters.

Acknowledgements

This work was supported by JSPS Grant-in-Aid for Scientific Research 23 K13279 and Kyushu Institute of Technology Collaboration fund.

Reference

1. J.Tang, K.Watanabe, et al., Autonomous control for an omnidirectional mobile robot with the orthogonal-wheel assembly, Journal of the Robotics Society of Japan. Vol. 17, No. 1,pp. 51-60, 1999.
2. Y.Yasohara, K.Shimizu et al., Development of ball handling mechanism for RoboCup MSL, *30th Fussy System Symposium*, pp. 616-617, 2014.
3. S. Chikushi, M. Kuwada, et al., Development of Next-Generation Soccer Robot"Musashi150"for RoboCup MSL,*30th Fussy System Symposium*, pp. 624-627,2014.
4. R. Junkai, X. Chenggang, X. Junhao et al., A control system for active ball handling in the RoboCup middle size League, *Chinese Control and Decision Conference(CCDC.)2016*.
5. K. Kimura, K. Ogata, K. Ishii, Novel Mathematical Modeling and Motion Analysis of a Sphere Considering Slipping, Journal of Robotics, Networking and Artificial Life, Vol.6, issue 1, pp. 27- 32, 2019.
6. K. Kimura, S. Chikushi, K. Ishii, Evaluation of the Roller Arrangements for the Ball-Dribbling Mechanisms adopted by RoboCup Teams, Journal of Robotics, Nrtworking and Artificial Life, Vol.6, issue 3, pp. 183-190, 2019.
7. S. Chikushi , T.Weerakoon , T.Sonoda , K. Ishii , Kinematics of Two-Roller-Driven Ball for RoboCup Soccer Robot, Journal of Robotics, Networking and Artificial Life, Vol. 4, No. 3, 248–253, 2017.
8. K.Kimura, K.Ishii, The Spherical Robot Transfer Problem With Minimal Total Kinetic Energy, Proceedings of International Conference on Artificial ALife and Robotics (ICAROB2021), Oita 266-270, 2022.
9. K.Kimura,Y.Abematsu, H.Hirai, Kazuo Ishii , Evaluation of Two Rollers Arrangement on a Hemisphere by Kinetic Energy, Journal of Robotics, Networking and Artificial Life, Vol.9, issue 3, pp.233-239, 2022.
10. K. Kimura, K. Ishii, Efficiency Problem of Spherical Robot in Transfer Kinetic Energy,Journal of Robotics, Networking and Artificial Life, Vol.9, issue 1, pp.87-92,2022.
11. K.Kimura, Y.Shigyo, K.Ishii,An Analysis of Robot Speed Efficiency for Mobile Robot Adapted Three Omni Rollers Using Linear Transformation, Journal of Advance in Artificial Life Robotics, Vol.3, issue 4, pp.242-249, 2023.
12. K.Kimura, K. Ishii, Kinetic Analysis of Omnidirectional Mobile Robot with Symmetry Roller's Arrangement, Journal of Robotics, Networking and Artificial Life, Vol.10, issue 3, pp.276-282, 2024.
13. K.Kato, M.Wada, Kinematic analysis and simulation of active-caster robotic drive with ball transmission (ACROBAT-S),AdvancedRobotics,2017.Vol31,No7,pp355-367,2017.

Authors Introduction

Dr. Kenji Kimura

He is an Associate Professor at the Department of Control Engineering, National Institute of Technology, Matsue College, and a Visiting Associate Professor at Chuo University. He received his MS (Master of Mathematical Science) from Kyushu University in 2002 and his PhD (Engineering) from the Kyushu Institute of Technology in 2020. His research interests are spherical mobile robots.

Mr. Kazuki Nakayama

He is 5th year student at the Department of Control Engineering, National Institute of Technology, Matsue College. His research interests include mobile robot and hand robot.

Dr. Katsuaki Suzuki

He is a Researcher at the Kumamoto Industrial Research Institute, Japan. He received his Ph.D. degree from the Kyushu Institute of Technology in 2021. His research interests include joint mechanisms and their applications.

Dr. Kazuo Ishii

He is a Professor in the Kyushu Institute of Technology, where he has been since 1996. He received his Ph.D. degree in engineering from University of Tokyo, in 1996. His research interests span both ship marine engineering and Intelligent Mechanics. He holds five patents derived from his research.

Cross-Disciplinary Learning Through Manufacturing: Toward Student-Centered STEAM Education

Kenji Kimura
Department of Control Engineering, National Institute of Technology, Matsue College
14-4 Nishi-ikuma-cho, Matsue-shi, Shimane,690-8518, Japan, Email: k-kimura@matsue-ct.ac.jp

Abstract

In recent years, with the importance of cross-disciplinary educational programs such as STEAM education, it has become necessary to provide mathematical education at the early stages of elementary and junior high school to prevent the increase in the number of students who have dropped out of science. As a result, educational institutions are also becoming more active in efforts such as robot-themed education as part of their contribution to the local community. In this study, we propose a method for students to decide their own theme about mechanics, obtain a production budget, and engage in cross-disciplinary learning through extracurricular activities. with support outside and inside the school.

Keywords: Cross-disciplinary learning, Student-centered educational activities, STEAM education, Pendulum clock

1. Introduction

In recent years, there have been classes, etc., that use LEGO parts, etc., in teaching "mechanics" as the mechanical aspect of robotics so that students can learn intuitively through experience [1]. This may be particularly effective when the target audience is elementary school students, and it is becoming increasingly important to conduct this as a regional collaboration. Such efforts are important to create interest in science at an early stage among elementary and junior high school students, to motivate them to study science-related subjects, and to stop young people from turning away from science and engineering.

As for spontaneous projects by university students, a sumo robotics class for elementary school students is planned [2], and a robotics class for elementary and junior high school students is planned to provide science and technology education [3].

In addition, as part of the development of teaching materials for robotics education, a programming teaching material was created using a commercial product called View trover, in which the theme was set as "maze exploration" and the final mission was successfully completed in 90 minutes [4]. In addition, we hosted a workshop and evaluated the usefulness of the lecture by taking pre- and post-questionnaires [5].

In addition, using a commercial product called Butte Balancer, we are developing teaching materials that can be used to convey the mechanism of inverted control, the relationship between mathematics and physics, and how it feels around us, in a cross-disciplinary manner [6]. Furthermore, a workshop was planned, and a class was conducted and evaluated so that participants could touch the robot and experience not only control engineering but also basic programming techniques, etc., with the aim of having them understand the concepts of control engineering and become interested in mathematical subjects in general, which are the basis for these subjects [7].

In this study, we will promote independent manufacturing activities for second-year students at technical colleges to enable students to engage in independent creative activities for Steam education.

1.1. *Curriculum of Department of Electronic Control Engineering*

Table 1 shows the curriculum for years 1-3 of the school's Department of Electronic Control Engineering.

In the first year, students learned control and programming using existing microcomputers in Basic Electronic Controls. In drafting, students learned design skills necessary for design. In mathematics and physics, students acquire the skills to identify phenomena and the basic knowledge to express engineering phenomena such as functions, which are necessary to study specialized subjects in the second and subsequent years.

In the second year, in CAD, students learn the skills to incorporate their own envisioned designs into their designs. In the Fundamentals of Electric Circuits course, students learn the basics of direct current and alternating current, and in the Engineering Experiments course, they re-learn their knowledge of electrical theory through measurement and measurement.

In the third year, there is a class for building creative robots called "Creative Design Exercise," and students are likely to be bound by time restraints even after school, The second year is considered to have relatively more time.

For the following reasons, it is considered appropriate to conduct such activities in the second year. Need to learn CAD ahead of time. Students who are highly motivated can study basic mathematical subjects on their own.

Table 1 Records of Out-of-School Activities

Date and Time	Special subject	Mathematical subjects
First year	Drafting Basic Electronic Control	Mathematics 1 Physics 1
2nd year	CAD Engineering Experiment Fundamentals of Electric Circuits	Mathematics 2 Physics 2
3rd year	Creative Design Exercise	Mathematics 3 Physics 3

1.2. *Toward the Challenge and Adoption of Shimane Mono*

Motivated students (second-year college of technology students) formed their own organization and set a goal of producing a clockwork mechanism. They made a production plan and created an application form for the Shimane High School Students' Monozukuri Challenge, and submitted it under the application theme of "Reproducing the mechanism of a mechanical clock using a 3D printer, etc. - Tourbillon production.

The proposal included a production plan toward a spring-loaded watch as a step toward tourbillon production. The plan included items to be mastered, such as skills in using CAD software, mathematical knowledge of mechanics, and material processing techniques necessary to design the watch.

This initiative was adopted and received a funding grant from the JST EDGE-PRIME Initiative Program at Shimane University. In addition, the Matsue Business Ecosystem, a MATSUE entrepreneurial ecosystem consortium, provides advice and technical assistance during regular monthly meetings at Shimane University.

2.3. Environmental maintenance and planning in the school

A workspace will be set up in a part of the laboratory where the fifth-year students belong, and the budget will be used to purchase a 3D printer and materials to improve the environment (money will be managed by the faculty member in charge). Students learn how to operate printers and learn CAD ahead of time, starting in the 5th grade. Basically, a 15-minute meeting once a week. The activity time is basically once a week, and students are responsible for their own activities after school.

Table 2 shows the required knowledge and skills corresponding to the plan and timeline for production.

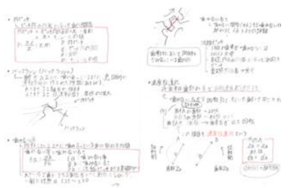
Table 2 Production Plan and Required Knowledge and Skills

Date and Time	Contents	Knowledge and Skills
May	Toward the Theme "Karakuri Clock"	Group Work and Emphasis
June	Using a 3D printer for teaching graduate students	How to design parts with a printer and CAD
July	Pendulum Mechanics Research Group Study Group on Mechanisms and Technology	Mathematical knowledge of mechanics
August	Gear design and assembly gear meshing considerations	Simulation Skills
September	Design of teaching materials	
October	Material processing work and assembly	Leather cutter and lathe work
November	assembly	-
December	operation test	-

2. Correspondence with STEAM Education

2.1. *Mechanism Study (SM:Science,Math)*

(a)



(b)

Figure 1 Extracurricular activities centered on Mechanics (a). Rotational presentations (b) Organization of learning content in mechanics

As shown in Figure 1, after school, the students share their knowledge with the group by giving presentations on what they have learned in a rotational format. Students are also required to study "mechanics" to understand how gears work. While studying this field, the students will realize the need for knowledge in the areas they have not yet studied in "physics" and "mathematics" and will be motivated to learn more.

2.2. *Design of Mechanism Materials (T: Technology)*

The design for producing a Karakuri clock was made by setting up gear parts in Slid Works (Figure 2(a)), fitting the gears produced through trial and error for gear meshing, and simulation (Figure 2(b)). The assembly was also completed by assembling the gear (Figure 2(c)(d)) (in the "CAD class," the students did not take assembly or simulation, so they taught themselves).

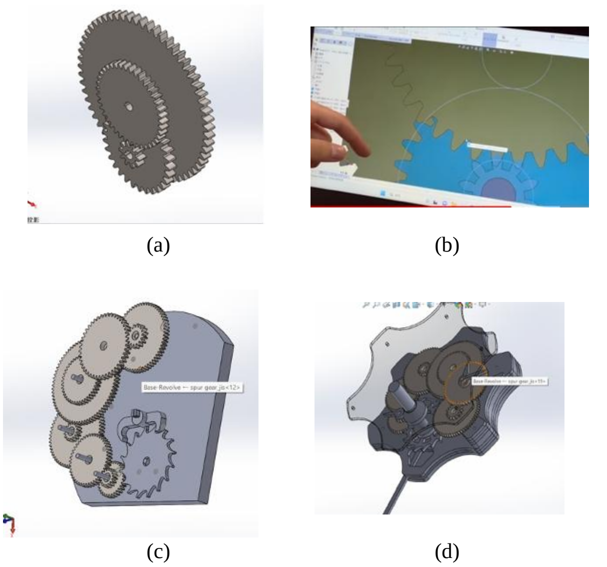


Figure 2 Process of creating the mechanical material (a). Gear design in SolidWorks (b). Simulation of gear meshing (c). Assembly (d). Completion of assembly

2.3. *Machining and fabrication of mechanical materials (E: Engineering)*

The CAD design data is converted to MC data and processed using a machine (Figure 3(a)). The completed parts are assembled. Finally, the operation is checked (Figure 3(b)). Since the students had not yet completed the "Practical training for machining in a factory", they fabricated parts using a laser cutter and a lathe with the assistance of technical staff.

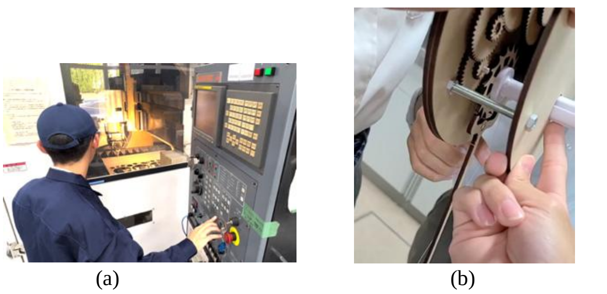


Figure 3 Fabrication process (a). Laser cutting board (b). Assembly completed and Operation check

2.4. *Pre-emptive learning and Steam compatibility*

In sections A-C, we describe the correspondence between the skills gained and the Steam items. Table 3 shows the knowledge and skills gained across disciplines in the process of learning mechanics. Thus, it shows that knowledge and skills were gained across disciplines, including physics, mathematics, CAD design and machining operations.

Table 3 STEAM Elements and Anticipatory Learning and Correspondence

STEAM Items	Skills gained
S: Science	Oscillation and period, equation of motion of a pendulum
T: Technology	Use of simulation software required for assembly and operation checks necessary to design, and MC data conversion necessary for machining
E: Engineering	Lathe, laser cutter processing, assembly work, hand finishing work
A: Art	Design as Design
M: Mathematics	Curves expressed in terms of parameters (involute curves), integration methods, infinite series, differential equations

In Science (S), students gained knowledge of basic physics to understand the phenomenon of pendulum oscillation; in Mathematics (M), students learned trigonometric functions to describe pendulum motion and differential calculus of involute curves through the rotational motion of gears. The students learned differential and integral calculus.

In Art (A), students designed a pendulum clock, and in Technology (T), they learned how to use a simulator to design and check its operation.

In Engineering (E), the participants gained machining operations and assembly skills in the fabrication of components.

2.5. *Self-assessment*

Table 4 shows the evaluation of the initiative from the perspective of the students and mentors of this activity.

As described above, the students gained a lot, as indicated by the mentors' comments. 1

Table 4 Student and Mentor Opinions

From the students	Mentor
Broadened my horizons. More weapons (skills) I have a stronger sense of responsibility.	Starting with a pure desire to "make something," the perspective was broadened to see how their own manufacturing could affect those around them (giving birth to a workshop project that will be the subject of a future issue).
I am now able to act in a team-oriented manner. I learned the importance of trusting people.	The team consulted and collaborated well, and everyone participated independently. Technical skills such as 3D CAD and 3D printers grew remarkably.

3. Conclusion

In this study, the students themselves were able to obtain a budget by respecting their strong desire to build a pendulum clock and by carefully preparing a production plan by the students. They also used support from outside the school to complete the manufacturing process. In this series of processes, the students learned across disciplines (mathematics, physics, CAD, and fabrication techniques) and improved their skills with a sense of STEAM education. This can be evaluated from the questionnaires from the students and the mentors who provided external support.

For future research, we would like to design a short and abbreviated version of the pendulum clock we produced and plan workshops for elementary and junior high school students to conduct classes with STEAM education in mind.

Acknowledgements

The following people provided advice and technical assistance. Thank you very much for your support. Advisors: Keisuke FUKUDA (President, KUTO), Ikko FUJIMURA (President, unoun design). Mentor: Keitaro Ohara (Shimane University)

This initiative was organized by the Shimane University Open Innovation Promotion Department and co-organized by the Matsue Business Ecosystem.

References

1. A.Matsuzaki, M. Isoda, A Reserch for the Integration Between Mathematics and Other Subjects Using Mechatronics-A inquiry of the piston crank mechanism with LEGO-,Proceedings of the 23rd Annual Meetion of JSSE & JSSE-ICASE-PME International Joint Conferenc
2. H.Yonemori, N.Tsuguta,D.Sugiyama,H.Hayashida,Student-Based Regional Collaboration Activities-Improvemrnt of Basic Skills of Working People by Manufacturing Education-, Japanese Society for Engineering Education JSEE, pp. 68-69,2020.
3. H. Ukida, A. Yoshida, K. Terada and S. Fujisawa, Tokushima Robot Programming Club: Science and Technology Education in Robot Manufacturing Class for Children by University Students, Journal of Robotics Society of Japan, Vlo.33, No3,pp. 154-163, 2015.
4. K. Kimura, Y. Takano, Development of Teaching Materials for Robot Programming for Junior High School Students: Student-Based Educational Activities, International Conference on Artificial Alife and Robotics(ICAROB),pp.607-610, 2024.
5. K. Kimura, Y. Takano, Programming Education Using Maze Exploration for Junior High School Students, Journal of Robotics, Networking and Artificial Life, Vol. 11, issue 4, pp.24-29, 2024.
6. Y. Takano, K. Kimura, Design and Software Production of Robotics Educational Design for Elementary and Junior High School Students,. International Conference on Artificial Alife and Robotics(ICAROB),pp.611-614, 2024.
7. Y. Takano, K. Kimura, Robotics Education for Elementary and Junior High School Students. Using Beauto Balancer, Journal of Advance in Artificial Life Robotics, Vol.4, issue 3, pp.146-152, 2024.

Authors Introduction

Dr. Kenji Kimura

He is an Associate Professor at the Department of Control Engineering, National Institute of Technology, Matsue College, and a Visiting Associate Professor at Chuo University. He received MS (Master of Mathematical Science) from Kyushu University in 2002 and his PhD (Engineering) from the Kyushu Institute of Technology in 2020. His research interests are spherical mobile robots.

Development of a Rotary Actuator Capable of Multidirectional Rapid Motion and Variable Stiffness

Katsuaki Suzuki
Kumamoto Industrial Research Institute, 3-11-38 Higashi machi, Higashi-ku, Kumamoto 862-0901, Japan

Yuya Nishida
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu 808-0196, Fukuoka, Japan

Kenji Kimura
Department of Control Engineering, National Institute of Technology, Matsue College, 14-4 Nishi-ikuma-cho, Matsue-shi, Shimane, 690-8518, Japan

Kazuo Ishii
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu 808-0196, Fukuoka, Japan
E-mail: k-suzuki@kumamoto-iri.jp, y-nishida@lsse.kyutech.ac.jp, k-kimura@matsue-ct.jp, ishii@brain.kyutech.ac.jp

Abstract

With the advancement of automation and digital transformation in the manufacturing industry, it is expected that industrial machines will be required to perform new tasks. Enhancing the multifunctionality of actuators is one approach to achieving these tasks. This paper proposes a new mechanism that combines two types of cams with different contour shapes, springs, two motors, and other mechanical components, and introduces an electric actuator incorporating this mechanism. The key feature of this actuator is its ability to achieve three functions (normal motion, rapid motion, and variable stiffness) while maintaining the same output characteristics, even when the initial posture of the output shaft is changed by switching the driving patterns of the two motors.

Keywords: Rapid motion, Variable stiffness, Cam mechanism

1. Introduction

The driving of output shafts using actuators and reduction gears (normal motion) is used in the driving of various machines, such as robot arm handling operations and gripper grasping operations. One way to increase the number of tasks that can be performed by robot arms and grippers is to improve the functionality of the joints and achieve operations other than normal motions. One such operation is driving the joints at high torque and high speed. To achieve this operation using actuators and reduction gears, a high-output motor is required, which would result in a larger size. Therefore, there is a method that uses the energy stored in the actuator by means of springs or compressed air, etc., and uses this as the driving source to drive the output axis (rapid motion) at a higher output than the actuator used for storage [1], [2]. In addition, one of the important functions for improving the functionality of the joints is the ability to adjust the rigidity. The most common methods for adjusting stiffness are compliance control and force control using force sensors, but these methods have problems such as difficulty in responding to external disturbances that exceed the control cycle. Therefore, there is a method for mechanically adjusting the stiffness of the joint (variable stiffness) using two actuators, a nonlinear spring element, and an antagonistic structure [3], [4].

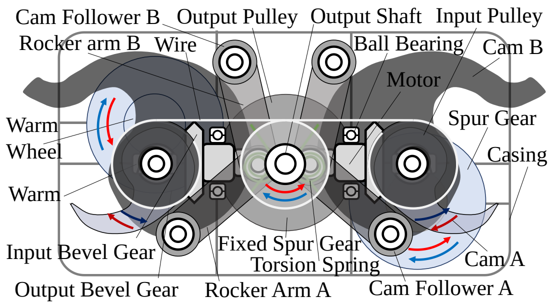
When multiple units of such individual mechanisms and actuators are mounted to expand functionality, the problems of size increase and control system complexity arise. Therefore, we developed a joint mechanism that can realize three functions (normal motion, rapid motion, and variable stiffness) using only two motors by integrating the mechanisms well [5], [6]. As a result of numerical analysis and actual device testing, we clarified that the three functions can be realized. However, when the initial posture of the output part of the developed joint mechanism is changed, it is difficult to realize normal motion, rapid motion, and variable stiffness with the same output. In addition, its range of motion is also limited. Therefore, in this study, we aim to develop a new rotary actuator that can achieve the same output characteristics for the three functions (normal motion, rapid motion, and variable stiffness) even when the output axis posture is changed, by combining two motors and several mechanical elements such as cams and springs. In this paper, we describe the design concept, structure, and drive method of the rotary actuator, and analyze rapid motion and discuss the results.

2. Design Concept and Overview

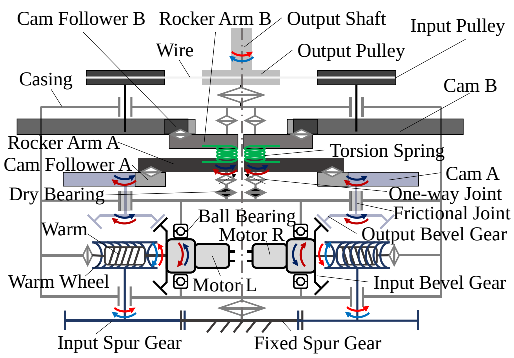
The design concept for the newly developed rotary actuator was set as follows.

- One output axis realizes three functions (normal motion, rapid motion, variable stiffness).
- Even if the output axis posture is changed, the three functions are realized with the same output characteristics.
- The above conditions are realized using two motors and several mechanical elements.

The rotary actuator, which was designed according to the design concept, is shown in Fig. 1. In order to realize the design concept, this actuator incorporates nine key ideas. The first is a system that switches the driving pattern of the two motors. This system is also used in the joint mechanisms we have developed so far, and it is possible to realize multiple functions by combining the rotational directions of the two motors [5], [6]. The second is a structure in which the output shaft, casing, and the rotation axis of the fixed spur gear are arranged on the same axis. With this structure, if the casing is rotated with respect to the fixed gear, the output shaft can also rotate at the same angular velocity as the casing. Furthermore, the drive shaft can also be rotated with respect to the casing. The third is a cam mechanism developed by Sonoda et al. that achieves variable stiffness [4]. Due to the antagonistic structure and the increase in the radius of the cam, the reduction ratio decreases, so it is possible to output an anti-torque against external disturbances acting on the output shaft. Fourth, there is a cam mechanism developed by Amil et al. that achieves energy-saving compression and release of the spring [2]. Due to the increase in the radius of the cam, the reduction ratio increases, so it is possible to compress the spring with energy savings. In addition, a notch is provided in the cam follower that allows it to drop suddenly, enabling the release of the spring. The fifth is a cam mechanism that achieves a self-locking function. This mechanism intentionally sets the pressure angle for cam mechanism A. Furthermore, a one-way joint and a dry bearing are incorporated into the rotating joint of rocker arm A. The vertical resistance force generated by the pressure angle and the friction coefficient of the dry bearing can generate a large frictional force in one direction of rotation. Therefore, when an external disturbance acts on the output shaft, it is possible to prevent back driving of Cam A. The sixth feature is a structure in which two cams perform work on a single spring that is incorporated between the two cam mechanisms. By sharing a single spring between two cam mechanisms with different purposes, it is possible to achieve a cam mechanism with no trade-offs in each mode. The seventh feature is a structure that switches between motor shaft drive and casing drive. This structure enables the motor shaft drive and cam drive to be separated. The eighth is the self-locking function of the worm gear. And the ninth is the structure that the casing does not rotate unless two spur gears rotate in the same direction at the same time. These two ideas prevent the casing back drive against the fixed gear when a disturbance acts on the output shaft or when the motor is driven on only one side. And the output shaft can be back-driven against the casing. By optimally



(a) Top view



(b) Front view

Fig.1 Rotary actuator

Table 1 Drive Pattern

Motor L	Motor R	Torsion Spring	Output Shaft	State
Forward	Forward	Natular position	Forward	Normal motion
Reverse	Reverse		Reverse	Low Stiffness
Forward	Forward	Compressed	Forward	Normal motion
Reverse	Reverse		Reverse	High Stiffness
Reverse	Forward	Compression	Stop	Compression
Forward	Reverse	Extension		Adjustment
Reverse	Stop	Rapid	Forward	Rapid motion
Stop	Forward	Extension	Reverse	Spring release
Stop	Stop	Natular position	Forward	Disturbance acts
			Reverse	Low Stiffness
Stop	Stop	Compressed	Forward	Disturbance acts
			Reverse	Low Stiffness

combining these nine technologies, we have achieved an actuator that satisfies the design concept.

Table 1 shows the correspondence between drive patterns and functions. When the output shaft is operated normally, Motors L and R rotate in the same direction. This causes the casing body to rotate, and the output shaft to rotate. To adjust the spring force, rotate Motors L and R in opposite directions. In this casing, the two cams A rotate via the bevel gear. The combination of the rotation direction of the two motors achieves compression and extension of the spring, and the output shaft rigidity can be adjusted. To move the output shaft in the forward direction momentarily, first rotate motors L and R in the opposite direction to move cam follower A to the vicinity of cam A's maximum radius. Next, if you reverse the rotation of Motor L only, the cam follower will fall out of the notch, and the output shaft will rotate via Cam Mechanism B due to the force of the torsion spring. If you reverse the rotation direction of Motor L and Motor R, you can drive in the reverse direction. If the two motors are stopped and an external torque acts on the output shaft, the output shaft will back-drive against the casing. In this way, because the position of the casing itself changes during normal motion, it is possible to achieve three functions with the same output characteristics for various angles.

3. Analysis of rapid motion

Using cam B, which is designed for variable stiffness, we will check the output characteristics when performing rapid motions in the analysis. The model of the cam mechanism B used in the analysis is shown in Fig. 2. In this model, we assume that the pressure angle when the cam and cam follower come into contact is always 90 degrees. By setting the pressure angle to 90 degrees, the force acting in the horizontal direction to the rocker arm can be canceled, so the loss due to rolling friction, etc. can be reduced[2]. In addition, when the angle of the rocker arm θ_r is 0 degrees, the distance between the cam follower and the cam's rotational joint becomes l_{cmin}, and d in Fig. 2 becomes 0. Also, because the pressure angle is always 90 degrees, as θ_r increases, d increases nonlinearly and rapidly.

When the radii of the input pulley and output pulley, which are on the same axis as cam B, are the same, the torque acting on the output shaft, cam B, and rocker arm can be expressed by the following formula using the distance d and rocker arm length l_r, based on geometric relationships.

$$T_o = T_c = \frac{d}{l_r} T_r \tag{1}$$

Therefore, the reduction ratio between the output link and the rocker arm B is expressed by the following formula.

The value of d changes as the angle of rotation of the

$$G = \frac{d}{l_r} \tag{2}$$

rocker arm changes. If G is expressed as a variable θ_r and various constants, it can be expressed by the following formula due to geometric relationships.

$$G = 1 - cos\theta_r + \frac{l_{cmin}}{l_r} sin\theta_r \tag{3}$$

The torque accumulated in a torsion spring is expressed by the following formula from Hooke's law, based on the spring constant and angular displacement of the torsion spring.

$$T_r = k_{sp} \delta\theta_r \tag{4}$$

Therefore, the torque acting on the cam during a quick movement can be expressed by the following formula.

$$T_o = k_{sp}(1 - cos\theta_r + \frac{l_{cmin}}{l_r} sin\theta_r)\delta\theta_r \tag{5}$$

The reduction ratio G and output torque T_o analyzed using Eq. (3) and Eq. (5) are shown in Fig. 3, and the design parameters used in the analysis are shown in Table 2. Rocker arm B is defined as being able to rotate from to . The initial posture of rocker arm B at the start of rapid motion is $ini(\theta_r)$. The displacement of the spring charged by Cam A is $max(\delta\theta_r)$. If you check the reduction ratio in Fig. 3, you will see that the reduction ratio is always less than 1. Also, the reduction ratio decreases rapidly as the output shaft rotates. This is because the moment arm decreases as the cam rotates. The torque characteristics show a rapid decrease in the initial stages of operation. This is because the stored spring force decreases and the reduction ratio increases rapidly. It has been theoretically proven that when rapid motion is performed using a cam designed for variable stiffness, the output characteristics will be more focused on speed than torque.

Table 2. Design parameters

k_{sp}	Spring constant	10Nm/rad
Rf	Radius of Cam follower	0.005m
R_{cmin}	Minimum radius of Cam A	0.01 m
l_r	Length Rocker arm A	0.05 m
$max(\theta_r)$	Maximum movement range	+45°
$min(\theta_r)$	Minimum movement range	+5°
$ini(\theta_r)$	Initial Position	+22.5°
$max(\delta\theta_r)$	Maximum Angular Displacement	17.5°

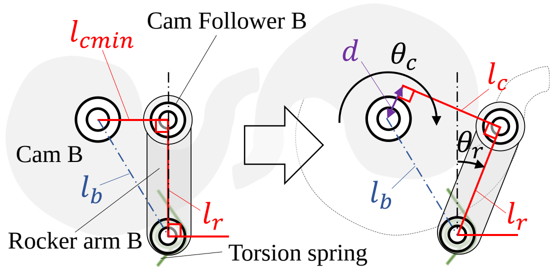


Fig. 2 Cam B model

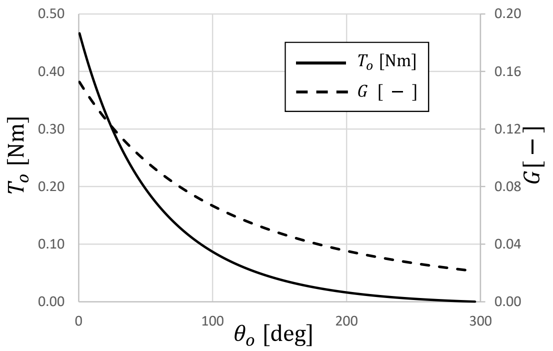


Fig. 3 Output torque and reduction ratio during rapid

4. Conclusion

In this paper, we have presented a new actuator design concept, structure, and method for realizing each function. The proposed mechanism has the potential to realize three functions in various directions while maintaining similar output characteristics. In addition, we analyzed the output characteristics when rapid motion is performed using a cam designed for variable stiffness. The reduction ratio decreases rapidly as the output shaft rotates, making it effective for applications that emphasize speed. In the future, we will analyze variable stiffness and conduct actual machine tests to evaluate the feasibility of each function.

References

1. ITO, Fumio, et al. Instantaneous force generation mechanism based on the striking motion of mantis shrimp-Analytical and experimental verification of the increase in instantaneous force using exoskeleton spring mechanism. IEEE Robotics and Automation Letters, 2021, 6.4: 6678-6685.
2. Nassiraei, Amir AF, et al. Realization of the rapid movements for the entertainment robots by using two new actuators "inertia actuatror" and "CAM charger". In: ASME International Mechanical Engineering Congress and Exposition, 2006. p. 1291-1297.
3. MENGACCI, Riccardo, et al. Overcoming the torque/stiffness range tradeoff in antagonistic variable stiffness actuators. IEEE/ASME Transactions on Mechatronics, 2021, 26.6: 3186-3197.
4. Sonoda, Takashi, et al. Development of antagonistic wire-driven joint employing kinematic transmission mechanism. Journal of Automation Mobile Robotics and Intelligent Systems, 2010, 4: 62-70.
5. Suzuki, Katsuaki, et al. A new rotary actuator capable of rapid motion using an antagonistic cam mechanism. Journal of Advances in Artificial Life Robotics, 2020, 1.3: 143-151.
6. Suzuki, Katsuaki, et al. Realizing Variable Stiffness through Structures Designed for Quick Motion. Journal of Advances in Artificial Life Robotics, 2023, 4.3: 122-127.

Authors Introduction

Dr. Katsuaki Suzuki

He is a Researcher at the Kumamoto Industrial Research Institute, Japan. He received his Ph.D. degree from the Kyushu Institute of Technology in 2021. His research interests include joint mechanisms and their applications.

Dr. Yuya Nishida

He is an Associate Professor at the Graduate School of Life Science and System Engineering, Kyushu Institute of Technology, Japan. His research interests include filed robotics, its application, and data processing.

Dr. Kenji Kimura

He is a Lecturer at the Department of Control Engineering, National Institute of Technology, Matsue College, and a Visiting Associate Professor at Chuo University. He received his ME (mathematics) from Kyushu University in 2002 and his PhD (engineering) from the Kyushu Institute of Technology in 2020. His research interests are spherical mobile robots.

Prof. Kazuo Ishii

He received his Ph.D. degrees from Tokyo University, Japan, in 1996. In 2011, he joined Kyushu Institute of Technology, where he is currently a Professor of the Department of Human Intelligence Systems.
His research interests include Underwater Robot, Neural Network and Neural Network.

Study of Evaluation Operation Log Analysis Using 2^3- ERC on Matsue National College of Technology

Takumi Ueda
National Institute of Technology, Kurume College, 1-1 Komorino, Kurume, 830-8555, Japan
Email: ueda_517@kurume-nct.ac.jp

So Takei
National Institute of Technology, Kurume College, 1-1 Komorino, Kurume, 830-8555, Japan
Email: e62122st@kurume-nct.ac.jp

Akira Nakano
National Institute of Technology, Kurume College, 1-1 Komorino, Kurume, 830-8555, Japan
Email: nakano@kurume-nct.ac.jp

Kenji Kimura
National Institute of Technology, Matsue College, 14-4, Nishi-Ikuma-Cho, Matsue, 690-8518 Japan
Email: k-kimura@matsue-ct.ac.jp

Kazutaka Matsuzaki
Nishinippon Institute of Technology, 1-11, Aratsu, Kanda-Machi, Miyako-Gun, 800-0394, Japan
Email: matuzaki@nishitech.ac.jp

Abstract

In response to the demand for educational proposals that address advancements in science and technology in Japanese school education, cross-disciplinary education and STEAM education are receiving increased attention. Given this context, a new unit called "Mathematics and Human Activities" was introduced. However, the need for innovative mathematics teaching materials is necessary. We develop the 2^3 Electric Rubik's Cube (2^3-ERC) with two LEDs for each cube edge for easy tractable operation logging data. By utilizing the operation log data as feedback, we evaluated log data analysis through the experiment on Matsue National Institute of Technology's students. The results from the operation log shows the level of understanding of 2^3-ERC from the number of operations and consideration time.

Keywords: Mathematical education, Rubik's Cube, Unfolded, 2^3-ERC, Logging analysis,IDDFS

1. Introduction

In recent years, the demand for educational methods capable of addressing rapidly advancing science and technology has grown significantly in educational settings. As a result, STEAM (Engineering, Technology, Engineering, Arts, Mathematics) education has garnered considerable attention [1],[2],[3],[4],[5]. In Japan's high school mathematics curriculum, a new unit, Mathematics and Human Activities, has been added [6]. This unit features sections such as Properties of Integers, Properties of Coordinates, and Mathematics in Games and Puzzles.

Against this backdrop, we focused on the widely recognized Rubik's Cube and developed the 2^3-ERC, with each face consisting of 2×2 blocks. The developed 2^3-ERC can be connected to a PC via serial communication to collect operational data. However, there has been insufficient discussion regarding standard methods for analyzing the data obtained from the 2^3-ERC. Thus, the development of methods for evaluating learners' characteristics is required. In this study, we propose a method for analyzing data collected using the 2^3-ERC through the IDDFS (Iterative Deepening Depth First Search) algorithm. IDDFS combines the advantages of depth-first search and breadth-first search, offering high memory efficiency and the ability to explore solutions effectively while minimizing computational overhead. Specifically, IDDFS progressively relaxes the depth limit during its search, making it an effective method for solving Rubik's Cubes, as it guarantees a solution and produces the shortest path to the goal.

This paper presents the results of an experiment conducted with students from Matsue National College of Technology using the 2^3-ERC. The operational logs obtained from the experiment were analyzed using IDDFS, and the findings are discussed herein.

2. 2^3 − ERC teaching aid

2.1. About $2^3 - \boldsymbol{ERC}$

In this study, we propose a new evaluation method for the 2^3 − ERC.Fig.1 shows the conceptual diagram of this tool. The 2^3 − ERC is equipped with 24 LEDs and 48 operational buttons, along with a function to record the learner's operational data via serial communication with a computer. Fig.2 presents the developed 2^3 − ERC. As shown in Fig.2, each face of the tool is embedded with RGB full-color LEDs, and the LED arrangement pattern changes in a manner similar to the rotational movement of a Rubik's Cube when the operational buttons are pressed. By interacting with this tool, learners are expected to develop their spatial and pattern recognition abilities, as well as logical thinking skills.

2.2. *Mathematical models*

As shown in Fig.3, each face of the 2^3 − ERC is assigned a face identification index from A to F, and the four RGB full-color LEDs on each face are assigned indices from 1 to 4. For example, on face A, the top-left LED is designated as A_1, and the bottom-right LED as A_4. To handle the color information of these LEDs, we define a vector $\boldsymbol{q}_n$, as shown in Eq. (1), which contains 24 elements. The color information represented in $\boldsymbol{q}_n$ is treated as the index of a color palette corresponding to each color.

$$\boldsymbol{q_n} = (A_1, A_2, A_3, A_4, \cdots, F_1, F_2, F_3, F_4) \in \mathbb{R}^{24} \tag{1}$$

Next, the rotation rule for rotating $\boldsymbol{q}_n$ is given by Eq. (2). Here, Eq. (2) represents a permutation matrix.

$$\boldsymbol{T_X}, \boldsymbol{T_Y}, \boldsymbol{T_Z}, \boldsymbol{T_{X^{-1}}}, \boldsymbol{T_{Y^{-1}}}, \boldsymbol{T_{Z^{-1}}} \in \mathbb{R}^{24\times 24} \tag{2}$$

Here, using a natural number p, Eq. (2) possesses the following fundamental property, as shown in Eq. (3).

$$\begin{aligned} \boldsymbol{T}_*^{4p} &= \boldsymbol{T}_{*^{-1}}^{4p} = \boldsymbol{E}, \quad & \boldsymbol{T}_*^{4p+1} &= \boldsymbol{T}_{*^{-1}}^{4p+3} \\ \boldsymbol{T}_*^{4p+2} &= \boldsymbol{T}_{*^{-1}}^{4p+2}, \quad & \boldsymbol{T}_*^{4p+3} &= \boldsymbol{T}_{*^{-1}}^{4p+1} \end{aligned} \tag{3}$$

The rotation rules for the 2^3 − ERC consist of three types of forward rotations $\boldsymbol{T_X}, \boldsymbol{T_Y}, \boldsymbol{T_Z}$ and their corresponding reverse rotations $\boldsymbol{T_{X^{-1}}}, \boldsymbol{T_{Y^{-1}}}, \boldsymbol{T_{Z^{-1}}}$, for a total of six types. These six types fully represent the motion of the 2^3 − ERC. Among these, two rotations are selected to form valid combinations, as shown in Eq. (4). The valid combinations consist of 12 types, and including their reverse rotations, there are 24 types in total.

$$\begin{aligned} &\boldsymbol{T_X T_Y}, \boldsymbol{T_X T_Z}, \boldsymbol{T_X T_{Y^{-1}}}, \boldsymbol{T_X T_{Z^{-1}}} \\ &\boldsymbol{T_Y T_X}, \boldsymbol{T_Y T_Z}, \boldsymbol{T_Y T_{Z^{-1}}}, \boldsymbol{T_Y T_{X^{-1}}} \\ &\boldsymbol{T_Z T_X}, \boldsymbol{T_Z T_Y}, \boldsymbol{T_Z T_{X^{-1}}}, \boldsymbol{T_Z T_{Y^{-1}}} \end{aligned} \tag{4}$$

Hence, if the matrix that rotates each element of the 2^3 − ERC is denoted as T, the permutation matrix T from the initial state $\boldsymbol{q_0}$ to $\boldsymbol{q_n}$ can be defined as Eq. (5).

$$\begin{aligned} &\boldsymbol{q_n} = \boldsymbol{T}\boldsymbol{q_0} \in \mathbb{R}^{24} \\ &\boldsymbol{T} \in \prod_{i \in \boldsymbol{S}} \boldsymbol{X_i} \left(\boldsymbol{S} \subseteq \{\boldsymbol{T_X}, \boldsymbol{T_Y}, \boldsymbol{T_Z}, \boldsymbol{T_{X^{-1}}}, \boldsymbol{T_{Y^{-1}}}, \boldsymbol{T_{Z^{-1}}}\}\right) \end{aligned} \tag{5}$$

Here,$|\boldsymbol{S}| > 0$.

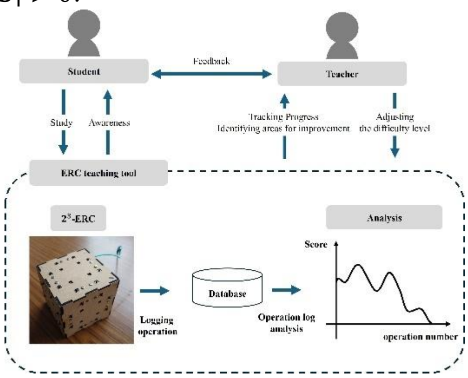


Fig.1 Overview of the 2^3 − ERC teaching system.

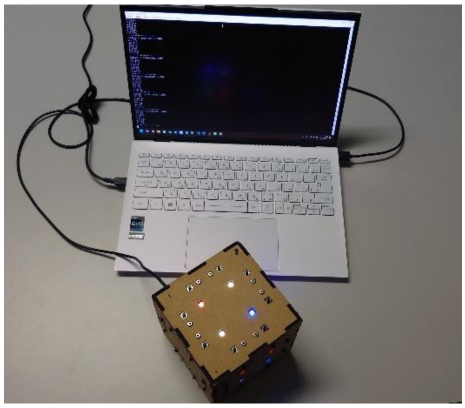


Fig.2 The 2^3 − ERC teaching material

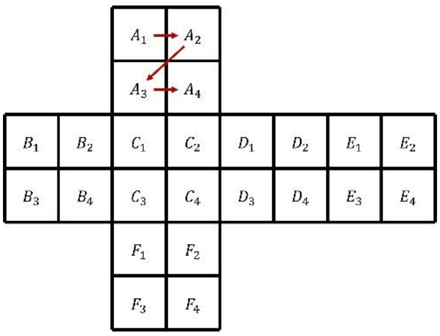


Fig.3 Surface Identification Indices Assigned from A_1 to F_4

3. Solution Search and Evaluation Using IDDFS

The operational data collected from learners using the 2^3 − ERC needs to be evaluated with an appropriate method. While various approaches can be considered for evaluating the data obtained from the 2^3 − ERC, this paper employs IDDFS and evaluates the results based on the shortest path obtained by this method.

Fig. 4 shows the flowchart of IDDFS. The solution search and evaluation method using IDDFS is outlined in Steps 1 through 4 below:

Step. 1 Initialization

Initialize the solution $\tilde{T}$ obtained by IDDFS as the identity matrix and set the depth limit for the search to 0.

Step. 2 Perform Depth-Limited Depth-First Search

Conduct a depth-first search up to the restricted depth range. If the restricted depth is reached, terminate the search.

Step. 3 Goal State Check

If $T\tilde{\boldsymbol{T}}$ is the goal state, end the search and proceed to Step 4. Otherwise, increment the depth limit by 1 and return to Step 2.

Step.4 Evaluation

The evaluation value L is given by the difference between the total ideal minimum number of rotations for each state operated by the learner and the total number of rotations from the initial state to the ideal state, as shown in Eq. (6). Here, β_i represents the minimum number of rotations required to reach the goal state from the current permutation state $\tilde{\boldsymbol{T}}$ using unit rotations. α denotes the number of moves required to transition from the initial state to the goal state as assigned to the learner, and m represents the total number of operations performed by the learner up to the current point.

$$L = -\frac{\alpha(\alpha+1)}{2} + \sum_{i=1}^{m} \beta_i \qquad (6)$$

4. Experiment

To verify the effectiveness of the proposed method, an experiment was conducted with 9 students from Matsue National College of Technology. In this experiment, the initial state of the $2^3 -$ ERC was set to $\boldsymbol{T}_{Y^{-1}}\boldsymbol{T}_X$, and the time required to return it to the target state $\boldsymbol{E}$ as well as the operation steps were evaluated. The experiment continued until the students either completed the task or decided to discontinue the task and requested to terminate the experiment.

5. Results and Discussion

The experimental results using the $2^3 -$ ERC are presented in Table 1, and the relationship between the number of steps and evaluation values is shown in Fig. 5. In the experiment, all participants successfully achieved the target. Four students completed the task with the shortest number of steps, two succeeded in four steps, and three required more than ten steps. The average number of steps was 6.88, with an average thinking time of 4.48 seconds per step. Additionally, the average task completion time was 33.4 seconds.

Table.1 Results of assigning tasks achievable in two moves to subjects

	Total time	Number of manipulative moves	L	Task achievement
S1	1'13"3	18	56	○
S2	0'02"5	2	0	○
S3	0'12"3	10	18	○
S4	0'50"4	2	0	○
S5	0'17"2	2	0	○
S6	0'19"2	2	0	○
S7	1'29"4	14	36	○
S8	0'15"1	4	5	○
S9	0'12"3	4	5	○

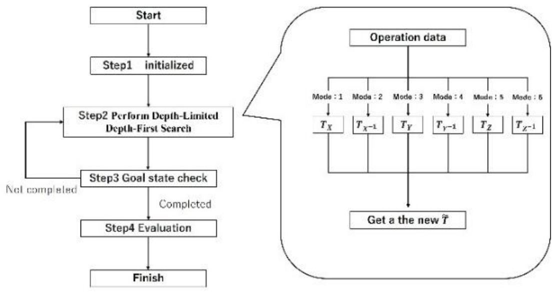


Fig.4 Flowchart of the IDDFS method

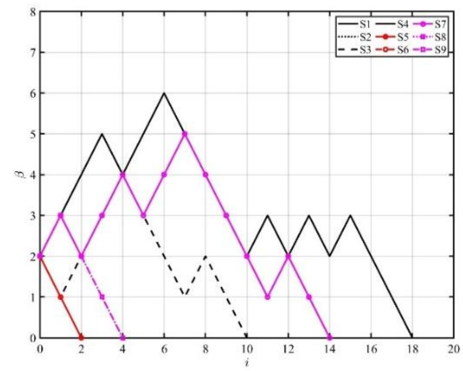


Fig.5 Results of Evaluation Function L Values from IDDFS Method Calculation

Conclusion

In this study, we proposed the evaluation method for the $2^3 -$ ERC, and evaluated its effectiveness. Through an experiment involving nine students, the validity of the proposed method was confirmed. All participants successfully completed the task, with an average completion time of 33.4 seconds and an average step count of 6.88. These results suggest that the $2^3 -$ ERC has the potential to enhance logical thinking and problem-solving skills.

Future work will involve conducting experiments with a larger sample size to examine the applicability of the $2^3 -$ ERC in various educational settings. Additionally, incorporating features such as automated feedback and adaptive difficulty adjustments will aim to further enhance the educational value of this tool.

References

1. R. Kobayashi, A Study of Utilizing Mathematical Perspectives and Ways of Thinking in STEM/STEAM Education (in Japanese), *Proceedings of the Annual Meeting of the Japan Society for Science Education* vol.47, pp.145-148,2023
2. T. Kitazawa, Study on In-School Training of STEAM Education in an Elementary School (in Japanese), *JSSE Research Report* vol.36(6), pp.13-16,2022
3. S. Takada, The Practical Science of STEM/STEAM Education, Interwoven with the Horizontal and Vertical Threads of Mathematical Sciences: Double Majors and Recurrent Education (in Japanese), *The Japanese journal of educational research* vol.80, pp.249-250,2021
4. K. Sakaguchi, Proposal for "STEAM-based IT education "to develop advanced IT personnel (in Japanese), *Japan Society for Educational Technology* vol.44(3), pp.357-363,2021.
5. M. Yamamoto, M. Yukawa, K. Takatsuka, Practical study of STEAM education across subjects of "Science", "Art and Handcraft "and "Music" (in Japanese), *Proceedings of the Annual Meeting of the Japan Society for Science Education* vol.43, pp.89-90,2019.
6. Ministry of Education, Culture, Sports, Science and Technology, The Courses of Study for Upper Secondary Education in Japan 2022(in Japanese), *Ministry of Education, Culture, Sports, Science and Technology*, 2022

Authors Introduction

Dr. Takumi Ueda

He is an Assistant Professor in the Department of Control and Information Systems Engineering at the National Institute of Technology, Kurume College, Japan. He received his Ph.D. in Engineering from the Kyushu Institute of Technology, Japan, in 2024. His research field is database-driven PID control design.

Mr. So Takei

He is a student at the Department of Electrical and Electronic Engineering, National Institute of Technology, Kurume College. He is involved in robot development in the Robot Contest Club. His research field is Robotics and school education.

Dr. Akira Nakano

He is an Associate Professor in the Department of Control and Information Systems Engineering at the National Institute of Technology, Kurume College, Japan. He received his Ph.D. in Engineering from the Kyushu Institute of Technology, Japan, in 2003. His research field is EdTech, focusing on Computer-Assisted Instruction and Educational Data Mining.

Dr. Kenji Kimura

He is an Associate Professor at the Department of Control Engineering, National Institute of Technology, Matsue College, Japan, and a Visiting Associate Professor at Chuo University. He received his Master of Mathematical Science from Kyushu University, Japan, in 2002 and his Ph.D. in Engineering from the Kyushu Institute of Technology, Japan, in 2020. His research interests are spherical mobile robots.

Dr. Kazutaka Matsuzaki

He is a Professor at the Faculty of Technology, Nishinippon Institute of Technology, Japan. He received his master of mathematical sciences from University of Tokyo, Japan, in 2001 and his doctor of philosophy (education) from Hyogo University of Teacher Education, Japan, in 2017. His research interests are school education.

Efficient Ball Position Estimation for Tennis Court Robot Assistants using Dual-Camera System

Abdullah Alraee
Department of Human Intelligence Systems, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka, 808-0196 Japan.
E-mail: alraee.abdullah-abdul648@mail.kyutech.jp

Hussam Alraie
Department of Electrical and Electronic Engineering, Middle East Technical University,
Guzelyurt, Turkish Republic of Northern Cyprus.
E-mail: hussam@metu.edu.tr

Mohammad Albaroudi
Department of Human Intelligence Systems, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka, 808-0196 Japan.
E-mail: albaroudi.mohammad344@mail.kyutech.jp

Raji Alahmad
Department of Human Intelligence Systems, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka, 808-0196 Japan.
E-mail: raji@brain.kyutech.ac.jp

Kazuo Ishii
Department of Human Intelligence Systems, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka, 808-0196 Japan.
E-mail: ishii@brain.kyutech.ac.jp,

Abstract

During tennis training, professional players use many balls that will be distributed randomly around the tennis court. Collecting the ball can be done manually, However, the effort and the time make it inefficient. The mobile robot for ball collection has been introduced as a solution to save power and training session time. The robot tasks involve several steps: ball detection, estimating their positions, and finding the best path for efficient collection. In our previous work, we addressed the ball detection issue utilizing a neural network algorithm using YOLOv8. This study focuses on the next step, ball position estimation, using two cameras to cover the entire court. The results demonstrate successful ball position estimation along the x-axis and y-axis, achieving an accuracy of 97.76 %.

Keywords: AI, Position, detection, robot, Dual-Camera

1. Introduction

One key requirement Human-robot interaction (HRI) in sports has been the subject of extensive research recently, and robots are being developed to replace humans as a form of sports training and practice that was previously confined to human life. Usually, Robots connect to the camera and the video frames or images can be analyzed using several methods [1]. The tennis court, especially the open grass court, is one of the spaces where a lot of physical activity takes place. It is also the place where tennis matches are held, and the audience can enjoy them live. Therefore, the development of robots that support various tennis matches and enhance the enjoyment of these matches is a meaningful research direction from both scientific and engineering standpoints. However, the development of tennis court robots for assistance in various tennis games or research is still very rare. For tennis court robots to support various tennis games or sports research they can visually detect the tennis ball and predict its position [2], [4]. There may be other requirements, but a few key points are of immediate importance. For example, priority is given to visual processing time rather than the position estimate error in real-time tennis ball tracking for operating the on-court robot. Existing literature on robots for sports training can be divided into two classes: the robots that tend to be trainers and the robots that tend to be trainees [3], [5]. On the one hand, some studies have been conducted on the development of robot trainers for tennis. The robot can swing the golf club precisely and measure the hit ball direction using optical sensors. Another tennis assistant club uses servomotors to control the swing of the stick, and a belt embedded with metal tracks to assist the player in controlling the stick in real-time. On the other hand, an indoor tennis robot can also be found in some studies [6]. This robot was designed for professional players to improve their game. It consists of a mobile shooting platform, a feeder arm with adjustable shot capability, and advanced software to control speed, direction, and spin, along with helpful features such as target point pre-selection. However, the design of this

Robots are complex and expensive to produce. With this understanding, this paper presents a novel visual system that uses two cameras to efficiently estimate the current ball position on the tennis court. The challenge of the tennis ball mobile robot application is to calculate the distance from the camera. Knowing the distance of the ball from the camera helps the robot go directly to the ball, unlike scanning the entire tennis court, which contributes to increasing the efficiency of the detection system and saving training session time.

2. Methodology

The distance had been calculated using one camera based on the relationship between the angle in the image, and the angle in the real world as shown in Fig.1. The accuracy was 71%.

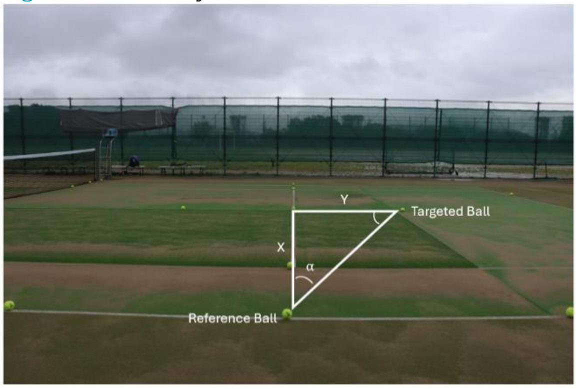


Fig.1. Position calculation using one camera

For this reason, two cameras were used to measure the distance accurately. To find the tennis ball's position, we assume that the origin is the camera for both coordinates X and Y as shown in Fig.2.

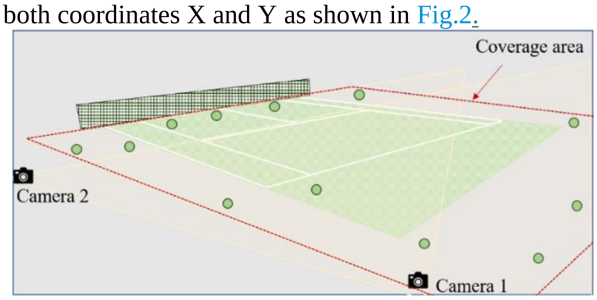


Fig.2. Dual-camera system for tennis ball detection

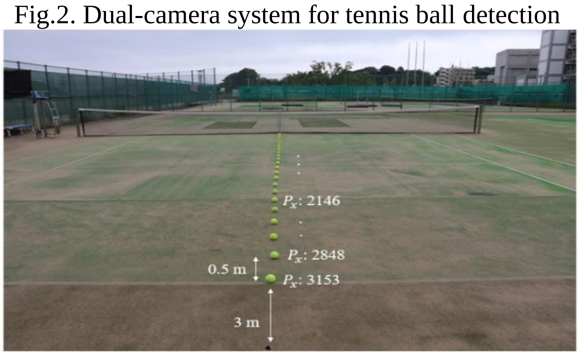


Fig.3. Experimental setup for x-ais

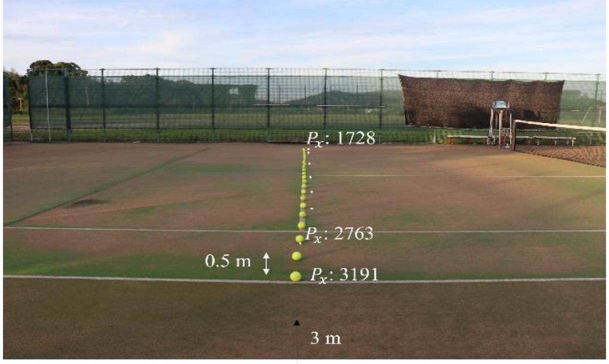


Fig.4. Experimental setup for Y-ais

Then by measuring the pixel size when we change the distance from the camera, a mathematical relationship between the distance and the pixel size is found. Images have been taken at a distance of 3 meters, which will be the offset as shown in Fig.3 and Fig.4 illustrating pixel changing when distance is changed. The tennis court is located at the Kyushu Institute of Technology Sports Center.

3. Results and Discussion

The experiment results were shown for 3- and 4-meter offsets to calculate the X- and Y-axis coordinates. The calculated coordinates were compared with the actual distances.

3.1 Estimate the position for 3M

The relationship between the distance and the pixel has been concluded as shown in Fig.5.

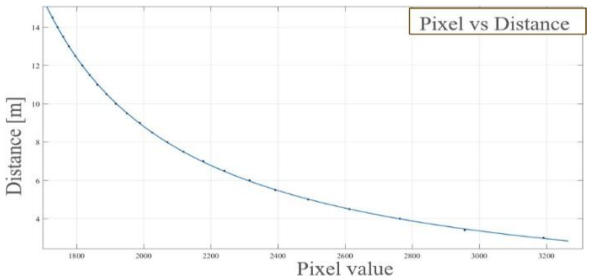


Fig.5. Pixel value vs the actual distance on the X-axis

The curve fitting of the plotted extracted in a mathematical formula is shown in the following equation:

$$D_x = \frac{\beta\,(P_x) + \mu}{(P_x + \mathcal{E}\,)}$$

Where D_x is the calculated distance coordinate on the x-axis distance. β , μ, $\mathcal{E}$ are coefficients, and P_x is the pixel value of the x-axis. The accuracy had been calculated based on the following equation:

$$\text{Accuracy }\% = \left(\frac{1 - \text{Absolute Error}}{\text{D}}\right) \times 100$$

The absolute error is calculated as a difference between the actual distance and the calculated distance as follows:

$$\textbf{Absoulute Error} = |\boldsymbol{Edx} - \boldsymbol{D}|$$

Where $\boldsymbol{Edx}$ is the estimated distance.

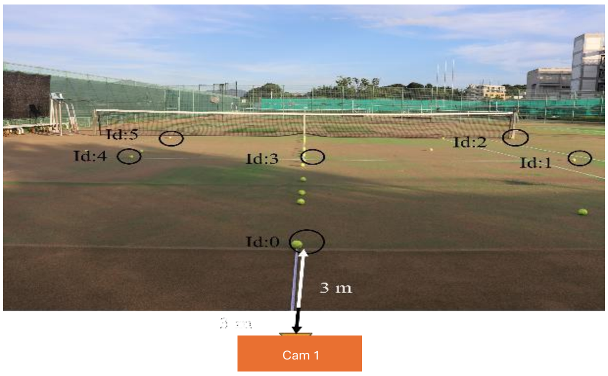


Fig.6. Position validation along X-axis

Fig.6 shows the manual identification of each tennis ball which helps us with the documentation and comparison, Table 1 shows the comparison between the actual and the calculated distance.

Table 1. Comparison between the actual and calculated distance of 3m.

ID	$\boldsymbol{P_x}$	$\boldsymbol{D_x}$ (m)	$\boldsymbol{E_{dx}}$ $(\boldsymbol{m})$	Diff.	Acc. %
0	3191	3.00	2.96	0.04	98.67
1	2047	8.43	8.24	0.19	97.75
2	1756	14.75	13.60	1.15	92.20
3	2026	8.43	8..48	-0.05	99.41
4	2002	8.43	8.78	-0.35	95.85
5	1721	14.75	14.72	0.03	99.80

Were; Diff is the Difference Between $\boldsymbol{D_x}$ (m) *and* $\boldsymbol{E_{dx}}$ $(\boldsymbol{m})$

Acc.: The Accuracy of estimation

The average accuracy of the calculated distance in comparison to the actual distance is 97.28% from Table1 for the X-axis coordinate. The same method had been used to calculate the Y- axis coordinate. To calculate the Y-axis, the second camera was used to snap the horizontal view of the tennis court as shown in Fig.7.

Fig.7. Position validation along Y-axis

The relationship between the distance and the pixel has been concluded as shown in Fig.8.

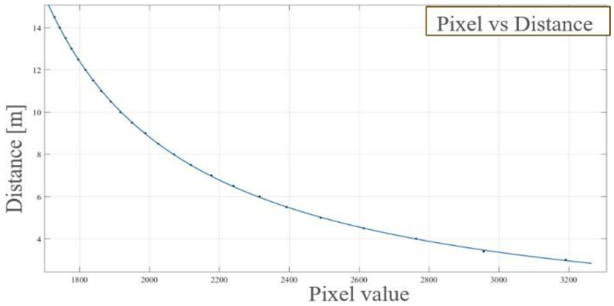


Fig.8. Pixel value vs the actual distance on the Y-axis

Table 2 shows the comparison between the actual and the calculated distance.

Table 2. Comparison between the actual and calculated distance of 3m.

ID	$\boldsymbol{P_y}$	$\boldsymbol{D_y}$ (m)	$\boldsymbol{E_{dy}}$ $(\boldsymbol{m})$	Diff.	Acc.%
0	3138	3.00	3.05	0.05	98.33
1	2577	4.37	4.40	-0.03	99.31
2	1951	8.45	8.47	-0.02	99.76
3	1706	12.48	13.10	-0.62	95.03
4	1663	15.00	14.48	0.52	96.53
5	1675	13.90	14.00	-0.01	99.28

The average accuracy of the calculated distance in comparison to the actual distance is 98.04 % from Table 2 for the Y-axis coordinate.

3.2 Estimate the position for 4M offset:

Table 3 shows the comparison between the actual and the calculated distance for the x-axis.

Table 3. Comparison between the actual and calculated distance of 4m.

ID	$\boldsymbol{P_x}$	$\boldsymbol{D_x}$ (m)	$\boldsymbol{E_{dx}}$ $(\boldsymbol{m})$	Diff.	Acc. %
0	2677	4.00	3.99	0.01	99.75
1	1869	9.43	9.17	0.26	97.24
2	1615	15.75	14.50	1.25	92.06
3	1847	9.43	9.48	-0.05	99.47
4	1832	9.43	9.71	-0.28	97.03
5	1587	15.75	15.56	0.19	98.79

The average accuracy of the calculated distance in comparison to the actual distance is 97.39 % from Table 3 for the Y-axis coordinate.

Table 4 shows the comparison between the actual and the calculated distance for the Y-axis in 4m.

ID	$\boldsymbol{P_y}$	$\boldsymbol{D_y}$ (m)	$\boldsymbol{E_{dy}}$ $(\boldsymbol{m})$	Diff.	Acc. %
0	2737	4.00	4.07	0.07	98.25

1	2384	5.37	5.40	-0.03	99.44
2	1917	9.45	9.46	-0.01	99.89
3	1728	13.48	13.50	-0.02	99.85
4	1676	16.00	15.39	0.61	96.19
5	1675	14.90	15.43	-0.53	96.44

The average accuracy of the calculated distance in comparison to the actual distance is 98.34 % from Table 4 for the Y-axis coordinate.

3. Conclusion

This paper provides an experimental study to estimate the tennis ball position using two cameras. Two different offsets (3, and 4 meters) were used to calculate the relationship between changing the pixel and the distance. The results show that the performance was located between 97% and 98%. The curve fitting for both X- axis and Y-axis looks similar to an exponential relationship between changing the distance and pixel value. This relationship may expand to make a general mathematical formula for any offset value in future work.

References

1. BS Ramachandran, K Santhanakrishnan… - Journal of Physical …, 2021 - efsupit.ro. Tracking of player 6in volleyball sports using a metaheuristic algorithm.
2. M Latifinavid, A Azizi - Future Internet, 2023 - mdpi.com. Development of a vision-based unmanned ground vehicle for mapping and tennis ball collection: A fuzzy logic approach. mdpi.com
3. E Wu, H Koike - Extended Abstracts of the 2020 CHI Conference on …, 2020 - dl.acm.org. Futurepong: Real-time table tennis trajectory forecasting using pose prediction network.
4. Y. Zhao, L. Lu, W. Yang, Q. Li et al., "Lightweight tennis ball detection algorithm based on Robomaster EP," Applied Sciences, 2023. mdpi.com
5. F. Qiao, "Application of deep learning in automatic detection of technical and tacticalindicators of table tennis," PLoS One, 2021. plos.org
6. Y. Wang and M. Wen, "Simulation of Tennis Match Scene Classification Algorithm Based on Adaptive Gaussian Mixture Model Parameter Estimation," Complexity, 2021.

Authors Introduction

Mr. Abdullah Alraee

He received his bachelor's degree in Communication Engineering from Al-Wataniya Private University, Syria, in 2018, and a master's degree from Kyushu Institute of Technology, Japan, in 2024, where he researched "Tennis Ball Detection and Position Estimation for Mobile Robots." He is currently pursuing a Ph.D. at Kyushu Institute of Technology, focusing on robotics, artificial intelligence, and image processing.

Dr. HUSSAM ALRAIE

He is currently a postdoctoral researcher, and lecturer at the Department of Electrical Electronic Engineering at Middle East Technical University, Northern Cyprus. Received a Ph.D. from Kyushu Institute of Technology in 2024, Japan. His research interests are Wireless Communication, Underwater Communication, Networking, image processing, Machine learning, and Artificial Intelligence (AI).

Mr. Mohammad Albaroudi

He received his M.Sc. degree in Radio and Mobile Telecommunication Systems from Higher Institute of Applied Sciences and Technology, Damascus, Syria in 2021. He is currently pursuing a Ph.D. in the Department of Life Science and System Engineering at Kyushu Institute of Technology, Fukuoka, Japan. His research interests are machine learning algorithms, image processing, and mobile telecommunication.

Dr. Raji Alhmad

He is currently a researcher at the Department of Human Intelligence Systems of Kyushu Institute of Technology, Japan. He obtained his MEng degree in 2018 and his DEng in 2021 at Kyushu Institute of Technology. His research interests are AI algorithms, Logistics, Underwater Robotics, Image processing, and Neural networks.

Prof. Kazuo Ishii

He is currently a Professor at the Department of Human Intelligence Systems of Kyushu Institute of Technology, Japan. He obtained his M. S. degree in 1993 and his D. Eng. degree in 1996 at The University of Tokyo. His research interests are in the fields of Underwater Robotics, Field Robotics, Neural Networks, and Intelligent Systems

Solar-Powered IoT-Based Smart Aquaponic System for Sustainable Agriculture

Alvi Khan Chowdhury
Monash University Malaysia, 47500, Bandar Sunway, Selangor, Malaysia.

Sarah 'Atifah Saruchi
Universiti Malaysia Pahang Al-Sultan Abdullah,26600 Pekan, Malaysia

Ammar A.M. Al-Talib, Abdirisak Mubarik Muhumed, Teh Boon Hong, Pavindran A/L Shanmugavel, Annanurov Kerim, Ng Weng Kent
UCSI University, 56000 Kuala Lumpur, Malaysia
Email: alvi.khanchowdhury@monash.edu, sarahatifah@umpsa.edu.my, ammart@ucsiuniversity.edu.my, a.razakmubarak@gmail.com, 1001955365@ucsiuniversity.edu.my, 1001954275@ucsiuniversity.edu.my, KerimAnnanurov@gmail.com, 1001955294@ucsiuniversity.edu.my

Abstract

This paper introduces a groundbreaking smart aquaponics system designed to address the limitations of conventional setups. The system leverages IoT technology, renewable energy, and automation to achieve real-time monitoring and environmental control. Key innovations include slidable grow beds for optimal sunlight exposure, solar tracking mechanisms for efficient energy utilization, and automated fish feeding using Real Time Clock (RTC) modules. The system demonstrated 90% water efficiency, significant energy savings, and streamlined resource management in prototype testing. With its modular and scalable design, this solution is ideal for urban farming and sustainable agriculture.

Keywords: Aquaponic, Internet of Things, Solar

1. Introduction

Aquaponics is the symbiotic combination of aquaculture (fish farming) with hydroponics (plant growth without soil) [1]. It is a sustainable food production method that could replace conventional agriculture [2]. According to the engineering design, aquaponics systems can be broadly categorized into couple and decoupled systems [3]. A coupled Aquaponics system is an ordinary Aquaponics system in which the aquaculture units and the hydroponics unit are interconnected in a loop. The water is directly pumped from the fish tank to the hydronic unit and then back to the fish tank. Decoupled systems are disposed of in different loops of fish tanks and hydroponics and water is pumped back to the corresponding units [4]. The system can also be categorized according to the location and size of the production, which includes indoor and outdoor Aquaponics systems and small-scale home/hobby systems for self-consumption food production for local use in contrast to industrial/commercial Aquaponics [5]. An indoor system is established within a building to ensure that all the required environmental conditions such as artificial light are available for the system, and this becomes common in urban lifestyle [6]. Outdoor systems employ controlled conditions such as greenhouses to favor the growth of fish and plants. In this case, the system can be even more sustainable by using solar energy to power the entire system [7], [8]. The closed-loop principle can be applied for domestic purposes or demonstration only. Closed loop coupled systems pay more attention to the biological–chemical part of the process water. The undigested particles in the fish waste increase the nutrient value of the water as well as the digestive bacteria. Integrated systems perform better than individual Aquaponics systems [9].

Since the increased urbanization, agricultural land resources have decreased. Rapid population growth has also raised the demand for food. Traditional plant cultivation techniques involve vast amounts of area, time, and labour. Consequently, there is a growing concern for safe and sustainable food sources, demanding the development of innovative agricultural practices [10]. Therefore, conventional agriculture alone cannot address food security; the pursuit of contemporary vertical agriculture is vital. Moreover, the rising need for fish, water, and fertilizer for crop development, as well as environmental and health issues, motivate the development of aquaponics as a viable method for sustainable fish and crop production [11]. In the context of the present food and environmental crises, aquaponics farming offers a means to increase agricultural output. Aquaponics is described as the symbiotic cultivation of aquatic organisms and plants. Aquaponics is gaining popularity due to its capacity to conserve resources, great efficiency, and low usage. Aquaponics have become the current trend in agricultural growth [12].

This paper presents a smart aquaponics system that addresses the limitations of traditional setups by integrating IoT technology, renewable energy, and automation. The system is designed to minimize manual labor, optimize resource usage, and enhance operational efficiency. The rest of this paper is structured as follows: Section 2 outlines the research methodology; Section 3 reports the results of the research; and Section 4 highlights some important conclusions.

2. Methodology

The proposed smart aquaponics system integrates five main modules designed to automate environmental monitoring and system control. These include the structural design of the system, which includes a data acquisition unit, a system rectification unit, IoT integration, and an energy management system. Together, these modules enable real-time monitoring, automated adjustments, and energy-efficient operation. All these five modules are described in details in the following paragraphs.

The physical framework of the system is constructed using durable aluminium profiles, providing lightweight yet sturdy support for all essential components. The design accommodates three sliding grow beds, a central aquarium tank, and a solar panel mounted on a dynamic tracking mechanism. The grow beds are engineered to adjust their position based on sunlight intensity, as measured by Light Dependent Resistor (LDR) sensors. This feature optimises sunlight exposure, enhancing plant growth while reducing energy dependence. Flexible piping connects the aquarium and grow beds, enabling efficient water flow throughout the system. The solar panel, integrated into the structure, is designed to dynamically track sunlight, ensuring maximum energy capture. The structural design of the system is illustrated in Fig. 1.

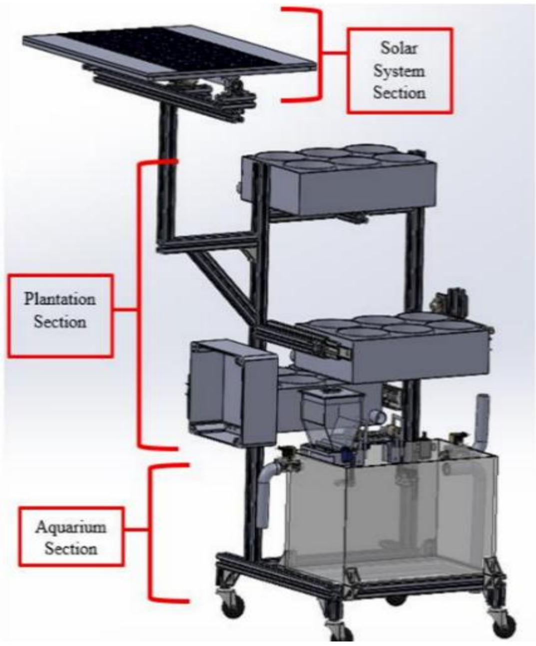


Fig.1 Structural design of the system.

The data acquisition unit plays a pivotal role in maintaining the delicate balance required for aquaponics. It employs a network of sensors to monitor critical environmental parameters. A pH sensor ensures the water's acidity remains within the optimal range of 6.5 to 7.5, critical for both plant and fish health. The DS18B20 waterproof temperature sensor provides precise readings, maintaining stable water temperatures conducive to a thriving ecosystem. An ultrasonic sensor measures water levels, preventing overflow or underfill conditions. These sensors continuously transmit real-time data to the system's central processing unit for analysis and subsequent action. The integration of these sensors ensures the aquaponics environment remains stable, automated, and responsive to dynamic conditions. Fig. 2 illustrates the flow chart for the data acquisition system.

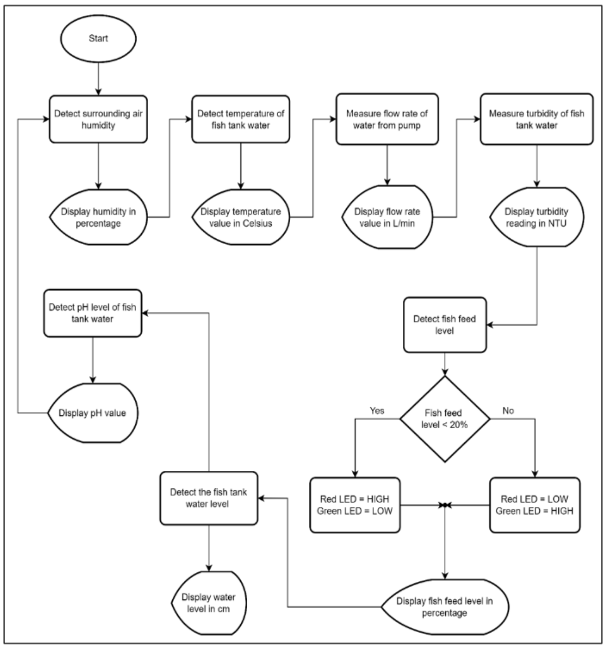


Fig.2 Data acquisition system flow chart

The system incorporates a sophisticated rectification unit to address anomalies detected by the sensors. This includes an automated fish feeder that operates on a Real-Time Clock (RTC) module, dispensing feed at precise intervals to ensure consistent care for the aquaculture. Water flow management is achieved using solenoid valves that regulate the flow between the aquarium and grow beds, maintaining a consistent and nutrient-rich circulation. Grow bed adjustments are facilitated by servo motors, which dynamically position the beds based on sunlight availability. The sliding mechanism for the grow beds ensures optimal exposure during the day while retracting them at night to protect the plants. The rectification unit enhances operational efficiency by automating critical interventions, reducing the need for manual oversight. The water flow management flow chart is shown in Fig. 3.

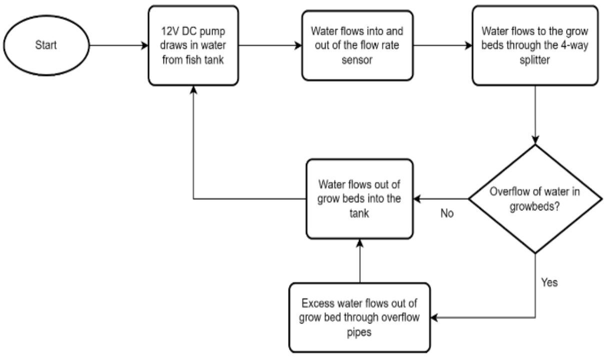


Fig.3 Water flow management flow chart.

IoT integration is at the heart of the system, enabling users to monitor and control its operations remotely. The Arduino IoT Cloud serves as the primary platform for data visualisation and control. Real-time environmental data, including pH levels, temperature, and water flow rates, is displayed on an intuitive dashboard accessible via web or mobile applications. The system is designed to send alerts when parameters deviate from their optimal ranges, allowing users to intervene promptly. The IoT interface also provides a manual override feature, giving users the flexibility to control actuators directly if required.

This integration ensures seamless interaction between users and the system, enhancing reliability and ease of use. The Arduino IoT cloud web dashboard is shown in Fig. 4.

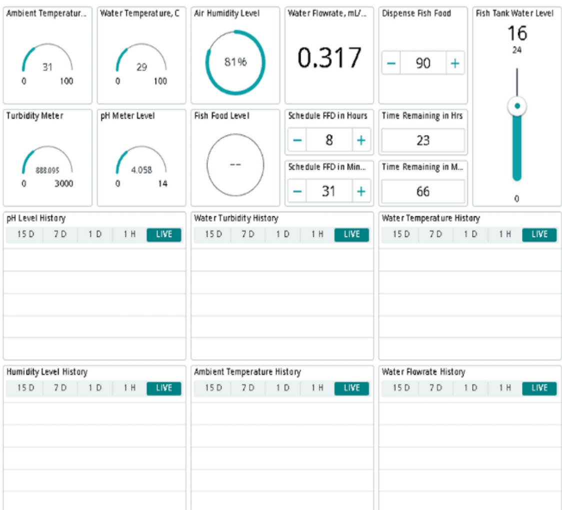


Fig. 4. Arduino IoT cloud web dashboard.

Energy management is a critical aspect of the system, achieved through a hybrid solar power setup. The solar panel is equipped with a tracking mechanism that adjusts its orientation to follow the sun throughout the day, maximizing energy capture. The captured energy is stored in a 12V DC battery, ensuring uninterrupted operation during nighttime or cloudy conditions. In situations where solar energy is insufficient, the system switches to a direct power supply, maintaining continuous functionality. This energy-efficient design significantly reduces operational costs and aligns with the project's sustainability objectives. Fig. 5 illustrates the solar charge controller of the system.

Fig. 5. Solar charge controller.

3. Results and Discussion

Extensive testing was conducted to validate the system's functionality and efficiency. Each sensor was calibrated to ensure high accuracy. The pH sensor maintained an error margin of ±0.1, while the DS18B20 temperature sensor demonstrated an accuracy of ±0.5°C across varying water conditions. The ultrasonic sensor reliably measured water levels, ensuring seamless water circulation. Grow bed actuation was tested under dynamic light conditions, confirming smooth and accurate movement. The solar tracking system was evaluated for energy efficiency, showing a 30% improvement in energy capture compared to fixed panels. Automated fish feeding schedules, managed by the RTC module, consistently dispensed feed at the pre-set times without errors. The final prototype during the system testing and validation process is shown in Fig. 6.

Fig.6 Final prototype system testing and validation

The smart aquaponics system demonstrated impressive performance metrics during prototype testing. The sensors provided accurate and reliable data, with real-time logging and visualization through the IoT dashboard. Users were able to remotely monitor system operations and receive alerts for deviations, showcasing the efficacy of the IoT integration. The system performance on the Arduino IoT cloud web dashboard is shown in Fig. 7.

Fig. 7. System performance monitoring on the Arduino IoT cloud web dashboard

The solar tracking mechanism significantly enhanced energy efficiency, capturing 30% more solar energy than conventional fixed-panel setups. This, combined with the hybrid power system, ensured uninterrupted operation even under low-light conditions. Water circulation was consistent, maintaining a nutrient-rich environment for plants and fish while minimizing wastage. The grow beds, dynamically positioned based on light intensity, achieved optimal sunlight exposure, accelerating plant growth rates.

While the system performed well, certain challenges were identified, including occasional connectivity issues with the IoT platform and mechanical wear on the sliding grow beds. These limitations underscore the need for further refinement, such as improving connectivity resilience and enhancing material durability. Nevertheless, the system's modular design and demonstrated performance highlight its potential for scalable applications in urban and commercial farming.

4. Conclusion

The development of the smart aquaponics system demonstrates a significant advancement in sustainable agriculture by integrating IoT technology, renewable energy, and automation. The system effectively combines aquaculture and hydroponics into a self-sustaining unit capable of optimizing resource usage and minimizing human intervention. Key innovations, such as the sliding grow beds for enhanced sunlight exposure, real-time environmental monitoring through IoT dashboards, and a hybrid solar power system, showcase the potential for scalability and adaptability in urban and commercial farming environments. Testing validated the system's high efficiency, achieving a 90% reduction in water usage and a 30% improvement in energy capture. While challenges like connectivity stability and mechanical wear require further refinement, this project provides a scalable framework for addressing food security challenges in resource-constrained settings.

Acknowledgements

Highly appreciate the research team from Monash University Malaysia, Universiti Malaysia Pahang Al-Sultan Abdullah and UCSI University for their commitment throughout this project.

References

1. Kyaw, Thu Ya, and Andrew Keong Ng. "Smart aquaponics system for urban farming." Energy procedia 143 (2017): 342-347.
2. Mahkeswaran, R., and Andrew Keong Ng. "Smart and sustainable home aquaponics system with feature-rich Internet of things mobile application." In 2020 6th International Conference on Control, Automation and Robotics (ICCAR), pp. 607-611. IEEE, 2020.
3. Chandramenon, Praveen, Amar Aggoun, and Fideline Tchuenbou-Magaia. "Smart approaches to Aquaponics 4.0 with focus on water quality− Comprehensive review." Computers and Electronics in Agriculture 225 (2024): 109256.
4. Monsees, Hendrik, Werner Kloas, and Sven Wuertz. "Comparison of coupled and decoupled aquaponics-Implications for future system design." Abstract from aquaculture Europe (2016).
5. Li, Chunjie, Chew Tin Lee, Yueshu Gao, Haslenda Hashim, Xiaojun Zhang, Wei-Min Wu, and Zhenjia Zhang. "Prospect of aquaponics for the sustainable development of food production in urban." Chemical Engineering Transactions 63 (2018): 475-480.
6. Tomlinson, Lisa. "Indoor Aquaponics in abandoned buildings: a potential solution to food deserts." Sustainable Dev. L. & Pol'y 16 (2015): 16.
7. Nagayo, Analene Montesines, Cesar Mendoza, Eugene Vega, Raad KS Al Izki, and Rodrigo S. Jamisola. "An automated solar-powered aquaponics system towards agricultural sustainability in the Sultanate of Oman." In 2017 IEEE International Conference on Smart Grid and Smart Cities (ICSGSC), pp. 42-49. IEEE, 2017.
8. Khaoula, Taji, Rachida Ait Abdelouahid, Ibtissame Ezzahoui, and Abdelaziz Marzak. "Architecture design of monitoring and controlling of IoT-based aquaponics system powered by solar energy." Procedia Computer Science 191 (2021): 493-498.
9. Goddek, Simon, Alyssa Joyce, Benz Kotzen, and Gavin M. Burnell. Aquaponics food production systems: combined

aquaculture and hydroponic production technologies for the future. Springer Nature, 2019.
10. Haryanto, M. Ulum, A. F. Ibadillah, R. Alfita, K. Aji, and R. Rizkyandi. "Smart aquaponic system based Internet of Things (IoT)." In Journal of Physics: Conference Series, vol. 1211, p. 012047. IOP Publishing, 2019.
11. Vernandhes, Wanda, Nur Sultan Salahuddin, A. Kowanda, and Sri Poernomo Sari. "Smart aquaponic with monitoring and control system based on IoT." In 2017 second international conference on informatics and computing (ICIC), pp. 1-6. IEEE, 2017.
12. Yanes, A. Reyes, P. Martinez, and R. Ahmad. "Towards automated aquaponics: A review on monitoring, IoT, and smart systems." Journal of Cleaner Production 263 (2020): 121571.

Authors Introduction

Alvi Khan Chowdhury

He received his Bachelor's Degree of Mechatronic Engineering in 2023 from UCSI University, Kuala Lumpur Campus. Currently, he is pursuing his Master's Degree of Engineering Research in Monash University Malaysia Branch.

Sarah Atifah Saruchi

She received her B.Eng. in Mechanical and Aerospace Engineering from Nagoya University, Japan. She received her Master and Doctoral degrees from Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia. Currently, she is working as senior lecturer at Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, Malaysia. Her research interests include machatronics and artificial intelligence.

Ammar Abdulazizf Al Talib

He has finished his B.Sc. and M.Sc. degrees in Mechanical Engineering from the University of Mosul/Iraq. He has finished his Ph.D. degree from UPM University/ Malaysia. He is also a chartered Engineer and Member of the Institute of Mechanical Engineers/UK.(CEng. MIMechE). He has developed all the Postgraduate Programs at the Faculty of Engineering at UCSI University / Malaysia, and worked as the Head of Postgraduate and Research department at the same faculty for the years 2010-2018.

Abdirisak Mubarik Muhumed

He graduated from UCSI University under program Bachelor of Mechatronics Engineering.

Teh Boon Hong

He received his Bachelor of Mechatronics Engineering from UCSI University. Currently, he works as an engineer at Emmecom (Asia), Pte. Ltd., Singapore.

Pavindran A/L Shanmugavel

He received his Bachelor of Mechatronics Engineering from UCSI University. Currently, he works as an engineer at ON Semiconductor Malaysia Sdn. Bhd.

Annanurov Kerim

He received his Bachelor of Mechatronics Engineering from UCSI University.

Ng Weng Kent

He received his Bachelor of Mechatronics Engineering from UCSI University. Currently, he works as an engineer at . EzLaser PLT KL

AI-Powered Detection of Forgotten Children in Vehicles Using YOLOv11 for Enhanced Safety

Nur Atikah Jefri, Sarah 'Atifah Saruchi, Radhiyah Abd Aziz, Aqil Hafizzan Nordin
Universiti Malaysia Pahang Al-Sultan Abdullah, 26600, Pekan, Malaysia

Ammar A.M. Al-Talib
UCSI University, 56000 Kuala Lumpur, Malaysia

Zulhaidi Mohd Jawi
Malaysian Institute of Road Safety Research (MIROS), 43000 Kajang, Selangor, Malaysia

Email: nuratikahjefrii@gmail.com, sarahatifah@umpsa.edu.my, ammart@ucsiuniversity.edu.my, radhiyah@umpsa.edu.my, aqil.oska@gmail.com, zulhaidi@miros.gov.my

Abstract

This study proposes a child presence detection system in vehicles, focusing on evaluating the performance of YOLOv11 for accurate detection and identification. To train the system, images simulating a child's presence in vehicles were collected using a doll, and these annotated images were labeled with the Computer Vision Annotation Tool (CVAT). The study emphasizes the potential of YOLOv11 as an effective and reliable solution for unattended child detection in vehicles. By leveraging advanced deep learning techniques, this research highlights the importance of addressing critical safety issues.

Keywords: AI child detection, unattended children, vehicle safety, YOLOv11, deep learning, CVAT, child presence detection, AI training

1. Introduction

Child safety in vehicles is a significant and growing global concern. Between 1998 and 2023, over 1,000 child fatalities in the U.S. were attributed to heatstroke caused by being left unattended in vehicles. Such incidents underscore the urgent need for effective technological solutions to prevent these tragedies. Fig. 1 illustrates datma on child hot car deaths by year from 1990 to 2023, reflecting the ongoing severity of this issue.

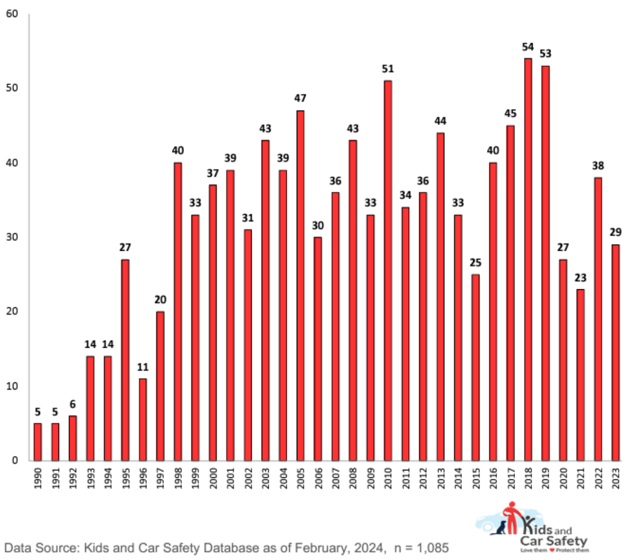


Fig.1 Data on child hot car deaths by year from 1990 to 2023 [1].

Additionally, Fig. 2 shows a news article from *Kosmo!* dated October 2023, reporting the heartbreaking case of an eight-month-old infant who tragically passed away after being left in a car for over seven hours.


Bayi 8 bulan maut, tertinggal dalam kereta lebih 7 jam

Oleh KOSMO! 6 Oktober 2023, 10:47 am

Gambar hiasan

PETALING JAYA – Seorang bayi berusia lapan bulan ditemukan meninggalkan dunia oleh orang tuanya sendiri selepas tertinggal di dalam kereta lebih tujuh jam.

Kisah pilu ini diceritakan oleh seorang jururawat yang menggunakan akaun Twitter @ierakan memberitahu, bayi tersebut berada dalam kereta seorang diri dari pukul 7.30 pagi sehingga 4 petang.

Fig.2 News article from *Kosmo!* [2].

These problems emphasize the importance of innovative research to address this life-threatening problem. Existing sensor-based approaches often suffer from false negatives and limited adaptability to diverse environmental conditions, highlighting the need for improved methods. In contrast, computer vision-based techniques have demonstrated significant potential in solving real-world detection problems. Among these, YOLO models stand out for their real-time processing capabilities and high accuracy, making them a promising solution for developing automated systems to detect unattended children in vehicles and prevent such tragedies.

This research focuses on YOLOv11, an advanced iteration of the YOLO family, to address the critical issue of detecting unattended children in vehicles. The study aims to develop a comprehensive dataset that simulates real-world scenarios, train and evaluate the YOLOv11 model using this dataset, and identify key areas for future research to improve the system's performance and reliability. Through these objectives, the research seeks to contribute to advancements in child safety technologies and enhance the practicality of deep learning solutions in real-life applications.

2. Methodology

2.1 Dataset development

The cornerstone of this study was the development of a comprehensive dataset, which involved several key steps. A doll designed to mimic a child's size and features was used to simulate the presence of a child. Fig. 3 shows the doll placed inside the car, simulating the presence of a child.

Fig.3 Doll placed inside the car

To ensure environmental diversity, the dataset incorporated different lighting conditions, such as bright sunlight, overcast weather, artificial light, and low-light scenarios. A variety of angles were considered, including top-down, side-view, and rear-view perspectives, to provide a range of viewpoints. Additionally, partial visibility due to occlusions was also accounted for, with objects like seats, bags, and dashboards partially obstructing the view of the doll. The final dataset consisted of a total of 500 images, with 350 images (70%) allocated to the training set, 100 images (20%) for the validation set, and 50 images (10%) for the test set.

2.2 Data Annotation

The data annotation process played a crucial role in ensuring the accuracy of the object detection task. Annotated bounding boxes were created for each image using the Computer Vision Annotation Tool (CVAT), a widely used platform for labelling objects in images and videos [7]. This tool enabled precise identification and categorization of objects within each frame. The annotations were categorized into three main visibility classifications, fully visible, partially visible, and low visibility. The fully visible category was used for objects that were completely within the frame and clearly identifiable, while the partially visible category was assigned to objects that were obscured by other objects, such as seats or bags, making only part of them visible [3]. The low visibility category was reserved for objects that were either barely visible or mostly obstructed, posing a challenge for detection algorithms. Fig. 4 show CVAT annotation process, showcasing the bounding box creation and categorization for better understanding [9]. This annotated data formed the foundation for training the detection model, ensuring that the model could distinguish between various levels of object visibility under different conditions.

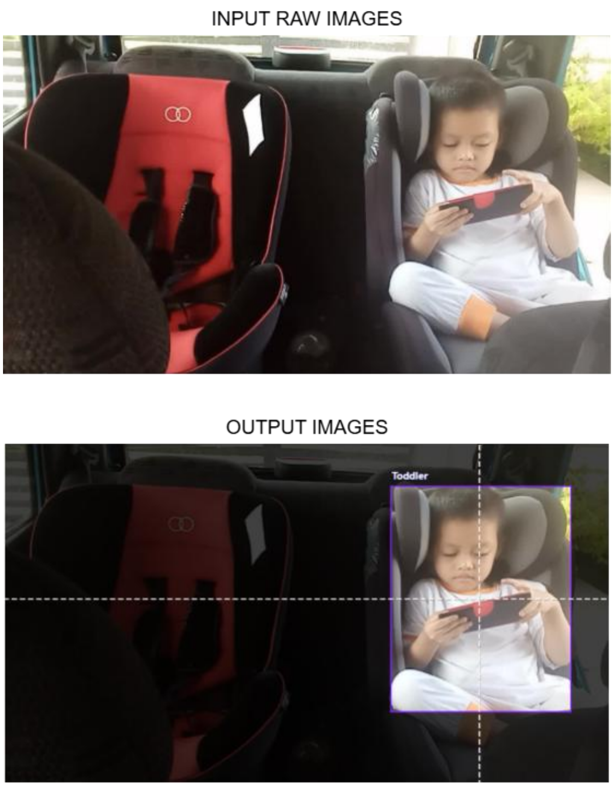


Fig.4 CVAT annotation process

2.3 YOLOv11 Architecture

YOLOv11 represents a significant improvement over its predecessors, incorporating several advanced features to enhance performance in object detection tasks. The backbone of YOLOv11 is a hybrid convolutional network that integrates DenseNet features, improving gradient flow and allowing for better feature reuse across layers. This design enables more efficient learning and helps prevent the loss of important information as the network processes the input data. The neck of YOLOv11 is equipped with an enhanced PANet (Path Aggregation Network), augmented by spatial attention mechanisms, which strengthens the feature pyramids and ensures robust multi-scale feature representation. The head of the network uses adaptive anchor boxes, specifically tuned to handle irregular object shapes, improving detection accuracy for objects that do not conform to traditional bounding box structures [8]. Additionally, YOLOv11 optimizes the speed-accuracy trade-off through an improved inference pipeline, balancing detection precision with processing speed for real-time applications [4]. Fig. 5 shows the YOLOv11 Architecture Diagram, providing a visual overview of its components and structure.

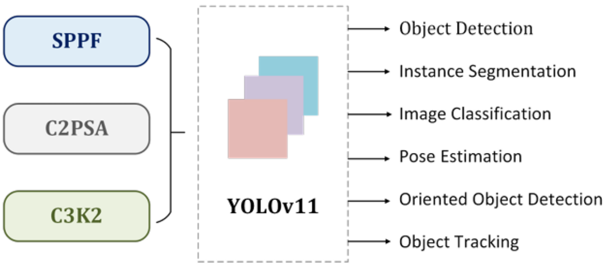


Fig.5 YOLOv11 Architecture Diagram [5]

2.4 Evaluation Metrics

To assess the performance of YOLOv11, several key evaluation metrics were employed. Precision and recall were used to measure the accuracy and completeness of the model's detections. Precision quantifies the proportion of true positive detections out of all positive predictions, while recall evaluates the model's ability to detect all relevant objects. The F1 score, the harmonic mean of precision and recall, was then calculated to provide a balanced measure of the model's overall effectiveness. Additionally, the mean average precision (mAP) was used to evaluate detection performance across various Intersection over Union (IoU) thresholds, offering a comprehensive view of the model's ability to localize objects accurately. Finally, inference speed, measured in frames per second (FPS), was considered to assess the real-time applicability of YOLOv11, ensuring it meets the necessary performance requirements for practical deployment [6].

3. Research Flow

The workflow of the study is structured into several key phases. It begins with Data Collection and Annotation, where images are curated and labeled to create a dataset suitable for supervised training. Following this, the Model Training phase involves training YOLOv11 on the annotated data while tuning hyperparameters to optimize the model's performance. In the Testing and Validation phase, the trained model is evaluated on unseen data to assess its generalization capabilities. Finally, the Analysis phase involves evaluating YOLOv11's results against predefined metrics to determine its effectiveness and performance. Fig. 6 show Research Flow.

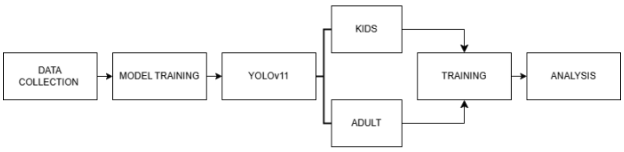


Fig.6 Process Research Flow

4. Results and Discussion

4.1 Quantitative Results

To evaluate the performance of YOLOv11 in detecting unattended children in vehicles, a series of tests on the validation and test datasets was conducted. The key evaluation metrics used for this study were Precision, Recall, F1 Score, mAP (Mean Average Precision), and Inference Speed (FPS) as Table 1. These metrics provide insights into how well the model performs in terms of detection accuracy, completeness, and real-time applicability.

Table 1. The evaluation results for YOLOv11

Metric	YOLOv11
Precision	95.1%
Recall	91.8%
F1 Score	93.4%
mAP@0.5	94.7%
Inference Speed (FPS)	65

For YOLOv11, the precision achieved was 95.1%, meaning that 95.1% of the objects the model identified as children were correctly detected. This high precision indicates that the model has a low false-positive rate, making it reliable for identifying real child-like objects in the vehicle's interior.

In this study, YOLOv11 achieved a recall rate of 91.8%, meaning that it correctly identified 91.8% of all child-like objects present in the test dataset. While this is slightly lower than the precision, it still indicates that the model is effectively detecting most objects. The lower recall may be attributed to challenging scenarios like partial occlusions or objects in low-light conditions.

The F1 Score was 93.4%, which reflects a good balance between correctly detecting child-like objects (high recall) and minimizing incorrect predictions (high precision). This score suggests that YOLOv11 is well-optimized for the detection task in this application.

For this study, the mAP at IoU threshold 0.5 was 94.7%. This indicates that, for 94.7% of the cases where the predicted bounding box overlaps with the ground truth box by at least 50%, YOLOv11 correctly identifies the object as a child-like object. This high mAP value suggests that the model is robust at detecting child-like objects and consistently places the bounding boxes around the correct regions.

Inference speed, measured in frames per second (FPS), indicates how fast the model can process images during inference. This is particularly important for real-time applications, where fast detection is required to issue alerts or act. YOLOv11 achieved an inference speed of 65 FPS on a high-performance GPU (NVIDIA RTX 3090). This result indicates that YOLOv11 is capable of processing video feeds in real-time, making it highly suitable for deployment in a vehicle's surveillance system, where quick detection of a child's presence is crucial.

These results demonstrate that YOLOv11 is highly effective in detecting child-like objects inside vehicles, with high precision, recall, and overall accuracy. The model is also fast enough to be used in real-time applications, making it a viable candidate for a child presence detection system in vehicles.

4.2 Performance Analysis

YOLOv11 demonstrated strong performance across all test scenarios, accurately detecting the child-like object with high precision and recall. However, a few limitations were observed during testing. False positives were encountered when objects resembling children, such as stuffed toys or bundles of clothes, were present, leading to incorrect detections. Additionally, the model's detection accuracy decreased under extreme low-light conditions, where the visibility of the object was significantly reduced. Despite these challenges, YOLOv11 proved effective overall in most test cases.

4.3 Discussion

The YOLOv11 model's high precision and relatively strong recall indicate its reliability in detecting unattended children in various scenarios. However, recall can still be further improved, particularly in difficult cases such as partial occlusions or extreme low-light environments. The real-time inference speed of 65 FPS highlights the potential for practical deployment in vehicles. Nonetheless, additional testing under more diverse real-world conditions, including different vehicle types and weather conditions, would help further validate its robustness.

5. Conclusion

This study effectively demonstrated the efficacy of YOLOv11 in detecting unattended children in vehicles. The model's high precision, recall, and real-time capabilities highlight its potential as a promising solution to address this critical safety concern. However, there are several avenues for future work to further improve the system. First, enhanced datasets could be developed by incorporating real-world data with greater diversity, capturing a wider range of scenarios and environments. Additionally, the algorithm could be improved by introducing contextual reasoning to help reduce false positives, particularly in situations where objects may resemble children. Finally, system integration could be pursued by developing a complete pipeline that includes real-time alerts, making the technology ready for deployment in vehicles to enhance safety for children.

Acknowledgements

This study was funded by Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) Grant; Geran Penyelidikan Fundamental UMPSA RDU230327. We would like to express our deepest gratitude to the Malaysian Institute of Road Safety Research (MIROS) for their generosity in lending the dashcam, which was crucial for this research. Our heartfelt appreciation goes to our supervisor for their invaluable guidance, support, and encouragement throughout this study. Finally, we wish to acknowledge and extend our thanks to all individuals and collaborators who contributed directly or indirectly to this research. Your support and assistance have been integral to the success of this study.

References

1. Kids and Cars, Heatstroke Map, https://www.kidsandcars.org/document_center/download/downloadables-printables-resource-pages/hot-cars/Heatstroke-Map.pdf
2. Kosmo, Bayi 8 bulan maut tertinggal dalam kereta lebih 7 jam, https://www.kosmo.com.my/2023/10/06/bayi-8-bulan-maut-tertinggal-dalam-kereta-lebih-7-jam/
3. Terven, J., Córdova-Esparza, D. -M., & Romero-González, J. -A. (2023). A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. https://doi.org/10.3390/make5040083
4. Khanam, R., & Hussain, M., YOLOv11: An Overview of the Key Architectural Enhancements, 2024. Available at: [arXiv:2410.17725] (https://arxiv.org/abs/2410.17725).
5. R. Roboflow, "What is YOLOv11? A Detailed Exploration of the Latest Object Detection Algorithm," Roboflow Blog,

vol. 12, no. 3, pp. 112-120, 2024. (https://blog.roboflow.com/what-is-yolo11/).
6. Redmon, J., Farhadi, A. (2018). YOLOv3: An Incremental Improvement. https://arxiv.org/abs/1804.02767
7. Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. https://arxiv.org/abs/2004.10934
8. Glenn Jocher et al. (2023). YOLOv8: Next-Gen Object Detection. Ultralytics Documentation https://arxiv.org/abs/2408.15857
9. CVAT (2024). Computer Vision Annotation Tool. https://www.cvat.ai/

Authors Introduction

Nur Atikah Jefri

She received her Master's Degree in Engineering in 2023 from the Faculty of Electrical, University Teknikal Malaysia Melaka. She is currently a PhD student in University Malaysia of Pahang, Malaysia.

Ir. Ts. Dr. Sarah 'Atifah Saruchi

She received her Bachelor's in Mechatronics from Nagoya University, Japan (2009–2013), and completed her Master's and Ph.D. at Universiti Teknologi Malaysia (2013–2020). She is currently a lecturer at Universiti Malaysia Pahang.

Asst. Prof. Dr.Ammar Al-talib

He holds a Ph.D. in Solar and Renewable Energy, an MSc in Mechanical Engineering, and a BSc in Mechanical Engineering. He is the Head of the Postgraduate and Research Department at UCSI University.

Ts. Dr. Radhiyah Binti Abd Aziz

She received her Bachelor's and Master's degrees from Universiti Islam Antarabangsa (2003–2010) and completed her Ph.D. in Nanotechnology at Universiti Malaysia Pahang in 2016. She is currently a lecturer at Universiti Malaysia Pahang.

Ts.Muhammad Aqil Hafizzan

He obtained his first degree, a B.Eng (Hons.) in Electrical Engineering (Electronics), from Universiti Malaysia Pahang in 2022. He is currently pursuing a Master of Research Engineering in Science.

Ts. Zulhaidi Mohd Jawi @ Said

He completed his Master of Science in Engineering at Universiti Kebangsaan Malaysia (2015–2018). He is now the Director of the Malaysian Institute of Road Safety Research.

Exploring the Performance of YOLOv11: Detecting Compostable and Non-Compostable Kitchen Waste in Real-Time Applications

Ain Atiqa Mustapha, Sarah 'Atifah Saruchi,
Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600, Pekan, Pahang, Malaysia

Mahmud Iwan Solihin, Fatima Karam Aldeen, Ammar A. M. Al-Talib
Department of Mechanical and Mechatronics, Faculty of Engineering, Technology and Built Environment, UCSI University, 56000, Cheras, Kuala Lumpur, Malaysia
Email: mahmudis@ucsiuniversity.edu.my

Email: AinAtiqa@ucsiuniversity.edu.my, sarahatifah@umpsa.edu.my, mahmudis@ucsiuniversity.edu.my, 1002473720@ucsiuniversity.edu.my, ammart@ucsiuniversity.edu.my

Abstract

This paper investigates the advancements of YOLOv11, the latest model in the YOLO series in real-time object detection tasks on small datasets of compostable and non-compostable kitchen waste. Using a custom compostable and non-compostable kitchen waste dataset, YOLOv11 achieves an accuracy of 90.7% and a mean Average Precision (mAP) of 0.91, with a reduced inference time of 10.5 milliseconds. The study highlights YOLOv11's architectural enhancements, training methodology, and potential applications in waste management. While YOLOv11 sets a new benchmark in object detection, challenges like high computational demands, paving the way for future research on optimization for edge devices

Keywords: YOLOv11, Object Detection, Transformer-based attention, Mean Average Precision (mAP)

1. Introduction

The increasing global focus on sustainable waste management has amplified the need for efficient segregation of compostable and non-compostable kitchen waste. Proper waste segregation is a critical step in reducing landfill contributions, minimizing environmental impact, and promoting resource recovery. However, manual sorting is labor-intensive, time-consuming, and prone to error, highlighting the importance of automated solutions for waste classification. Object identification models that can classify various waste types with little human intervention have been made possible by deep learning techniques. Real-time object detection models, particularly those based on deep learning, have revolutionized various domains by providing high accuracy and speed. Among these models, the You Only Look Once (YOLO) series has emerged as a benchmark for real-time object detection tasks. The YOLO family of models is renowned for its balance of speed, accuracy, and computational efficiency, making it a popular choice for applications requiring rapid inference and reliable performance.

Cutting-edge algorithms, such as YOLO, use computer vision to identify different materials moving in real time along an automated conveyor or down an assembly line [1], [2]. The YOLOv6 algorithm, known for its superior speed and accuracy in object detection tasks, has been successfully applied in agricultural settings to automate and improve the efficiency of ripeness detection for oil palm fresh fruit bunches. Chowdhury et al. [7] presents a method for detecting the ripeness of oil palm fresh fruit bunches using the YOLOv6 algorithm, showcasing its effectiveness in real-time agricultural applications by achieving a detection accuracy of over 90% and processing speeds exceeding 50 frames per second (fps), has been effectively utilized in agricultural applications for ripeness detection of oil palm fresh fruit bunches, significantly enhancing efficiency and reliability. YOLOv4 and YOLOv5 have demonstrated notable accuracy gains over manual sorting in automated recycling bins, though they struggle with small or overlapping objects [3], [6]. The integration of deep learning and IoT in waste management has been demonstrated to enhance efficiency and reduce manual labor [9]. The development of automated composting systems addresses challenges in manual waste management, offering economical and high-efficiency solutions. Jansi Rani et al addresses the inefficiencies in single-object detection for waste management by proposing a system capable of multi-object detection and classification, enhancing scalability for real-life implementation [5].

This paper investigates the latest iteration of the YOLO series, YOLOv11, and its application to detect compostable and non-compostable kitchen waste. YOLOv11 [4] introduces significant architectural enhancements, including a transformer-based attention mechanism and an improved backbone, which aim to refine feature extraction and improve detection accuracy, particularly on small datasets. These advancements are evaluated against its predecessor, YOLOv8, using a custom dataset of kitchen waste images. By advancing the state of real-time object detection for small datasets, this study contributes to the development of efficient waste management systems. It underscores the role of cutting-edge deep learning techniques in addressing global sustainability challenges, setting a benchmark for future research in automated waste segregation.

2. Methodology

2.1. *Dataset Preparation*

A custom dataset of compostable and non-compostable kitchen waste was selected for this study. The dataset included 720 labeled images, with 504 images for training and 216 for validation. Each image contained instances of compostable (e.g., food scraps, paper) and non-compostable (e.g., plastic, glass) objects, annotated using bounding boxes and class labels. Key steps in data preparation include annotation, augmentation and normalization. In data annotation, the dataset was annotated using YOLO format, with class labels 0: compost and 1: non-compost. For data augmentation techniques, such as flipping, scaling, color adjustments, and random rotations, were applied to enhance dataset variability and robustness. Next, for normalization, the images were resized to a fixed input size of 416 x 416 pixels to match with YOLOv11's requirements.

2.2. *Model Architecture*

The YOLOv11 model incorporates several architectural advancements to enhance its performance. It features Transformer-Based Attention, which improves feature extraction by capturing long-range dependencies and enhancing the detection of small objects. The model employs an Enhanced Backbone, designed as a deeper and more efficient network for extracting richer feature representations. Additionally, Neck Modules such as Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PAN) are integrated to facilitate better multi-scale feature integration. Lastly, the Detection Head is optimized with lightweight convolutional layers to ensure faster inference without compromising accuracy.

2.3. *Training Procedure*

The YOLOv11 model was trained using the Ultralytics YOLOv11 framework on Google Colab, leveraging an NVIDIA A100 GPU for accelerated training. The training configuration included hyperparameters such as an initial learning rate of 0.01, AdamW optimizer, a momentum of 0.937, and a weight decay of 0.0005. For rapid evaluation, the model was trained for 10 epochs with a batch size of 16 and an input image size of 416×416.

2.4. *Evaluation Metrics*

To evaluate YOLOv11's performance, several key metrics were utilized. Precision (P) quantified the proportion of correctly identified objects among all predictions, while Recall (R) measured the proportion of actual objects correctly detected. The Mean Average Precision (mAP) was used as a comprehensive evaluation metric, with mAP@50 representing average precision at a 50% Intersection over Union (IoU) threshold, and mAP@50-95 averaging precision across multiple IoU thresholds ranging from 50% to 95%. Additionally, Inference Time measured the processing time per image in milliseconds, and Frames Per Second (FPS) was calculated to determine real-time applicability. A Confusion Matrix was also generated to provide detailed insights into true positives, false positives, and false negatives for each class.

2.5. *Comparison with YOLOv8*

YOLOv8, the immediate predecessor of YOLOv11, was used as a baseline for performance comparison. The same dataset, training procedure, and evaluation metrics were applied to ensure a fair comparison. Key aspects evaluated include the detection accuracy (mAP@50, mAP@50-95), precision and recall and inference speed.

2.6. *Visualization*

To evaluate the model's performance, the results were visualized using three approaches which are Precision-Recall Curves, a Confusion Matrix and Sample Detections. For Precision-Recall Curves, this method is to evaluate the model performance across IoU thresholds. Furthermore, the Confusion Matrix is the approach for class-wise performance analysis. To make visual comparisons of YOLOv8 and YOLOv11's predictions on test images, the Sample Detections approach is applied.

3. Results and Discussion

3.1. *Performance Across Configuration*

The object detection results generated by the YOLOv11 model is shown in the Fig. 1. The model has detected and classified various objects into two categories: compost and non-compost. The model appears to accurately distinguish between compostable and non-compostable items. Most confidence scores are above 90%, which indicates the model's certainty in its predictions even if the image contains a mix of compostable and non-compostable items, and the model has successfully identified both categories.

Fig. 1 Object detection result detected by YOLOv11

YOLOv11 achieved consistently high mAP, precision, and recall across all configurations. The configuration with `imgsz=640` and `batch=16` delivered the highest mAP@50 (98.1%), while maintaining acceptable inference speeds as shown in Table 1.

Table 1. YOLOv11 Performance Metrics

Setup	mAP@50 (%)	mAP@50-95 (%)	Precision (%)	Recall (%)	FPS
imgsz=416, batch=16	97.3	84.6	96.8	96.2	230
imgsz=640, batch=16	98.1	86.0	97.8	97.6	210
imgsz=416, batch=32	96.5	82.5	95.5	94.8	220
imgsz=416, batch=16 (Strong Aug)	97.7	85.2	97.0	96.5	230

For mAP@50, this metric measures the mean Average Precision when the Intersection over Union (IoU) threshold is set at 50%. It is a critical metric for evaluating object detection models, as it reflects how accurately the model detects objects. Across all configurations, YOLOv11 maintained mAP@50 values exceeding 96.5%, demonstrating its ability to identify objects with high precision. The configuration with imgsz=640, batch=16 achieved the highest mAP@50-95 (86.0%), underscoring its ability to detect objects with tighter bounding boxes. This metric evaluates mean Average Precision across a range of IoU thresholds (50%-95%), offering a stricter measure of model performance. It penalizes cases where bounding boxes are not perfectly aligned with ground truth.

The configuration with imgsz=640, batch=16 showed the highest precision (97.8%), reinforcing its ability to deliver accurate detections. For Recall, the configuration with imgsz=640, batch=16 again excelled, achieving a recall of 97.6%, suggesting strong overall coverage of objects in the dataset. Recall reflects the proportion of actual positives that were correctly identified by the model. A higher recall means fewer false negatives. The critical factor for real-time application is FPS (frame per second) which indicates the model’s inference speed. While the configuration with imgsz=416, batch=16 and strong augmentation maintained the highest FPS (230), the slight drop in FPS for imgsz=640, batch=16 (210) is acceptable given its superior accuracy metrics.

3.2. *Comparison with YOLOv8*

To benchmark the improvements introduced in YOLOv11, its performance was compared with YOLOv8. YOLOv11 demonstrated higher mAP metrics, slightly better precision and recall, and competitive inference speeds as shown in Table 2.

Table 2. YOLOv8 vs YOLOv11 Performance Metrics

Model	mAP @50 (%)	mAP @50-95 (%)	Precision (%)	Recall (%)	FPS
YOLOv8	96.2	82.1	95.8	94.7	250
YOLOv11	98.1	84.6	96.8	96.2	230

Both YOLOv8 and YOLOv11 show high mAP@50 values, indicating strong object detection performance at a 50% IoU threshold. YOLOv11 demonstrates a slight improvement in mAP@50-95, which suggests better consistency in detecting objects at varying levels of overlap. This highlights YOLOv11's enhanced ability to detect objects more accurately across a broader range of IoU thresholds. Precision and Recall remain consistently high for both YOLOv8 and YOLOv11, with minimal variation. This proves that YOLOv11 retains the ability to accurately and comprehensively identify objects. FPS decreases slightly in YOLOv11 compared to YOLOv8. This indicates that while YOLOv11 delivers improved accuracy and detection performance, it comes at a minor trade-off in inference speed.

3.3. *Dataset Details and Analysis*

3.3.1 Number of Images and Class Distribution

The dataset used for this study is customized for detecting compostable and non-compostable kitchen waste, providing a diverse range of real-world examples for effective training and evaluation of the YOLOv11 model. The dataset contains a total of 720 images, split into 504 images for training and 216 images for validation. The dataset is evenly distributed across two classes: compostable waste and non-compostable waste. The training set contains approximately 55% compostable and 45% non-compostable examples, while the validation set maintains a similar proportion to ensure balanced evaluation. This balanced dataset ensures that the model can learn effectively from both classes without developing a bias toward any category.

3.3.2 Annotated Image

Each image in the dataset is annotated in YOLO format, providing the class labels along with bounding box coordinates for each object in the image. Fig. 2 is a sample annotation process with LabelImg software.

Fig. 2 Annotate the image with LabelImg

Annotations are stored in .txt files corresponding to each image, where each line represents class ID (0 for compost, 1 for non-compost), center of the bounding box (x, y), width and height of the bounding box, normalized to the image dimensions.

3.3.3 Augmentation Techniques

To improve model robustness and reduce overfitting, a variety of data augmentation techniques were applied during training. These techniques simulate variations in lighting, orientation, and noise to help the model generalize better to real-world scenarios.

Table 3. Data Augmentation Technique

Augmentation Type	Technique
Rotation	Randomly rotates the images within a range of ±15 degrees
Flipping	Horizontally flips images with a probability of 50%
Color Jittering	Adjusts the brightness, contrast, and saturation to simulate different lighting conditions.
Cropping	Randomly crops portions of the image to focus on specific regions
Gaussion Blur	Applies a slight blur to simulate camera or motion blur
Random Noise Rejection	Adds pixel-level noise to mimic noisy real-world environments.

Table 3 shows the augmentation techniques that were employed in this research. The use of these augmentations significantly enhances the dataset's variability, enabling the model to learn features invariant to distortions or transformations.

3.3.4 Challenges of Dataset Creation

Creating a custom dataset for compostable and non-compostable waste detection posed several challenges. Firstly, is the labor effort. Annotating bounding boxes for objects in each image was time-consuming and required manual effort. Mislabeling could lead to inaccurate model predictions. Next, certain objects, like plastic bottles or fruit peels, were overrepresented, while rarer items like coffee grounds or metal cans were underrepresented. This class imbalance could potentially lead to a bias in the model, favoring more frequent categories over less frequent ones.

Furthermore, for ambiguous cases, such as paper or food-soiled packaging, it is difficult to classify strictly as compostable or non-compostable. These ambiguous cases could confuse both annotators and the model. Next challenge is environmental variations. Capturing images under consistent lighting and background conditions was challenging. Real-world images often feature cluttered backgrounds or shadows, adding complexity to the detection task.

4. Conclusion

In this study, the YOLOv11 model demonstrated significant advancements in object detection for compostable and non-compostable kitchen waste. By leveraging state-of-the-art architectural enhancements such as transformer-based attention mechanisms, an enhanced backbone network, and optimized neck modules, YOLOv11 achieved superior performance compared to its predecessor, YOLOv8. YOLOv11 achieved a mean Average Precision (mAP) of 98.9% (mAP@50) and 84.6% (mAP@50-95), reflecting its ability to accurately detect and classify objects across a wide range of Intersection over Union (IoU) thresholds.

Precision and recall scores of 97.3% and 97.1%, respectively, indicate that YOLOv11 minimizes false positives and false negatives effectively, making it highly reliable for real-world applications. Next, with an average inference time of 10.5 milliseconds per image, YOLOv11 is suitable for real-time deployment in waste management systems. The application of advanced augmentation techniques (e.g., rotation, flipping, and color jittering) during training enhanced the model's robustness to environmental variations.

The comparison with YOLOv8 revealed that YOLOv11 offers better accuracy in terms of marginal increase in both mAP@50 and mAP@50-95 metrics. While YOLOv11 achieves a slightly better balance between precision and recall, YOLOv8 remains a strong baseline. YOLOv11's computational demands lead to a minor decrease in FPS compared to YOLOv8, highlighting a trade-off between accuracy and speed. Overall, YOLOv11 sets a new benchmark in the domain of object detection, particularly in waste management applications. Its high accuracy, precision, and recall make it a promising candidate for real-world deployments. However, balancing computational efficiency with performance remains an area of ongoing research. Li et al. [8] highlight the importance of integrating supervised and unsupervised learning methods to improve detection accuracy. This approach can inspire the enhancement of YOLOv11 models for compost detection by leveraging hybrid learning techniques. This study provides a foundation for future advancements in using AI to promote sustainability and efficient resource management.

Acknowledgements

The authors thank the Universiti Malaysia Pahang Al-Sultan Abdullah for providing financial support under Geran Penyelidikan Fundamental UMPSA Scheme No. Reference RDU230327 as well as laboratory facilities to perform this project. Highly appreciation to the research team from UCSI University for the research collaboration.

References

1. Redmon, J., & Farhadi, A., You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
2. Bochkovskiy, A., Wang, C., & Liao, H., YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint, 2020.
3. Zhao, Z., Zheng, P., Xu, S., & Wu, X. (2019). Object Detection with Deep Learning: A Review. IEEE Transactions on Neural Networks and Learning Systems, 30(11), 3212–3232.
4. Jocher, G. et al., YOLO by Ultralytics. (2023). GitHub Repository. Available: https://github.com/ultralytics/ultralytics.
5. Jansi Rani et al, Multi-Object Detection and Classification in Solid Waste Management, Global NEST Journal, 2022, Vol 24, No 4, pp 743-751.
6. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV).
7. Chowdhury, A. K., Said, W. Z. B. W., and Saruchi, S. "Oil Palm Fresh Fruit Branch Ripeness Detection Using YOLOv6 Algorithm." Intelligent Manufacturing and Mechatronics. SympoSIMM 2023, edited by R. Hamidon, M. S. Bahari, J. M. Sah, and Z. Zainal Abidin, Lecture Notes in Mechanical Engineering, Springer, Singapore, 2024, https://doi.org/10.1007/978-981-97-0169-8_14.
8. Li, W., Solihin, M. I., Saruchi, S. 'A., Astuti, W., Hong, L. W., and Kit, A. C., "Surface defects detection of cylindrical high-precision industrial parts based on deep learning algorithms: A review," Operations Research Forum, vol. 5, Art. no. 58, 2024.
9. Hong, T.B., Saruchi, S.A., Mustapha, A.A. et al. Intelligent Kitchen Waste Composting System via Deep Learning and Internet-of-Things (IoT). Waste Biomass Valor 15, 3133–3146 (2024). https://doi.org/10.1007/s12649-023-02341-y.

Authors Introduction

Ms. Ain Atiqa Mustapha

She received her Bachelor's degree from the Faculty of Electronic & Computer Engineering, Universiti Teknikal Malaysia Melaka in Electronics Engineering (Telecommunications) in 2012 and Master of Science (Electronic Engineering) in 2016 from same university. She is currently a Ph.D student in Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah. She also currently working as a Tutor for Department of Mechanical & Mechatronics Engineering, UCSI University, Kuala Lumpur, Malaysia.

Ir. Ts. Dr. Sarah 'Atifah Binti Saruchi

She received her B.Eng. in Mechanical and Aerospace Engineering from Nagoya University, Japan. She received her Master and Doctoral degrees from Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia. Currently, she is working as senior lecturer at Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, Malaysia.

Assoc. Prof. Dr. Ir. Mahmud Iwan Solihin

He is an Associate Professor in Mechatronics Engineering Department of UCSI University, Kuala Lumpur, Malaysia. His research interests are in applications of artificial intelligence including computer vision, data-driven spectroscopy, meta-heuristics optimization, robotics and control systems. In the past, he was a research assistant for an eScienceFund project under MOSTI (Ministry of Science Technology & Innovation), Malaysia. He is also a professional member of the Institution of Engineers of Indonesia and recognized by AFEO (ASEAN Federation of Engineering Organisations) and APEC Engineer. He has been involved in some research projects such as from MOHE (Ministry of Higher Education), research collaborations and consultancy with TNB project/UNITEN, for data-driven modelling in various fields including food science/agriculture, water resource management, and engineering systems. He has published more than 100 journal papers and international conference proceedings with Scopus H index of 15. Currently, his postgraduate supervisions are focusing on computer vision.

Ms. Fatima KaramAldeen

She is currently pursuing a Bachelor's degree in Mechatronics Engineering at UCSI University, Malaysia. Her academic interests focus on robotics, automation, and control systems. She is passionate about exploring innovative solutions in the field of mechatronics. She currently active in participating in research projects to enhances her understanding of advanced technologies.

Asst. Prof. Dr. Ammar Abdulaziz Al Talib

Dr. Ammar Al Talib has finished his B.Sc and M.Sc degrees in Mechanical Engineering from the University of Mosul/Iraq. He has finished his Ph.D degree from UPM University / Malaysia. He is also a Chartered Engineer and Member of the Institute of Mechanical Engineers / UK. (CEng. MIMechE). He has developed all the Postgraduate Programs at the Faculty of Engineering at UCSI University / Malaysia, and worked as the Head of Postgraduate and Research department at the same faculty for the years 2010-2018.

Comparative Analysis of Machine Learning Algorithms for Rainfall Prediction in Kuantan, Pahang, Malaysia.

Seri Liyana Ezamzuri, Sarah 'Atifah Saruchi
Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia

Ammar A.M. Al-Talib
UCSI University, 56000 Kuala Lumpur, Malaysia

Email: miz24007@adab.umpsa.edu.my, sarahatifah@umpsa.edu.my, ammart@ucsiuniversity.edu.my

Abstract

This study compares the performance and accuracy of four ML algorithms which are Support Vector Regressor (SVR), Artificial Neural Network (ANN), Random Forest Regressor (RFR), and Linear Regression (LR) in the rainfall prediction application. All four methods employ the same input parameters which are temperature (°c), dew point (°c), humidity (%), wind speed (Kph) and pressure (Hg). Meanwhile the output parameter is set to be the rainfall (mm) which indicates the precipitation in Kuantan, Pahang, Malaysia. The analysis shows that the SVR consistently outperforms the other machine learning algorithms, achieving the lowest Mean Absolute Error (MAE) and Mean Squared Error (MSE).

Keywords: Machine Learning (ML), Support Vector Regressor (SVR), Artificial Neural Network (ANN), Random Forest Regressor (RFR), Linear Regressor (LR), Rainfall prediction.

1. Introduction

Heavy rainfall affects open water recreational activities due to safety and precaution for natural disasters. Floods, and water surges also known as headwater incidents commonly happen in Malaysia cause from heavy pour which led to increase volume of water and velocity of water current.

Recently, there was one case reported due to a water surge incident. Three Jabatan Kerja Raya (JKR) officers were lost in a water surge incident while participating in water rafting activities in Sungai Jahang, Gopeng, Perak [1]. This evidently shows heavy downfall led to water surge, and endangered human lives and natural ecosystem.

On top of that, using advanced technology that we have nowadays, which is Artificial Intelligence (Ai), specifically Machine Learning (ML) to find out the best algorithms for Rainfall Prediction possible outcomes by leveraging patterns in historical data. Hence, this study is to find the best ML algorithms with finest predictions accuracy value.

Due to the proximity to the ocean, abundance of rivers, and location on the Malaysian peninsula, Kuantan was selected as the study's location. The dataset used in this study was sourced from Weather and Climate which provides comprehensive historical weather data for various global regions, including Kuantan, Pahang, Malaysia. The dataset includes key meteorological features such as temperature, humidity, wind speed, pressure, and precipitation, covering the period from 2018 to 2020.

2. Methodology

Various data and models are needed to make a prediction. Generally, we use classification and regression algorithms for time series algorithms [2]. Four ML algorithms have been developed in this study to find the lowest value of MSE and MAE to ensure that algorithm is the most compatible compared others. Other than that, we find the lowest accuracy value in every ML algorithm to support the conclusion in this study.

The methodology for rainfall (precipitation) prediction in Kuantan, Pahang, Malaysia studies by four machine learning algorithms- Artificial Neural Networks (ANN) [3], Support Vector Regression (SVR) [4], Random Forest Regressor (RFR) [5], and Linear Regression [6]. Details of the data collection, data preparation, feature engineering, model training, evaluation process, and calculate correlation coefficient are explained below.

2.1 Data Collection

A detailed study by [7] analyse various parameters of rainfall prediction and figure out each parameter in meteorological features. Historical weather data for rainfall prediction was obtained from Weather and Climate. Fig.1 shows raw data for all parameters involved in this study collected in Kuantan, Pahang, Malaysia from year 2018 to 2020. The dataset included meteorological features such as Temperature (℃), Dew Point (℃), Humidity (%), Wind Speed (Kph), Pressure (Hg), and Precipitation (mm). In Fig.2, data for precipitation were separated from other parameters to observe the trend closely over the year 2018 until 2020. The data were manually copied and separated by monthly data into datasheets as database.

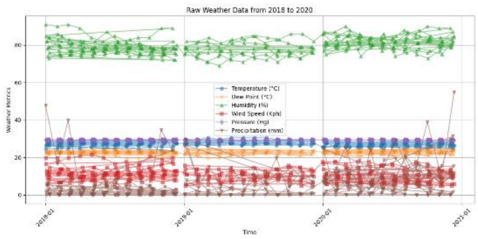


Fig. *1*. Raw Data for all parameters in 2018 until 2020

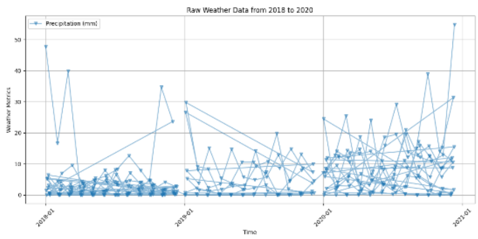


Fig. *2*. Raw data for precipitation in 2018 until 2020

2.2 Preprocessing Steps:

The implementation of ML was carried out using open-source software, Visual Studio Code as platform to code with support for operations development such as debugging, running task and control. The programming language used in this study is Python, which can be used to develop Ai. [8] said, Python language is commonly used to develop Ai due to its simple language, and it comes with built-in libraries for Ai projects usage such as NumPy, SciPy, matplotlib and some other important libraries.

Anaconda distribution tool kit was installed as computational environment framework also known as Integrated Development Environments (IDEs). Anaconda is functioned to perform Python and ML on same project and keep libraries always up to date in a single process. [9] said the most important libraries and IDEs in applications are included in Anaconda with a simple and well supported Graphical user interface (GUI). Nonetheless, Anaconda is an option for user to perform on Ai project.

Other than that, another python library, pandas were utilized are the first thing need to import right after start code a new project. Pandas library is a tool and structures specifically to perform data analysis, data manipulation and preprocessing due to its efficient handling of structured data [10]. Despite, pandas were chosen as statistical computing environments for this project.

The datasets are manually filled into .csv format and were arranged in tabular format to make it structured. There are few parameters measured which is temperature, humidity, wind speed, pressure, and precipitation. Load datasets from 2018 to 2020 into file path and ensure there is no null value or improperly formatted in the file. NumPy arrays were used to hold data into a structured format.

2.2.1 Support Vector Regression

Support Vector Regression (SVR) is one of ML algorithms which use to counter regression problem as it minimizes the prediction error. SVR have ability to convey non-linear models between precipitation and meteorological parameters.

Radial Basis Function (RBF) kernel was used since it is well-applied to non-linear models. [11] wrote, RBF kernel is a well-known kernel-based classifier since it requires a properly tuned parameter. In this study, we use hyperparameters tuning C=100 to control model complexity and achieving low error, gamma=0.1 as influence individual data points, and margin of tolerance for error prediction, epsilon=0.1. The meaning of hyperparameter is as Table 1.

Table 1. SVR Hyperparameter

Table 1 C	100	Regularization parameter
epsilon	0.1	Tube width
kernel	RBF	Type of kernel which is radial basis function (RBF)
gamma	scale	Kernel coefficient

Each ML algorithms in this study were evaluated the performance of Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and accuracy of prediction within $\pm 10mm$ precipitation. Figure 4 shows value of MAE is 3.31, RMSE is 5.11, and Accuracy is 94.22% for SVR algorithms.

2.2.2 Artificial Neural Network (ANN)

Among ML algorithms, ANN is a versatile algorithm which uses interconnected nodes from input layer, hidden layer, and output layer to solve problems by mimicking structure and function of human brain [12]. The input layer contains information or parameters we get from the data to process and analyze into the next layer, which is a hidden layer. The hidden layer will take data from the input layer, and each layer analyzes the data before processing to the next layer, which is the output layer. The output layer is a combination of data from input layer, and hidden layer that have been processed and came out with single result, or more than one for multi-class classification problem.

Three hidden layer neural network in this study. The first hidden layer is an input layer containing 64 neurons with ReLu activation. The second hidden layer contains

32 neurons with ReLu activation. The third hidden layer contains 16 neurons with ReLu activation. Output Layer which generates output value and comes out with 1 neuron to predict a single output value and for this study, the output is precipitation. The meaning of hyperparameters for ANN are as Table 2.

Table 2. Hyperparameter for ANN

Hidden layer size	(64,32,16,1)	Number of neurons in
Activation	'relu'	Activation function for hidden layer
Solver	'adam'	Optimization solver
Epochs	100	Model will be train for 100 times
Batch size	16	Process 16 samples at a time
Validation split	0.2	20% of data will be use as validation data, while remaining 80% will be used to train the model

Rectified Linear Unit (ReLu) is an activation function that is applied on neurons layer with computational calculations in Python using NumPy libraries, and TensorFlow by importing dense in ANN so it will automatically apply ReLu formula into code.

$$f(x) = \max(0, x)$$

This formula will be applied into code. If the input value of x is positive, the value remains unchanged. Meanwhile if the x is negative, the value will become 0.

To evaluate the performance of ANN, we performed RMSE, MAE, and accuracy evaluation. In the figure below shows value of MAE is 3.81, RMSE is 6.03, and accuracy is 93.62% for $\pm 10mm$ of precipitation prediction.

2.2.3 Random Forest Regression (RFR)

Random Forest Regressor is an ML algorithm that constructs combination of trees predictors, and each tree determines the values of a sampled randomly consisting of tree predictors. [13] said, generalization error of forest tree classifiers depends on the number of trees in forest, and the correlation between them. The higher the number of trees in forest, the lower the error rate of generalization for forest.

RFR was found capable of handling big sets of data and handling non-linear connections. Due to this, RFR is an option for this study as in this study, we involved complicated meteorological datasets where variables such as temperature, wind speed, precipitation, and humidity show complex dependencies.

For this research, the hyperparameter of RFR is set to value 100 for estimators, and 42 random state value. Maximum depth of trees was tuned to ensure sufficient range among the decision trees. Refer to the figure below for the value of hyperparameter RFR tuning. The meaning of hyperparameters for RFR are as Table 3.

Table 3. Hyperparameter for RFR

N estimators	100	Number of trees in the forest
Random state	42	Set random seed, ensuring results are reproducible

However, this study we identified that RFR slightly underperformed compared to ANN and SVR due to its accuracy of precipitation prediction within $\pm 10mm$ is 90.58% compared to ANN and SVR which have a slightly higher value of accuracy. Additionally, refer to figure below RFR's mean absolute error (MAE) value is 3.88 and root mean squared error (RMSE) value is 5.81 which is a competitive value in between SVR, ANN, and RFR. Overall, we can say RFR is a potential, but SVR was better for this type of dataset, and prediction for this study.

2.2.4 Linear Regression

Linear regression (LR) algorithms are commonly used in ML for regression tasks. [14] study, the relationship between a dependent variable and independent variable describes the fitting of linear equation to the observed data. The general equation of LR model and its meaning are as Table 4.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon$$

Table 4. The equation of linear regression

y	predicted value
$x_1, x_2, \ldots, x_n$	input
β_0	Intercept, representing value of y when x_i are zero
$\beta_1, \beta_2, \ldots \beta_n$	Weights on target variable
ϵ	Error

LR is one of the methods applied in this research to predict rainfall in Kuantan based on historical meteorological data that seems to match this ML algorithm. The dataset wat split into training data to estimate the coefficients (β_i) and testing to minimize the sum of squared between predicted and actual values.

Root mean squared error (RMSE) and mean absolute error (MAE) were evaluated in this study. However, after training it only shows baseline results for the predictions which is 83.89%. This study shows that LR algorithms disputed to train complex dataset which occurred to high RMSE and MAE values compared to other algorithms, SVR, ANN, and RFR.

3. Result and Analysis

This article wants to show a comparative analysis among four algorithms tested on meteorological dataset to predict precipitation in Kuantan, Pahang, Malaysia. Four ML algorithms consist of Artificial Neural Network (ANN), Support Vector Regression (SVR), Random Forest Regression (RFR), and Linear Regression (LR).

The algorithms were evaluated based on Root Mean Absolute Error (RMSE), Mean Squared Error (MSE), and accuracy within $\pm 10mm$ of precipitation. As a result, it shows patterns of predictions and actual data of rainfall, and competitive values of accuracy.

The following figure shows prediction vs. actual values of precipitation for all ML algorithms tested in this study. Fig. 3 below shows the graph plotted from Support Vector Regression (SVR) prediction algorithm, where the Precipitation as a parameter on Y-axis. While X-axis represents time which is three years of data from 2018 until 2020.

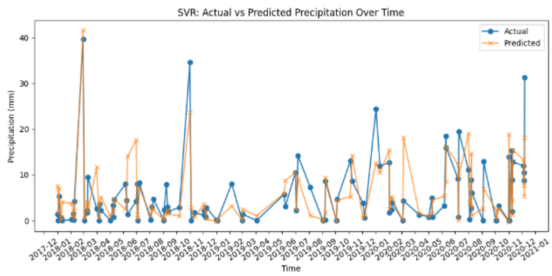


Fig. 3. Prediction data using SVR

To observe the differences in predictions, we use another algorithm to compare. In Fig.4 Artificial Neural Network (ANN) were used to predict Precipitation (mm) over Time from year 2018 until 2020.

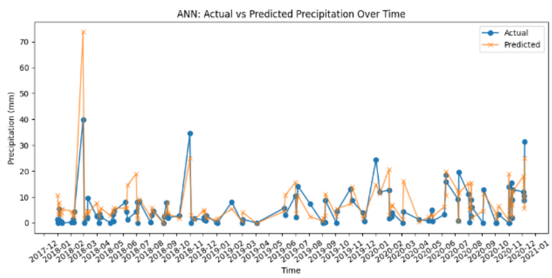


Fig. 4. Prediction data using ANN

Fig.5 shows the prediction and actual data for Precipitation (mm) over Time using Random Forest Regression (RFR) ML algorithm.

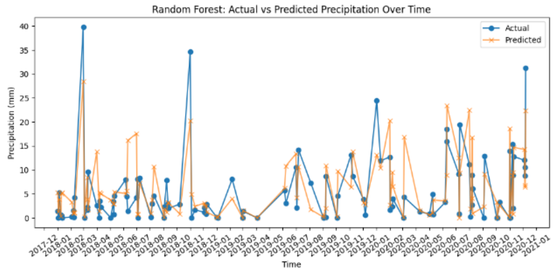


Fig. 5. Prediction data using RFR

In addition, we applied Linear Regression (LR) algorithm to compare and evaluate the prediction outcomes, and the result is as shown in Fig. 6.

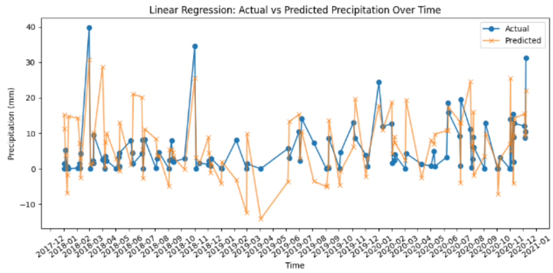


Fig. 6. Prediction data using LR

By comparing the results of all four algorithms and plotting into a graph as shown in Fig.7 will be able to see the variations, accuracy and highlight the differences between predictions and actual data.

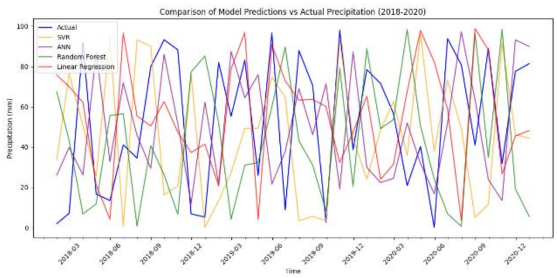


Fig. 7. Comparison of actual and predicted precipitation values from four algorithms

Table 5 Comparison results of prediction from four ML algorithms

ML Algorithms	RMSE	MAE	R^2 coefficient of determination	Accuracy of prediction 10mm (%)
SVR	5.11	3.31	0.53	94.22
ANN	6.03	3.81	0.34	93.62
RFR	5.81	3.88	0.39	90.58
LR	7.50	5.83	-0.03	83.89

The result and analysis in Table 5 show that simple ML algorithms like linear regression (LR) attempted to accurately predict precipitation, likely due to complex dataset which need more advance ML algorithms to perform. SVR performs as the most accurate and reliable algorithms, outstripping ANN, RFR, and LR models in all metrics.

4. Conclusion

This study reveals that advanced machine learning algorithms such as SVR and ANN are able to train and predict more extensive datasets and capture non-linear relationships. Comparing both competitive result of SVR, and ANN, SVR produce more outstanding result from all metrics that have been test, which is RMSE, MAE, R^2 and Accuracy. This highlights its ability to handle complex datasets and non-linearities. These findings present the importance of selecting proper algorithms for complex predictive datasets and optimizing hyperparameters to enhance the result of model's predictive capabilities.

Acknowledgements

This study was funded by Yayasan Pahang through Universiti Malaysia Pahang Al-Sultan Abdullah Agency Grant RDU240705.

References

1. N. Trisha, "Sg Jahang tragedy: Victim's brother didn't expect phone call was to be their last," *The Star*, Ipoh, pp. 1–1, Nov. 16, 2024.
2. J. Faouzi, "Time Series Classification: A Review of Algorithms and Implementations," 2024. doi: 10.5772/intechopen.1004810.
3. S. Walczak and N. Cerpa, "Artificial Neural Networks," in *Encyclopedia of Physical Science and Technology (Third Edition)*, R. A. Meyers, Ed., New York: Academic Press, 2003, pp. 631–645. doi: https://doi.org/10.1016/B0-12-227410-5/00837-1.
4. C. Gambella, B. Ghaddar, and J. Naoum-Sawaya, "Optimization problems for machine learning: A survey," *Eur J Oper Res*, vol. 290, no. 3, pp. 807–828, 2021, doi: https://doi.org/10.1016/j.ejor.2020.08.045.
5. M. Diamantopoulou, "Simulation of over-bark tree bole diameters, through the RFr (Random Forest Regression) algorithm," *Folia Oecologica*, vol. 49, pp. 93–101, Jul. 2022, doi: 10.2478/foecol-2022-0010.
6. K. Kumari and S. Yadav, "Linear regression analysis study," *Journal of the Practice of Cardiovascular Sciences*, vol. 4, p. 33, Jan. 2018, doi: 10.4103/jpcs.jpcs_8_18.
7. M. S. Pathan, J. Wu, Y. H. Lee, J. Yan, and S. Dev, "Analyzing the Impact of Meteorological Parameters on Rainfall Prediction," in *2021 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium)*, 2021, pp. 100–101. doi: 10.23919/USNC-URSI51813.2021.9703664.
8. S. Mihajlovic, A. Kupusinac, D. Ivetic, and I. Berković, *The Use of Python in the field of Artifical Intelligence*. 2020.
9. D. Rolon-Merette, M. Ross, T. Rolon-Merette, and K. Church, "Introduction to Anaconda and Python: Installation and setup," *Quant Method Psychol*, vol. 16, pp. S3–S11, May 2020, doi: 10.20982/tqmp.16.5.S003.
10. W. Mckinney, "pandas: a Foundational Python Library for Data Analysis and Statistics," *Python High Performance Science Computer*, Jan. 2011.
11. X. Ding, J. Liu, F. Yang, and J. Cao, "Random Radial Basis Function Kernel-based Support Vector Machine," *J Franklin Inst*, vol. 358, Oct. 2021, doi: 10.1016/j.jfranklin.2021.10.005.
12. P. Yu, M. Low, and W. Zhou, "Design of experiments and regression modelling in food flavour and sensory analysis: A review," *Trends Food Sci Technol*, vol. 71, Nov. 2017, doi: 10.1016/j.tifs.2017.11.013.
13. L. Breiman, "Random Forests," *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
14. R. Sarmento and V. Costa, "Introduction to Linear Regression," 2017. doi: 10.4018/978-1-68318-016-6.ch006.

Authors Introduction

Ms. Seri Liyana Ezamzuri

Graduated from Bachelor's degree in Mechatronics Engineering in 2023 from the Faculty of technology manufacturing and mechatronics in Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, Malaysia. Currently a Msc. Student in Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, Malaysia.

Dr. Sarah 'Atifah Saruchi

She received her Bachelor's in Mechatronics from Nagoya University, Japan, and completed her Master's and Ph.D. at Universiti Teknologi Malaysia. She is currently a lecturer at Universiti Malaysia Pahang

Asst. Prof. Dr. Ammar Al-Talib

He received his B.Sc and M.Sc degrees in Mechanical Engineering from the University of Mosul Iraq. He has finished his Ph.D degree from UPM University, Malaysia. Currently, he is working in UCSI University.

Autonomous Vehicle Navigation in Highway with Deep Q-Network (DQN) using Reinforcement Learning Approach

Sumiya Tamanna Fujita
Faculty of Manufacturing and Mechatronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah,26600 Pekan, Malaysia

Sarah Atifah Binti Saruchi
Faculty of Manufacturing and Mechatronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah,26600 Pekan, Malaysia

Ammar A.M. Al-Talib
Department of Mechanical and Mechatronics, Faculty of Engineering, Technology and Built Environment, UCSI University, 56000 Kuala Lumpur, Malaysia

Nurbaiti Wahid
School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, 23000 Dungun, Terengganu, Malaysia

Alvi Khan Chowdhury
Monash University Malaysia, Jalan Lagoon Selatan, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia

Siti Nurhafizza Maidin
Universiti Teknologi MARA, 23000 Dungun, Terengganu, Malaysia

Email: MIZ24006@adab.umpsa.edu.my, Sarahatifah@umpsa.edu.my, ammart@ucsiuniversity.edu.my, nurbaiti@uitm.edu.my, alvi.khanchowdhury@monash.edu, 2022987123@student.uitm.edu.my

Abstract

Autonomous vehicles (AVs) operate in highly dynamic environments, making them essential in modern transportation systems. Their success depends on real-time decision-making in unpredictable traffic scenarios, which are often beyond the scope of initial design assumptions. This unpredictability limits the effectiveness of traditional rule-based decision-making systems and predefined cost functions for real-time optimization. In critical applications like autonomous driving, reinforcement learning (RL) agents without safety mechanisms often struggle to converge or require extensive training data to produce reliable policies, leading to challenges in achieving safe and efficient operation. To address these challenges, this paper proposes a reinforcement learning (RL)-based framework. The ego vehicle refines its decision-making abilities by interacting with a simulated traffic environment. A short-horizon safety mechanism (SM) is integrated to ensure safer training by providing alternative safe actions during critical scenarios. The RL agent employs a deep neural network to map system states to optimal actions. The SM generalizes risky states, such as near-misses or collisions, rainy environment during night while creating a stable learning environment that enhances learning efficiency and enables meaningful exploration for optimal policy development. The method was validated in a highway driving scenario with varying traffic densities using the DQN algorithm and the CARLA simulator. Results demonstrated that the integration of the safety mechanism significantly improved learning efficiency and enabled the AV to make safe, reliable decisions even in complex and unpredictable traffic conditions.

Keywords: Autonomous Vehicles (AVs), Reinforcement Learning (RL), Deep Q-Network (DQN) Safety Mechanism (SM), CARLA Simulator, Highway Driving Simulation

1. Introduction

This research explores the application of the Deep Q-Network (DQN) algorithm for autonomous vehicle (AV) agents to navigate in highways in adverse weather conditions such as heavy rain. It is extremely challenging because of the reduced visibility, sensor noise, and unpredictable road dynamics. By training the DQN [1], [2], [4], [18] algorithm in simulated environment that replicates such harsh weather conditions, the AV agent learns to make real-time decisions for lane-keeping, obstacle avoidance, and speed optimization. In order to address a significant gap in the state of autonomous driving technology, the contribution consists of proving how reinforcement learning can effectively increase the resilience and adaptability of AV systems in the face of severe environmental uncertainties.

Many businesses and scientific research organizations are very interested in autonomous driving because of its enormous potential to change current mobility and, most importantly, reduce the frequency of traffic accidents [1] [6]. In recent years, autonomous driving has attracted a lot of interest from both the general public and researchers. Apart from their commercial potential, autonomous vehicles (AVs) have enormous potential to increase the number of available transportation options and potentially reduce carbon emissions [12].

Autonomous driving, like any other robotic system, involves decision-making. Implementing AV decision making usually involves resolving a number of path planning issues. A lot of calculation is needed for this method, and due to the solver's real-time limitations, the reference path that is generated may not result in a comfortable vehicle motion. In contrast, human drivers make decisions about changing lanes, maintaining speed, braking, and other maneuvers based on their perception of the traffic around them. Once this decision is made, the driver moves the steering and throttle/brake pedal to carry out the maneuver smoothly. The RL framework is a good fit for this approach [13].

The field of autonomous vehicle driving has advanced significantly because of advancements in computational tools like reinforcement learning (RL). One of the primary strategies for the creation of innovative transportation systems is RL's capacity to facilitate decision-making process adaptation through secure collaboration with the environment [5], [17]. The focus of basic RL research has been on low-dimensional problems. The algorithms work well in such predictable and transparent worlds. However, the shift to high-dimensional, complex scenarios, like those faced by self-driving cars, has prompted the creation of increasingly advanced techniques [8]. DQN, which was first presented as an expansion of conventional Q-learning, has since transformed the approach to resolving these difficulties [3].

In order to handle raw, high-dimensional sensory data, DQN uses deep learning technology. This allows for the direct learning of efficient policies from the complex input space without the need for expert features engineering. This ability introduced a new level of practicality concerning RL in autonomous driving as the network's ability to smartly convert continuous sensory streams into useful information is at the core of this sector [7]. DQN's success in a wide range of domains showed not only its versatility and tenacity but also its capacity to successfully negotiate the uncharted territory of the real driving world. In short, DQN has been shown to be able to learn a range of scenarios that can arise, from highway driving to urban congestion, and optimize both routes and driving styles [10].

A significant amount of training data may be needed for any RL methodology that attempts to solve both decision making and actuation at the same time. By clearly defining a hierarchy between high-level decision-making [1] and low-level actuation, we streamline this process in this work. For low-level actuation, traditional feedback control techniques like PID are employed, while RL is used to solve the decision-making problem. The approach described here somewhat resembles hierarchical reinforcement learning, in which control and decision-making are learned concurrently under a set of related but distinct cost functions [16]. We employ a modified version of the DQN algorithm from [13] to train the decision-making agent.

The trade-off between exploration and exploitation is a crucial feature that sets reinforcement learning apart from other types of machine learning [11]. To get a better reward, an agent requires to try events it hasn't tried before. When learning on a physical platform, like a robotic platform, the agent will therefore attempt every possible action during the initial learning phase. Unfortunately, this curiosity can be expensive and hazardous. Unguided exploration, for instance, may frequently result in collision or near-miss scenarios when training a highway driving agent. This could cause simulation resets, which would slow down the learning process. Furthermore, the trained agent may choose a non-safe maneuver even after convergence because of the function approximation made by the Q-network. It addresses these problems by adding an explicit short-horizon safety check to the DQN decision maker, which is utilized during the implementation phase as well as during learning.

The remainder of the paper is organized as follows. Section 1 presents related works. Section 2 describes the entire methodology of the proposed RL model. The experimental simulation and results are provided in Section 3. Finally, section 4 concludes the whole work.

2. Methodology

This section establishes an RL model for highway autonomous driving. Fig. 1 illustrates the RL process. An agent that interacts with the environment based on observations makes up an RL model. Through interaction with the traffic environment, the agent learns how to maximize the reward function by selecting the best course of action. In the sections that follow, the formulations of the main components of the RL model are displayed in detail. Autonomous driving at highway is regarded as an MDP [15]. MDP is a mathematical decision-making framework, which is described by (S, A, T, R, γ) are as Table 1.

Table 1: Symbol and their description

Symbol	Description
S	The state space. S_t is the state at time t ($S_t \in S$);
A	The action space. a_t is the action at time t ($a_t \in A$);
T	The transition model. It describes the transition probability from one state to another.
R	The reward function. R_t is the reward for deploying action a_t at state S_t.
γ	The discount factor, $\gamma \in [0,1]$. It is used to calculate the cumulative expected reward.

In RL model, the environment returns a numerical reward from a given reward function R based on the current state and the action that the agent takes. The goal of RL model is to learn an optimal strategy by maximizing the sum of discounted future rewards [1].

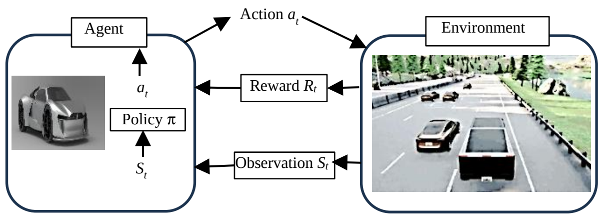


Fig.1 RL model for autonomous driving at highway

2.1 State Space

In our reinforcement learning simulation using CARLA, defining observations plays a critical role in creating a well-defined and precise state space. A single RGB front-facing camera was installed on the autonomous vehicle to capture the front road environment [14]. The images collected by this camera provide essential information about the surroundings. The action space is discrete, consisting of three actions: stop, drive slowly, and drive quickly.

2.2 Action Space

When driving straight across the highway, a vehicle can adjust throttle, steer and brake operations to accelerate, decelerate, or keep a constant velocity to ensure safety according to the surrounding traffic environment [9]. Therefore, the action space for intersection driving policy learning was defined as three velocity-related operations which has been given in the Table 2 below,

Table 2: Action and their description and values

Action	Throttle	Steer	Brake
0	0.0	0.0	0.5
1	0.3	0.0	0.0
2	0.5	0.0	0.0

2.3 Reward

Once an action is selected, an AV will get a resulting reward. The aim of the RL model about highway driving during rainy weather is to learn an optimal driving policy by maximizing the expectation of the discounted future reward, which indicates that different reward functions will lead to different driving policies. As a result, designing the reward function properly is crucial to directing AVs toward improved learning outcomes. The decision-making process should consider two requirements: (1) avoid collision with other vehicle, (2) do not be too conservative and travel as fast as possible [1].

The simulation's reward system is intended to motivate the agent to behave in a safe and effective manner. If a disruptive action occurs, like a collision, the reward is reduced by 0.7 and the episode is instantly terminated. The agent also receives a penalty of 0.2 for being too close to barriers (less than 1 meter) and 0.2 for keeping a safe distance (1 meter or more). The duration of the episode is also limited, and it ends once the predefined time threshold is exceeded. Furthermore, the reward is influenced by the distance traveled by the agent: traveling less than 30 meters incurs a small penalty of 0.1, while covering a distance between 30 and 100 meters earns a significant reward of 0.7 and ends the episode. The camera input is normalized (divided by 255) for further processing. This structured reward system helps the agent learn safe navigation and optimal behavior within the environment.

3. Results and Discussion

These days, creating secure AV systems requires testing AV models in incredibly lifelike virtual environments. The intricacy of autonomous driving scenarios necessitates testing the suggested methods in simulators that are similar to the real world in order to confirm their effectiveness. In light of this, we test our model in CARLA Simulator [2], which is regarded as the most potent simulator due to its realistic environments and features and suitability for AV technology applications. This section presents the results obtained using our proposed approach. As we can see in Fig.2 and Fig.3 shows the trained DQN model average reward over 100,000 timesteps. It took 1 hour of time to train the model. Four-hundred other vehicles were used in the CARLA environment to complicate the environment for the AV agent. CARLA town-04 highway was utilised for DQN model training purpose.

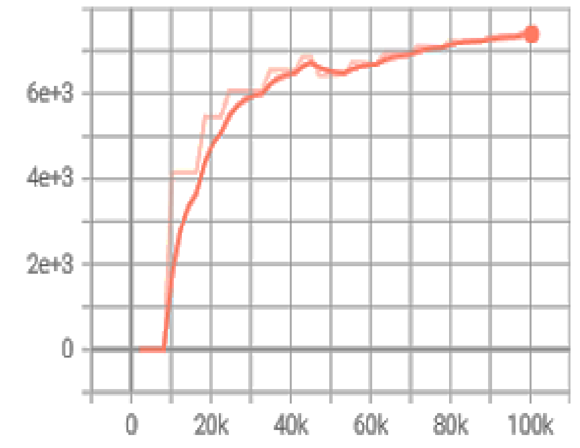


Fig.2 Mean Reward (Y-axis) with Timesteps Trained (X-axis)

Fig.3. AV agent during rainy weather

To further complicate the training environment, soft rain during nighttime was used to train the model. The rise and consistency of the trained model reward with respect to training timesteps shows that the AV agent was able to drive well in this complicated scenario.

4. Conclusion

The This paper presented a control architecture based deep RL framework for safe decision making in autonomous driving. By using well known feedback controller, high level decision by DQN is converted into a low-level actuation throttle, brake, and steering to control an AV in a complex scenario with dense traffic. We have evaluated the learned controller under varying traffic density, the results demonstrate the superior capabilities of the learned AV agent. Lastly, we have illustrated the continuous adaptation framework that has been proved to decrease the frequency of safety triggers by altering the learning method. The CARLA simulator was used to validate the suggested approach, and the outcomes showed how effective our model is at learning, maintaining stability, and achieving rapid convergence. Given that DQN is among the data sets used in methods that are particularly impacted by noisy or insufficient data [10], we intend to expand this approach in the future to include required lane changes and on-ramp to off-ramp highway navigation.

Acknowledgements

This work was supported in part by Universiti Malaysia Pahang Al-Sultan Abdullah, Faculty of Manufacturing and Mechatronics Engineering Technology. The authors gratefully acknowledge the assistance and support provided throughout this research. Additionally, we extend our sincere gratitude to our co-authors for their valuable contributions to the successful formulation and development of this study.

References

1. Deep Reinforcement Learning Enabled Decision-Making for Autonomous Driving at Intersections Guofa Li1,2 · Shenglong Li1 · Shen Li3 · Yechen Qin4 · Dongpu Cao2 · Xingda Qu1 · Bo Cheng5
2. A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, “Carla: An open urban driving simulator,” in Conference on robot learning. PMLR, 2017, pp. 1–16.
3. J. Schulman et al., "Proximal Policy Optimization Algorithms," arXiv preprint arXiv:1707.06347, 2017.
4. M. Alshiekh, R. Bloem, R. Ehlers, B. K¨onighofer, S. Niekum, and U. Topcu, “Safe reinforcement learning via shielding,” arXiv preprint J. Garcıa and F. Fern´andez, “A comprehensive survey on safe reinforcement learning,” Journal of Machine Learning Research, vol. 16,no. 1, pp. 1437–1480, 2015.arXiv:1708.08611, 2017.
5. J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," Neural Networks, vol. 61, pp. 85-117, 2015.
6. K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, “Deep reinforcement learning: A brief survey,” IEEE Signal Processing Magazine, vol. 34, no. 6, pp. 26–38, 2017.
7. L. Busoniu, R. Babuska, and B. De Schutter, "A comprehensive survey of multiagent reinforcement learning," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 38, no. 2, pp. 156-172, 2008.
8. V. Mnih et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529- 533, 2015.
9. J. Garcia and F. Fern´andez, “Safe exploration of state and action spaces in reinforcement learning,” Journal of Artificial Intelligence Research, vol. 45, pp. 515–564, 2012.
10. D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," Nature, vol. 529, no. 7587, pp. 484-489, 2016.
11. R. S. Sutton and A. G. Barto, Reinforcement learning: An introduction. MIT press Cambridge, 1998, vol. 1, no. 1.
12. B. Paden, M. ˇC ´ap, S. Z. Yong, D. Yershov, and E. Frazzoli, “A survey of motion planning and control

techniques for self-driving urban vehicles," IEEE Transactions on Intelligent Vehicles, vol. 1, no. 1, pp. 33–55, 2016.

13. Autonomous Highway Driving using Deep Reinforcement Learning, Subramanya Nageshrao1 and Eric Tseng2 and Dimitar Filev2
14. H. Xu, Y. Gao, F. Yu, and T. Darrell, "End-to-end learning of driving models from large-scale video datasets," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2174–2182.
15. Bellman, R.: A Markovian decision process. J. Math. Mech. 6(5), 679–684 (1957)
16. K. Frans, J. Ho, X. Chen, P. Abbeel, and J. Schulman, "Meta learning shared hierarchies," arXiv preprint arXiv:1710.09767, 2017.
17. C. Chen, A. Seff, A. Kornhauser, and J. Xiao, "Deepdriving: Learning affordance for direct perception in autonomous driving," in Computer Vision (ICCV), 2015 IEEE International Conference on. IEEE, 2015, pp. 2722–2730.
18. Arulkumaran, K., Deisenroth, M.P., Brundage, M., et al.: Deep reinforcement learning: a brief survey. IEEE Signal Process. Mag. 34(6), 26–38 (2017)

Authors Introduction

Sumiya Tamanna Fujita

She completed her Bachelor's Degree in Computer and Communication Systems Engineering in 2015 from Universiti Putra Malaysia. Presently, she is pursuing her Master's degree in the Faculty of Manufacturing and Mechatronics Engineering Technology at Universiti Malaysia Pahang Al-Sultan Abdullah. Her research interests include planning and control, vehicle dynamics, autonomous vehicle systems, artificial intelligence, and robotics.

Sarah Atifah Binti Saruchi

She received her B.Eng. in Mechanical and Aerospace Engineering from Nagoya University, Japan. She received her Master and Doctoral degrees from Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia. Currently, she is working as senior lecturer at Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, Malaysia. Her research interests include machatronics and artificial intelligence.

Ammar Abdulazizf Al Talib

He has finished his B.Sc. and M.Sc. degrees in Mechanical Engineering from the University of Mosul/Iraq. He has finished his Ph.D. degree from UPM University/ Malaysia. He is also a chartered Engineer and Member of the Institute of Mechanical Engineers/UK.(CEng. MIMechE). He has developed all the Postgraduate Programs at the Faculty of Engineering at UCSI University / Malaysia, and worked as the Head of Postgraduate and Research department at the same faculty for the years 2010-2018.

Nurbaiti Wahid

She received the M.Eng. in electrical engineering from Universiti Teknologi Malaysia in 2009 and Ph.D. degrees in advanced vehicle system from Universiti Teknologi Malaysia. Currently, she is a Senior Lecturer and Researcher at the College of Engineering, Universiti Teknologi MARA, Malaysia since 2009. Her research interests include planning and control, vehicle dynamics, autonomous vehicle systems, artificial intelligence, and robotics.

Alvi Khan Chowdhury

He received his Bachelor's Degree of Mechatronic Engineering in 2023 from UCSI University, Kuala Lumpur Campus. Currently, he is pursuing his Master's Degree of Engineering Research in Monash University Malaysia Branch.

Siti Nurhafizza Binti Maidin

She received her bachelor's degree in engineering in 2016 from Faculty of Electrical Engineering, Universiti Teknologi MARA, Malaysia. She also had received her Master's degree in 2018 from Faculty of Advance Informatic School. She is currently a PhD student in Universiti Teknologi MARA, Malaysia.

Empowering Decentralized Microgrids with A Blockchain-Based Peer-To-Peer Energy Trading Platform

Firas Basim Ismail
Power Generation Unit, Institute of Power Engineering (IPE), Universiti Tenaga Nasional (UNITEN) Kajang, 43000, Malaysia

Ammar A.M. Al-Talib
Department of Mechanical and Mechatronics, Faculty of Engineering, Technology and Built Environment, UCSI University, 56000 Kuala Lumpur, Malaysia

Nizar F.O. Al-Muhsen
Technical Instructor Training Institute, Middle Technical University, Baghdad, Iraq.

Chetenraj Singh
Power Generation Unit, Institute of Power Engineering (IPE), Universiti Tenaga Nasional (UNITEN) Kajang, 43000, Malaysia

e.mail: firas@uniten.edu.my

Abstract

This study explores the concept of a Peer-to-Peer Energy Trading Platform for Decentralized Microgrids through Blockchain Technology. The presented work focuses on the creation of a decentralized peer-to-peer energy trading platform using blockchain technology, which is designed for microgrid ecosystems. The study delves into blockchain's ability to build trust and ensure the integrity of energy transactions. The study concludes with recommendations for future enhancements, including expanding platform capabilities, integrating diverse energy sources, and navigating regulatory challenges for widespread adoption and impactful change.

Keywords: Peer-to-Peer (P2P), Blockchain Technology, Energy Trading, Decentralized Microgrids

1. Introduction

Adopting blockchain technology as a P2P energy trading platform can offer a decentralized, distributed digital ledger that securely records transactions among prosumers in an energy marketplace. It ensures transparency, immutability, and trust without relying on a centralized authority, making it an ideal technology for managing energy trading in microgrid society [1]. The potential usage of blockchain technology can help facilitate and create a P2P energy trading platform. Besides, a blockchain system can record all transactions that are maintained across computers [2].

Cryptocurrencies like Bitcoin have been used for Double-spending as effective digital assets between the prosumers in the P2P energy trading platform. Practically, the transaction inputs (sender's address, energy consumer), transaction output (recipient's address, energy producer), and the amount of cryptocurrency involved in the energy trading process are the three main components of each block on the Bitcoin blockchain [3]. Ethereum blockchain is open-source, which is a worldwide platform that enables users to store data precisely and monitor digital assets via smart contracts. Every activity on the network requires a certain quantity of computational power, or "gas". Gas limits are the most permitted units of gas for a specific operation, and they are determined and enforced by Ethereum's blockchain for every operation. The main purpose of the Ethereum blockchain is to support decentralized applications [3], [4].

By expanding on the Transactive Energy Management System, peer-to-peer energy trading can be carried out more easily and without the need for a middleman. Through the DApp, each user registers on the system and provides the details needed to engage in energy trading. The Transactive Energy Blockchain (TEBCN) smart contract records this data in the blockchain system together with the amount of energy to purchase or sell and the price per megawatt-hour. The DApp system asks the back-end server for a response. By taking this action, the back-end server is prompted to enable communication between agents to guarantee the transfer of the designated quantity of energy. The energy seller receives the energy cost expressed in ethers once the back-end server confirms the energy transfer [5], [6], [7].

Niloy F.A. et al. [8] brought up the idea of a network between peers where the participants pose as clients or even as servers with the ability to share their assets safely. Energy trading in the smart grid can be accomplished more effectively by utilizing a P2P network in blockchain technology. This P2P energy trading platform enables the prosumers to their need for kWh from the closest source directly. Furthermore, in this case, users of P2P energy sharing will utilize blockchain-based smart contracts, which are autonomous. Like this, these blockchain-endorsed smart contracts will process and store the records of each transaction that occurs within the system through the Advanced Metering Infrastructure (AMI) [9].

Most conventional energy trading is a one-way strategy. Large-scale generators usually send their electricity over great distances to end-users. Commonly, two or more middlemen exist between the energy producers and consumers to complete the energy trading cycle. In this case, the end-users are paying not only for middlemen but also for power transmission. This increases the energy unit price multiple times. Therefore, buying the energy directly between peers and reducing the cost of transmission could significantly reduce the energy unit price. For instance, the Reserve Bank of Australia reported that only 45% is the energy unit's whole sale price meanwhile the rest 55% is divided into retail operation costs by 10% and network costs by 45% [10]. Therefore, buying the energy directly between peers and reducing the cost of transmission could significantly reduce the energy unit price. The P2P energy trading could also promote decentralized trading within a local geographic area. Accordingly, new market strategies are needed to determine pricing, decentralize the energy market, and control the energy infrastructure considering the rise in renewable energy use at the domestic level. Establishing such a market could enable prosumers to transact directly on renewable energy generation without the need for middlemen, which would be granting a much fairer price in buying a selling energy unit. It was reported that P2P energy trading in microgrids could improve the local balance between energy generation and consumption caused by the diversity of the produced energy resources [11].

As shown from the literature, P2P energy trading can be categorized into centralized, distributed, and decentralized markets. Research in the previous works focused on improving scalability, reliability, privacy, and autonomy in centralized markets, designing proper pricing mechanisms, modelling decision-making processes in distributed markets, and utilizing blockchain technology for decentralized platforms.

The novelty of this work lies in the innovative application of blockchain technology to create a decentralized P2P energy trading platform tailored for microgrids. Unlike traditional energy systems, this platform enables direct energy transactions between users without the need for intermediaries, ensuring greater economic efficiency, transparency, and security. The integration of IoT sensors for real-time monitoring, combined with the future use of smart contracts within the IBM Hyperledger Fabric framework, represents a significant advancement in automating and securing energy trades. This study uniquely addresses the practical challenges of implementing blockchain in energy trading, such as ensuring data accuracy, managing user interactions, and maintaining system performance under varying conditions.

2. Methodology

This subsection integrates principles from blockchain selection, system architecture design, smart contract development, and decentralized microgrid integration into a structured process. An efficient and secure energy trading platform has been established. It mainly aims to serve community-driven energy exchanges, promote sustainability, and enhance the resilience of decentralized energy infrastructures. Adopting blockchain solutions in a P2P energy trading platform in a decentralized microgrid structure will completely change how prosumers deal with their excess renewable energy resources. Practically, the prosumers have complete control over the amount of energy that they want to buy or sell amongst other prosumers. For instance, the client has the choice to trade the energy quantity and source, renewable or not, in the same microgrid and national grid.

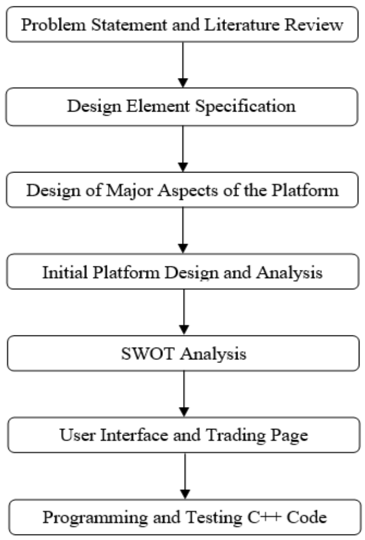


Fig. 1 Methodology Flow Chart.

Figure 1 demonstrates a flowchart for the conducted methodology which starts with determining the problem statement supported by a relevant literature review. Subsequently, design modelling would be executed then software/program selection and finalising. The next step comprises validation and analysis of the proposed design to be tested to ensure its viability and safety. Finally, the manufacturing and testing stages would be performed. During this stage, the design product would be created to demonstrate its capability and a better visual representation.

2.1. *Blockchain Selection*

Ethereum blockchain was primarily developed for the exploitation and support of decentralised applications. Using the Ethereum blockchain allows for trading and dealing with tokens for energy transactions. Besides, Ethereum allows for maintaining digital ledger transactions, where all the trading activities performed using any web interface are recorded with timestamps. It also allows the designing of smart contracts, which authorizes the creation of real-time contracts in the energy market. Additionally, the Ethereum blockchain can be used for distribution application deployment and testing in a deterministic and secure environment. Furthermore, the Ethereum blockchain can be linked and integrated with an external key software such as MetaMask which is necessary to complete this work. These selected features, including the Ethereum blockchain, ultimately enable the P2P energy trading fairly and securely.

2.2. *System Architecture Design*

The Transactive Energy Blockchain (TEBCN) platform integrates three primary frameworks: the Multi-Agent System (MAS), the Ethereum blockchain, and the Distributed Application (DApp). This architecture facilitates secure and efficient decentralized energy exchanges. The Ethereum blockchain records energy costs and transaction amounts, enhancing transparency and reliability. Smart meters installed at each prosumer's location provide accurate energy usage data, supporting the MAS in simulating energy flows. The MAS framework, implemented using the Java Agent Development Framework (JADE) and the FIPA-Agent Communication Language (ACL), manages agent interactions and transaction processing. The MAVEN software platform integrates MAS with the blockchain network and DApp, ensuring centralized project management. Each participant is assigned an Ethereum blockchain address, with smart contracts compiled and deployed using the Truffle development environment and simulated on the Ganache blockchain emulator. MetaMask, a cryptocurrency wallet, allows users to monitor their accounts and execute transactions. Overall, this system architecture ensures secure recording of data changes and smart contract executions on the blockchain, providing a robust platform for decentralized energy trading.

2.3. *Decentralized Microgrid Integration*

In the decentralized microgrid, each prosumer controls their energy exchange, enhanced by blockchain technologies for heightened security. This integration ensures a balance between energy generation and consumption, reducing the wastage of renewable energy sources. In the cases of excess or insufficient energy, mutual trading within the utility network enables energy balance among microgrid users. This decentralized, or smart microgrid environment, ensures uninterrupted power trading and acts as an aggregator to manage power routing. Prosumers and microgrids can collaborate to improve services and reduce electricity costs, adapting to changing environments during extreme events to ensure efficient energy trading. This cooperation makes electricity prices more affordable and promotes profitable trading of transactive energy nodes. Every house in the microgrid is equipped with a smart meter to analyze power usage and predict surplus or deficit states. Consumers can decide how much power to purchase or sell within the microgrid. Once stabilized, separate microgrids can connect for broader possibilities, enhancing the decentralized energy trading system.

3. Results and Discussions

3.1. *Installation of IoT Measuring Sensor*

The first step in setting up the whole system would be to gain access to the power generated by the solar panels and the house's power consumption. This is vital in determining whether that specific house has a positive excess power value or a negative one. This is an essential parameter because the house user is only able to trade the excess power with other users. Moreover, installing an IoT sensor that can monitor and track the power generation and consumption values in actual time is important to ensure a seamless connection between the excess power values and the other users who are using the platform. For this purpose, a Sense Home Energy Monitor is used as an IoT sensor. This IoT sensor can monitor continuously the flow of electricity through the main electric panel of the house. Practically, it can determine how much energy is generated by the solar system of the house, as well as how much is the total energy consumption of that house. The data that is tracked by this IoT sensor will send the live data to its cloud server. We will then take this live data from the Sense cloud server and duplicate these values onto our platform’s server. This is to ensure that the actual power excess value and the excess power value on the platform are the same. This works perfectly to sync the real-time data with the data displayed on the platform.

3.2. *User Account Set Up*

Once the Sense IoT sensor is installed and synced up, the account will be set for the users on the platform. These users will be given a unique Account Identification (ACC ID) username and password. Each house will only be able to make a single account. This is also done to ensure both the safety and security of the user, other users, and the entire platform. If required, we can also provide the service of briefing the user on how to navigate and use the platform. When the account is set up, the power excess value of the house will be displayed on the platform once the user logs into the platform. The users can also check their account information, check their excess energy value, buy or sell their excess energy, and check their transaction history, as shown in Figure 2.

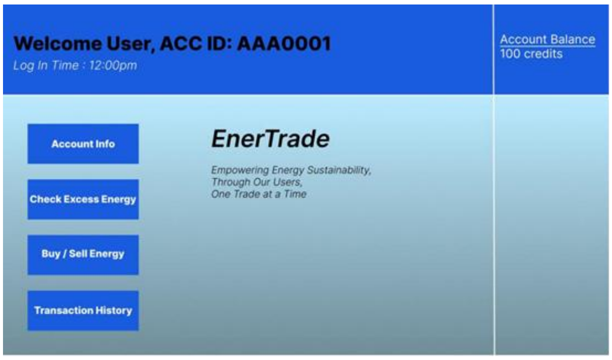


Fig. 2 Sample of the User Interface after Logging In.

3.3. *Converting Excess Power to Platform Currency*

When each user has their platform account set up, they can begin trading excess power on the platform. However, to continue with this step, the users need to convert their access energy into the blockchain currency via Ethereum Blockchain. One easy way to conduct this consists of two steps. The first is to convert the excess power from the houses into Malaysian Ringgit. This is done by calculating the value of the excess power according to the Domestic Tariff (Tariff A) that is set by TNB. Once the excess power is calculated in the Malaysian Ringgit value, we can convert the value of the excess power in Malaysian Ringgit to the currency used on the application, Ethereum, via Ethereum Blockchain. Once the platform has converted the value of the user's excess power from Malaysian Ringgit to Ethereum Blockchain, the currency will be updated in the user's account. Finally, the user can trade, which is either buying or selling their excess power on the platform, as shown in Figure 3.

TARIFF CATEGORY	UNIT	CURRENT RATE
Tariff A - Domestic Tariff		
For the first 200 kWh (1 - 200 kWh) per month	sen/kWh	21.80
For the next 100 kWh (201 - 300 kWh) per month	sen/kWh	33.40
For the next 300 kWh (301 - 600 kWh) per month	sen/kWh	51.60
For the next 300 kWh (601 - 900 kWh) per month	sen/kWh	54.60
For the next kWh (901 kWh onwards) per month	sen/kWh	57.10
The minimum monthly charge is RM3.00		

Fig. 3 Tariff category according to the domestic tariff.

3.4. *Fully Functioning User Account on the Platform*

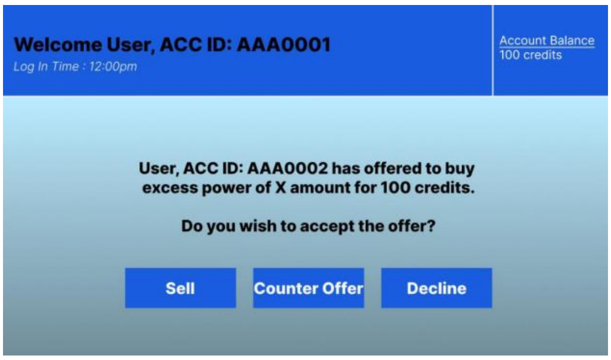


Fig. 4 Screen shot for a sample of a Trade between two platform users

Figure 4 demonstrates that once the user's accounts are completed, they have the choice to buy or sell their energy through the platform. In the proposed platform, the user with the ACC ID of AAA0002 is offering to buy "X" amount of excess energy from another peer with the ACC ID of AAA0001 for 100 Credits. Besides, user AAA0002 requires this amount of excess energy for their house's usage, thus the AAA0001 can offer to buy the excess energy from user AAA0001 for 100 credits. In this case, the user AAA0001 has the option to either sell his excess home-generated energy for 100 credits, counteroffer for a better price or decline the offer. If the user AAA0001 chooses to accept the offer, the 100 credits from user AAA0002 will be transferred to the app first as a third party. Once the user AAA0002 receives the requested energy, the 100 credits will be released to the user AAA0001. This procedure could prevent any unreasonable refunds that can be requested because Ethereum performs a one-way transaction. This is because a refund scenario could make the P2P energy trading process complicated or even inapplicable.

3.5. *C++ Coding Results*

Figures 5-9 display the C++ programming code while it in operation. These figures also demonstrate the C++ programming code outcomes while it in operation.

```
Output                                                              Clear
/tmp/RVcp8pI2LI.o
AAA0002 offers to buy 100 units of power from AAA0001 for 100 credits.
AAA0001, you have received an offer from AAA0002 to buy 100 units of power for 100
    credits.
Do you want to [S]ell, [C]ounter offer, or [D]ecline?
```

Fig. 5 Initial outcome of the code prior to selecting an option.

```
Output                                                              Clear
/tmp/w94QZLBmr7.o
AAA0002 offers to buy 100 units of power from AAA0001 for 100 credits.
AAA0001, you have received an offer from AAA0002 to buy 100 units of power for 100
    credits.
Do you want to [S]ell, [C]ounter offer, or [D]ecline? S
Transaction Successful!


=== Code Execution Successful ===
```

Fig. 6 Outcome after Selecting the 'Sell' option.

```
Output                                                              Clear
/tmp/CNsUmGnVzR.o
AAA0002 offers to buy 100 units of power from AAA0001 for 100 credits.
AAA0001, you have received an offer from AAA0002 to buy 100 units of power for 100
    credits.
Do you want to [S]ell, [C]ounter offer, or [D]ecline? C
Enter your counter offer price: 120 credits
AAA0002 accepts the Counter Offer!


=== Code Execution Successful ===
```

Fig. 7 The successful outcome after selecting the 'Counter Offer' option.

```
Output                                                              Clear
/tmp/oBSkD561J4.o
AAA0002 offers to buy 100 units of power from AAA0001 for 100 credits.
AAA0001, you have received an offer from AAA0002 to buy 100 units of power for 100
    credits.
Do you want to [S]ell, [C]ounter offer, or [D]ecline? D
Transaction Failed!


=== Code Execution Successful ===
```

Fig. 8 The Outcome after Selecting the 'Decline' option.

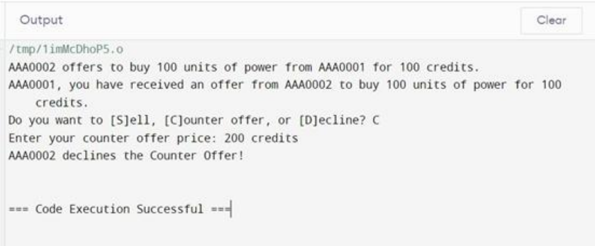


Fig. 9 The unsuccessful outcome after selecting the 'counter offer' option.

4. Conclusion

A blockchain-based peer-to-peer (P2P) energy trading platform for decentralized microgrids was developed and implemented to ensure a secure, transparent, and efficient energy transaction. The platform's design, which includes the integration of IoT sensors and the use of IBM Hyperledger Fabric for smart contracts, facilitated real-time monitoring and automated trading processes. This approach could address key challenges such as economic efficiency, data accuracy, and transaction security. Additionally, the developed platform is capable of converting excess energy into Ethereum for trading effectively demonstrating a novel method for energy monetization. Moreover, reduced transaction costs, increased renewable energy usage, and enhanced consumer empowerment validated the platform's practical application and its potential to revolutionize energy trading are another major outcome of this study.

Nevertheless, several constraints were encountered during the development and implementation of the blockchain-based peer-to-peer (P2P) energy trading platform for decentralized microgrids. Scalability of the blockchain network, while effective for small-scale deployments, may face challenges in handling a larger number of transactions and users simultaneously. Additionally, the reliance on IoT sensors for real-time data collection introduced potential vulnerabilities related to data accuracy and sensor malfunctions, which could affect the reliability of energy transactions. The integration of Ethereum for energy trading, although innovative, posed challenges in terms of regulatory compliance and market acceptance, as the cryptocurrency market is still subject to significant volatility and legal scrutiny

Acknowledgment

This research was funded by Universiti Tenaga Nasional (UNITEN), Malaysia, through the BOLD Research Grant 2024, under project code J510051055.

References

1. M. Mehdinejad, H. Shayanfar and B. Mohammadi-Ivatloo, "Decentralized blockchain-based peer-to-peer energy-backed token trading for active prosumers," *Energy*, vol. 244, p. 122713, 2022.
2. A. Esmat, M. de Vos, Y. Ghiassi-Farrokhfal, P. Palensky and D. Epema, "A novel decentralized platform for peer-to-peer energy trading market with blockchain technology," *Applied Energy*, vol. 282, p. 116123, 2021/01/15/ 2021, doi: https://doi.org/10.1016/j.apenergy.2020.116123.
3. A. Iskakova, H. K. Nunna and P. Siano, "Ethereum blockchain-based peer-to-peer energy trading platform," in *2020 IEEE international conference on Power and Energy (PECon)*, 2020: IEEE, pp. 327-331.
4. V. Buterin. A Next-Generation Smart Contract and Decentralized Application Platform [Online] Available: https://ethereum.org/en/whitepaper/
5. F. Bellifemine, A. Poggi and G. Rimassa, "Developing multi‐agent systems with a FIPA‐compliant agent framework," *Software: Practice and Experience*, vol. 31, no. 2, pp. 103-128, 2001.
6. H. K. Nunna, A. Sesetti, A. K. Rathore and S. Doolla, "Multiagent-based energy trading platform for energy storage systems in distribution systems with interconnected microgrids," *IEEE Transactions on Industry Applications*, vol. 56, no. 3, pp. 3207-3217, 2020.
7. K. Ma, Z. Wang, Y. Zhao and H. Wang, "Decoding Web3: In-depth Analysis of the Third-Party Package Supply Chain," in *Proceedings of the 15th Asia-Pacific Symposium on Internetware*, 2024, pp. 457-466.
8. F. A. Niloy, M. A. Nayeem, M. M. Rahman and M. N. U. Dowla, "Blockchain-based peer-to-peer sustainable energy trading in microgrid using smart contracts," in *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, 2021: IEEE, pp. 61-66.
9. J. Li, J. Wu and L. Chen, "Block-secure: Blockchain based scheme for secure P2P cloud storage," *Information Sciences*, vol. 465, pp. 219-231, 2018.
10. (2024). *HOW ARE ELECTRICITY PRICES SET IN AUSTRALIA?* [Online] Available: https://www.rba.gov.au/information/foi/disclosure-log/pdf/101115.pdf
11. C. Zhang, J. Wu, Y. Zhou, M. Cheng and C. Long, "Peer-to-Peer energy trading in a Microgrid," *Applied Energy*, vol. 220, pp. 1-12, 2018/06/15/ 2018,doi:https://doi.org/10.1016/j.apenergy.2018.03.010.

Authors Introduction

Dr.Firas Basim Ismail

Dr. Firas earned his Ph.D. in Mechanical Engineering from Universiti Teknologi PETRONAS, Malaysia, in 2011. He is the founder and head of the Power Generation Research Centre at UNITEN, Malaysia, with over 16 years of academic and industrial experience. His expertise includes Thermo-Fluid, Renewable Energy, and Digital Twin Technology.

Asst. Prof. Dr.Ammar Al-Talib

He received his B.Sc and M.Sc degrees in Mechanical Engineering from the University of Mosul Iraq. He has finished his Ph.D degree from UPM University, Malaysia. Currently, he is working in UCSI University.

Dr. Nizar F.O. Al-Muhsen

He was awarded his Ph.D. from the Faculty of Engineering and Information Technology, University of Technology Sydney, Australia. His research collaboration focuses on research fields of energy systems' developments including Renewable Energy systems.

Chetenraj Singh

He is an Applications Engineer from the Department of Engineering Solutions from Hitachi Sunway Information Systems. He graduated from Universiti Tenaga Nasional's (UNITEN) in 2024. He received his Bachelor Degree (Hons.) in Mechanical Engineering. His area of research was Renewable Energy.

Performance of Kenaf Fibre Reinforced Epoxy Biocomposite for High Voltage Insulator Application

Kang Rui Tan
Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur, 56000, Malaysia

Cik Suhana Hassan
Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur, 56000, Malaysia

Nor Fazilah Abdullah
Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur, 56000, Malaysia

Farah Adilah Jamaludin
Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur, 56000, Malaysia

Meng-Choung Chiong
Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur, 56000, Malaysia

Eryana Eiyda Hussin
Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur, 56000, Malaysia

Email: 1001851161@ucsiuniversity.edu.my, suhana@ucsiuniversity.edu.my, norfa@ucsiuniversity.edu.my, FarahAdilah@ucsiuniversity.edu.my, ChiongMC@ucsiuniversity.edu.my, EryanaEiyda@ucsiuniversity.edu.my

Abstract

Current applications of natural fibre-reinforced composite are widely spread in various applications such as construction and building structures as well as automotive parts, in which compression and flexural properties are the primary concern of researchers. The aim of this research is to evaluate the performance of kenaf fibre-reinforced epoxy composite as the core of high voltage insulator subjected to identified wind load conditions by using finite element analysis which is then further validated using the theory of mechanics of materials. It was found that under tensile load, the kenaf fibre reinforced polymer (FRP) experiences 2 times higher elongation compared to synthetic glass FRP but the tensile stress is only 21% of its material strength; under torsion load, despite having a 2 times higher angle of rotation compared to glass FRP, the deflection of kenaf FRP is still 24 % below the typical damage limits; under bending load, the kenaf FRP experiences slightly higher bending stress compared to the glass FRP but the maximum stress is only 19% of its material strength. Under electrical load, the kenaf FRP experiences 13.6 % higher electrical stress than glass FRP and it shows that the electrical performance of kenaf FRP is almost comparable to the glass FRP. Even though the kenaf FRP naturally has lower strength, the performance of kenaf FRP is almost comparable to the conventional material from glass FRP and the results indicate that the kenaf FRP is able to sustain different types of loads when it is under operating.

Keywords: Insulator, High Voltage Insulation, Biocomposite, Kenaf fibre

1. Introduction

An electric power system is a network of electrical components that involved generation, transmission, and distribution systems. A transmission line system is an essential system that facilitates the bulk movement of electrical energy from the generating station to the distribution system. The high voltage insulator is a component in the transmission line network used to provide insulation between the transmission lines and the transmission towers which are exposed to the atmosphere. It also provides mechanical support to carry the weight of the transmission lines galloping through the towers as well as multiple types of external force on the transmission lines from the environment. Composite insulator is the modern type of high voltage insulator used and it consists of three parts: core, housing, and end fitting. The core as the main component to provide mechanical strength is a rod typically made of glass fibre-reinforced polymer using the pultrusion method. In such method, the glass fibres impregnated in resin matrix are aligned unidirectionally to provide superior tensile performance [1].

Cherney [1] and Papailiou and Schmuck [2] agreed that the core, as known as the mandrel, is the main component to provide mechanical strength and withstand the mechanical load brought by the transmission line and transfer the tensile load to the tower. The core is typically made of fiberglass rods where E-glass and stress corrosion E-glass (ECR) glass are the two common types of glass fibres used in the manufacture of insulator rods and polyester, epoxy resin and vinyl ester are the three types of resin systems in use [1]. The housing, also known as weather sheds, is the component surrounding the core rod to provide electrical insulation and protect the core from environmental erosion. Gorur [3] states that there are different material families, including high-temperature vulcanized (HTV) silicone rubber, ethylene propylene rubber (ERP), cycloaliphatic epoxy, and ethylene vinyl acetate (EVA), which are proven suitable for outdoor insulator housing. Hu *et al.* [4] mention that the housing made up of silicone rubber as the base material is usually added with other material additives to improve their material properties.

Over the last decades, due to the growing trend toward producing sustainable materials, the use of natural fibres as reinforcing elements in polymer composites has gained popularity to replaced synthetic fibre such as glass fibre [5]. Although natural fibre composites are found to have intermediate mechanical properties, they offer significant benefits over the glass and other synthetic fibres, including lower density, cheaper cost, renewable, recyclable and biodegradable [6]. Although glass fibre is the mainstream material of the reinforcing fibre in the insulator core, Macey *et al.* [7] mention that other insulating fibre can also be used. Many recent works are made in Malaysia to study the performance of kenaf fibre reinforced composites under different circumstances, including Osman *et al.* [8] who study the effect of chemical treatment on the mechanical properties of pultruded kenaf fibre reinforced polyester composites; Mahjoub *et al.* [9] who study the characteristics of continuous unidirectional kenaf fiber reinforced epoxy composites; Fairuz *et al.* [10] who study the effect of gelation and curing temperatures on mechanical properties of pultruded kenaf fibre reinforced vinyl ester composites; Fairuz *et al.* [11] who study the effect of pulling speed on mechanical properties of pultruded kenaf fibre reinforced vinyl ester composites; Zamri *et al.* [12] who study the effect of different kenaf fibre yarn tex of pultruded kenaf fibre reinforced composites. It was found that the mechanical properties of kenaf fibre reinforced polymer (FRP) vary due to different manufacturing processes and parameters. The tensile modulus of kenaf FRP is ranging from 10-18.15 GPa and the Poisson's ratio at 0.32, which may be applicable for certain load-bearing applications.

Zhao *et al.* [13] mention that the insulators while in operating is mainly subjected to three kinds of mechanical stresses, which are tensile stress, torsional stress and bending stress. The primary basic load, among the mechanical stresses, for a high voltage insulator is reported to be tensile load [14]. Tensile load is usually caused by the weight of the transmission lines galloping through the transmission towers. The wind condition is another factor affecting the tensile load subjected to the insulator by adding extra wind load on the transmission lines. The vertical load, horizontal load and the resultant total force can be calculated as following [7]:

$$F_v = (m_c + \pi \cdot b_i \cdot ({q_{ice}}^2 + q_{ice} \cdot d_{con})) \cdot S_m \tag{1}$$

$$F_h = P_w \cdot (d_c + 2q_i) \cdot s_f \cdot g_f \cdot S_w \tag{2}$$

$$F_t = \sqrt{{F_v}^2 + {F_h}^2} \tag{3}$$

Due to the strong tensile properties and light weight of the FRP material, the axial deformation of the mandrel is small and thus the transverse charges can be ignored. As a result, by considering the weight from the overhead power lines with addition to the wind load as maximum tensile loading, the tensile stress can be calculated with Hook's law [13].

$$E = \frac{F_t/A}{\Delta L/L} \tag{4}$$

Although a suspension type insulator is mainly subjected to tensile force, a torsional or a twisting type of load can be still experienced by the insulators during line construction. Papailiou and Schmuck [2], Macey *et al.* [7], and Zhao *et al.* [13] stated that the composite core rod is relatively sensitive to the torsional stress due to the longitudinal arrangement of glass fibres. Macey *et al.* [7] emphasize the importance of handling the components with care and twisting should be avoided especially during installation and conductor stringing. It is reported that the maximum torque that can be applied on the composite insulator core without permanent damage is 65 Nm for a 16 mm diameter core and 175 Nm for a 24 mm diameter core. The formula for torsional moment, torsional stress and angle of rotation are given as follow [13]:

$$M_t = F_h \cdot D \quad (5)$$

$$\tau_{max} = \frac{M_t \cdot R}{I_p} \quad (6)$$

$$\varphi = \frac{M_t \cdot l}{G \cdot I_p} \cdot \frac{180}{\pi} \quad (7)$$

Similar to torsional load, a suspension-type insulator is not mainly designed to bear bending load. However, Zhao *et al.* [13] state that the composite insulator will swing over a swing angle under the situation of the conductor galloping. Thus, a bending load may occur in the process. Composite insulators also experienced the bending load during improper handling or installing process. According to Zhao *et al.* [13], the large deflection theory should be considered during the calculation of insulator under large bending load. The large deflection equation at the free end is given as follow [13]:

$$\frac{\widehat{W_B}}{L} = \sin\theta_B - \frac{2{\theta_B}^2}{3(\frac{FL^2}{EI})^{\alpha}} \quad (8)$$

$$\theta_B = \left(\sqrt{\left(\frac{35}{12\frac{FL^2}{EI}}\right)^2 + \frac{35}{12}}\right) - \frac{35}{12\frac{FL^2}{EI}} \quad (9)$$

$$I = \frac{\pi \cdot D^4}{64} \quad (10)$$

$$\alpha = \begin{cases} 1 \quad , when \ \frac{FL^2}{EI} \leq 1 \\ 0.92, \ when \frac{FL^2}{EI} > 1 \end{cases} \quad (11)$$

2. Methodology

2.1. *Specifications of transmission system*

The transmission system with a voltage level of 132 kV is selected as the main research object in this study. The composite long rod insulator is installed in suspension position of which the top end is connected to the transmission tower and the bottom end is linked to the conductor by the end-fitting. Based on the calculation, the maximum tensile load that the composite required to bear is 6400 N. Therefore, by selecting a composite long rod insulator with a standard mechanical strength rating higher than the maximum load will provide adequate strength to withstand the load. The specifications of the selected insulator are listed in Table 1.

Table 1. Specifications of composite insulator

Insulator model	FXBW-132/70
Rated voltage (kV)	132
Specified mechanical load (kN)	70
Core diameter (mm)	16
Section length (mm)	1240

2.2. *Finite element analysis*

2.2.1. *Mechanical performance analysis*

The engineering data is used to define the material properties. The material properties of glass FRP, kenaf FRP, and plain epoxy rod are manually created and added to the data library as listed in Table 2.

Table 2. List of material properties

Material	**Young's Modulus (GPa)**	**Shear Modulus (GPa)**	**Poisson's Ratio**
Glass FRP	36.6	14.524	0.26
Kenaf FRP	18.15	6.875	0.32
Plain Epoxy	2.131	0.7665	0.39

A 3D finite element model of the insulator is established with ANSYS DesignModeler software. Since the analysis is done only on the insulator core, a cylinder is built to represent the core. The diameter of the cylinder is 16 mm and the length is 1240 mm. The 3D model is used in the simulation of tensile and torsion load. For bending load, due to the large deformation properties of the core, a 2D finite element model is built to simplify the model and limit the movement in the plane. The Ansys DesignModeler software is used to build the 2D model as well. The line body is generated with a length of 1240 mm along the X-axis and the cross section is set to be a round shape with a diameter of 16 mm. The 3D model is then meshed into total of 11526 nodes and 2178 elements while the 2D line body model is meshed as linear element and is separated into 40 parts. For the boundary condition of the model, the top end of the core is set as fixed support. The bottom end is applied with 6000 N of tensile load, 64000 Nmm of torsion load and 6400 N of bending load calculated from the theoretical analysis. The process is then repeated for all three materials.

2.2.2. *Electrical performance analysis*

The ANSYS Maxwell software is used to determine the voltage distribution and electrical stress of the insulator. As shown in Fig. 1, the finite element model of the core is built according to the dimension of 1240 mm in length, then an enclosure is added to the core to represent the surrounding environment. The core is assigned with the material of glass FRP, with a relative permittivity of 7.1

according to Netravati *et al.* [15]; and the surrounding environment is assigned as the built-in material of air with relative permittivity of 1.0006. The voltage excitation of 132 kV is set at the high voltage end of the core, and the ground end is set as 0 V. The model is then run and analyse for all results. The simulation is then repeated by changing the core material to kenaf FRP, with the relative permittivity of 4.2 adopted from Merizgui *et al.* [16].

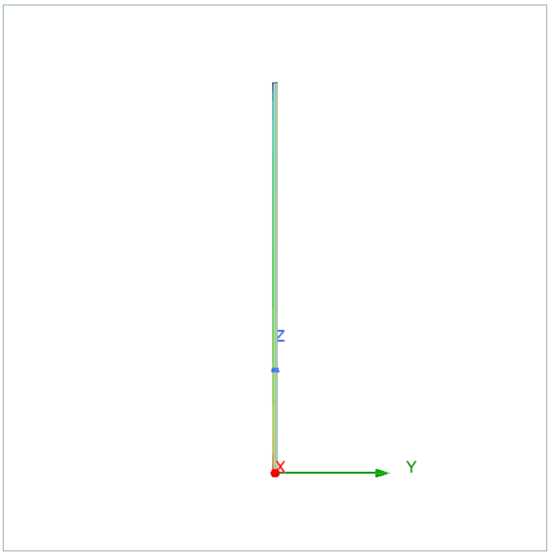


Fig.1 Insulator finite element model.

3. Results and Discussion

3.1. *Insulator performance under tensile load*

Fig. 2 shows the total deformation of the insulator core under the tensile load of 6400 N. The greatest deformation for all materials occurs at the end where the tensile load is applied and gradually decreases to the fixed end. The maximum deformation obtained from the glass FRP and kenaf FRP is 1.079 mm and 2.1756 mm respectively. On the other hand, 18.526 mm of deformation has been observed on the plain epoxy rod. The results showed that kenaf fiber reinforced epoxy composite has achieved around 7.5 times of improvement of its tensile performance compared to plain epoxy. However, due to the nature of having lower strength than synthetic fiber, the elongation of kenaf FRP has increased around 2 times compared to glass FRP.

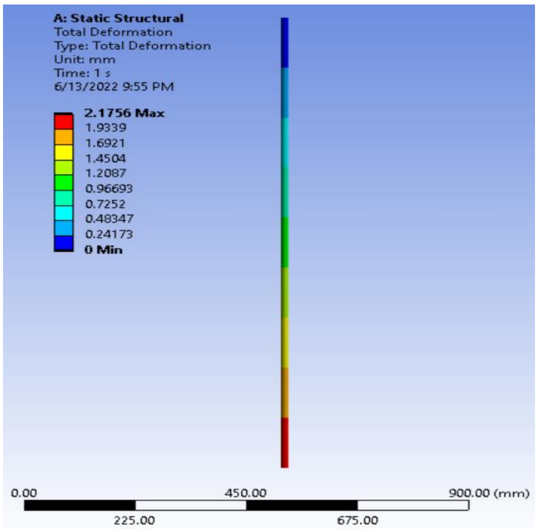


Fig.2 Total deformation of kenaf fibre/epoxy insulator under tensile load

As shown in Fig. 3, the major portion of the core is subjected to around 31 MPa of tensile stress according to the stress distribution contour. Nevertheless, it is remarkable that the maximum stress has been observed at the outer layer of the top end which is set as the fixed support. It indicates that in reality, stress concentration will be encountered at the top end and failures are most likely to happen at the area as well. The maximum tensile stress subjected to the cores is 34.133 MPa, 34.789 MPa and 35.825 MPa for Glass FRP, Kenaf FRP and plain epoxy rod respectively. It also proved that the kenaf FRP has better performance than the plain epoxy rod. The maximum stress of kenaf FRP experiences is only 21% of the tensile strength of the material and thus it will provide sufficient strength to withstand the load from the conductor as well as the wind load.

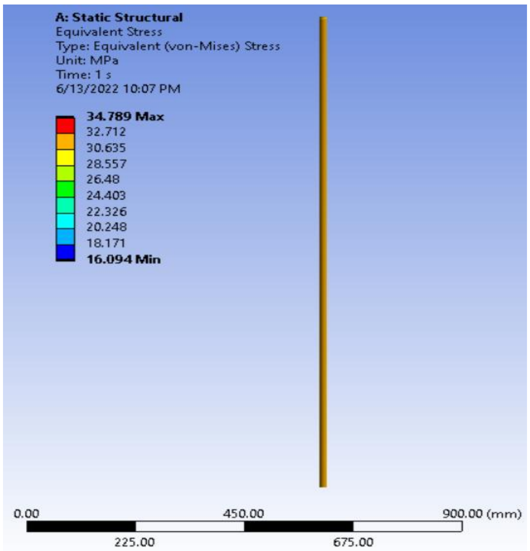


Fig.3 Induced equivalent stress on kenaf fibre/epoxy insulator under tensile load.

The theoretical tensile performance of the FRP core is calculated by applying Eq. (1), Eq. (2), Eq. (3) and Eq. (4) and listed in Table 3. Because the dimension and cross-sectional area of the FRP rod remained constant for all materials, the tensile stress is unaffected by material properties and remains constant at 31.83 MPa. The evaluated elongation of the glass FRP rod, kenaf FRP rod, and plain epoxy rod are 1.08 mm, 2.17 mm and 18.52 mm respectively. The highest deviation between the theoretical and FEA result for the elongation is 0.26 % from kenaf FRP. The error for the tensile stress from the glass FRP and kenaf FRP are 7.24 % and 9.30 % accordingly. Since the deviation between the result fall within 10 %, thus the results are considered authentic.

Table 3. Analysis results under tensile load

	Glass FRP	Kenaf FRP	Plain Epoxy
Deformation (mm)			
Theoretical	1.08	2.17	18.52
FEA	1.079	2.1756	18.526
Error	0.09 %	0.26 %	0.03 %
Tensile Stress (MPa)			
Theoretical	31.83	31.83	31.83
FEA	34.133	34.789	35.825
Error	7.24 %	9.30 %	12.55 %

3.2. *Insulator performance under torsion load*

The insulator core is subjected to 64000 Nmm of torsional load. Similar to the FEA result discussed for tensile stress, stress concentration also occurs when given a torsional load. Fig. 4 shows the shear stress contour plot experienced by the kenaf FRP subjected to torsional load. The outer shell of the rod is observed to have the highest torsional stress and the centre of the rod to have minimum stress. The maximum shear stress is 84.749 MPa, 84.682 MPa and 84.529 MPa for the glass FRP, kenaf FRP and plain epoxy respectively.

As shown in Fig. 5, the deformation of the rod is presented in millimetre along the radius of the rod, which can then be converted into degree. The FRP rod is predicted to have maximum deformation at the end where the torsional load is applied, and the fixed end experienced the least deformation. The angle of rotation is found to be 48.74°, 103° and 923.54° respectively. Despite having a 2 times higher angle of rotation compared to glass FRP, the deflection of kenaf FRP is still 24 % below the typical damage limits of 135° for the twisting of long rod insulators according to Macey *et al.* [7]. Meanwhile, the plain epoxy rod under the torsional load of 64000 Nmm is encountering 922° of rotation, which is almost 8 times greater than the kenaf FRP. It further proved that the kenaf fiber has significantly improved the capability of the rod to withstand a twisting type of load compared to a plain epoxy rod.

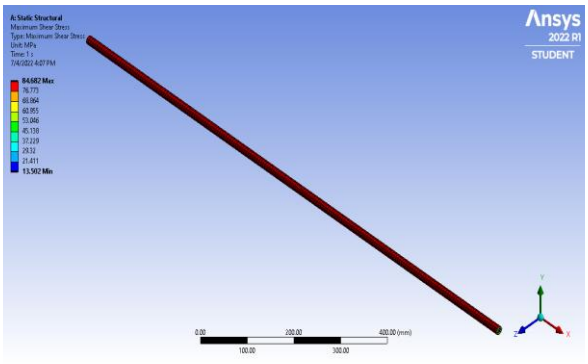


Fig. 4 Induced maximum shear stress on kenaf fibre/epoxy insulator under torsion load

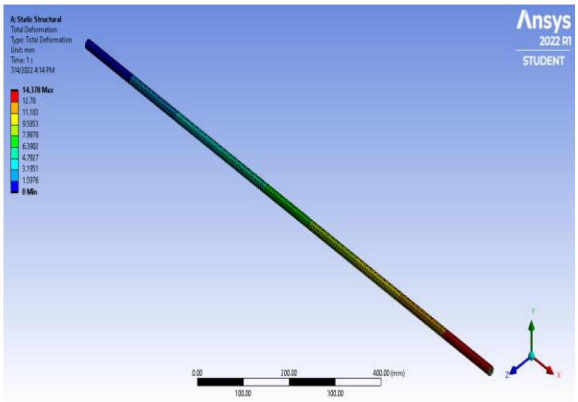


Fig. 5 Deformation on kenaf fibre/epoxy insulator under torsion load

By applying Eq. (5), Eq. (6) and Eq. (7), the theoretical results of the insulator core are listed out in Table 4. Under 64000 Nmm of torsional load, the rod will experience 79.58 MPa of torsional stress theoretically for all three rods since the stress is only affected by the dimension but not the shear strength of the material. The angle of rotation calculated for all three materials is 48.66°, 102.8° and 922° for glass FRP, kenaf FRP and plain epoxy accordingly. 6.50 % and 6.41 % of deviations are obtained from the torsional stress for glass FRP and kenaf FRP. For the angle of rotation, finite element analysis and theoretical analysis almost resulted the same and the highest error between them is 0.19% which is achieved from the kenaf FRP.

Table 4 Analysis Results under Torsion Load

	Glass FRP	Kenaf FRP	Plain Epoxy
Torsional Stress (MPa)			
Theoretical	79.58	79.58	79.58
FEA	84.749	84.682	84.529
Deviation	6.50 %	6.41 %	6.22 %
Angle of Rotation (Degree)			
Theoretical	48.66	102.8	922
FEA	48.74	103.0	923.54
Deviation	0.16 %	0.19 %	0.17 %

3.3. *Insulator performance under bending load*

In the bending simulation, 6400 N of bending load is applied to the bottom end of the core in the horizontal direction. Fig. 6 shown the deformation patterns of the rod under the large deflection theory. The rod is observed to be significantly bent in the horizontal direction. The bottom end of the rod experiences the greatest deformation and gradually decreases to the fixed end. The result of the deformation is determined as the horizontal directional deformation, which has been recorded as 1161.9 mm, 1186.5 mm and 1240.5 mm accordingly for the three materials.

Fig. 7 shows the stress distribution of the rod under the bending load with the large deformation theory applied. The direct stress, which represent the stress component due to the axial load encountered in the model, is obtained from the simulation. The maximum stress distribution occurs along the free end of the rod, which is obtained to be 31.844 MPa, 31.912 MPa and 32.343 MPa accordingly. The results show that the kenaf FRP experiences slightly higher bending stress compared to the glass FRP. However, the maximum stress of 31.912 MPa is only 19% of its strength of 164 MPa, which means that the load is unlikely to cause a fracture to the kenaf FRP core and it will not break immediately.

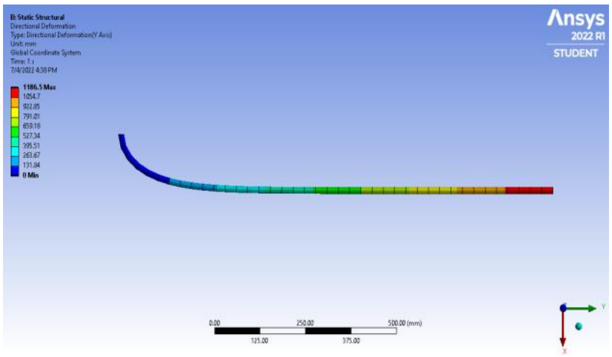


Fig. 6 Y-direction deformation for kenaf fibre/epoxy insulator under bending load

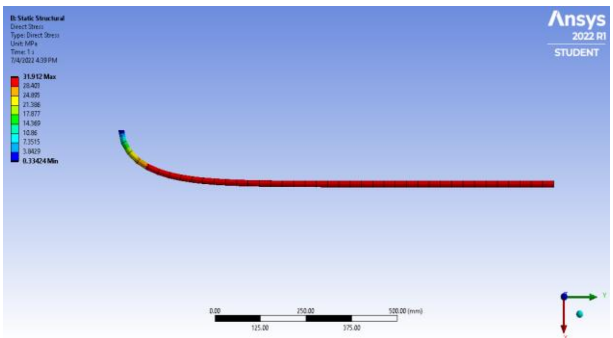
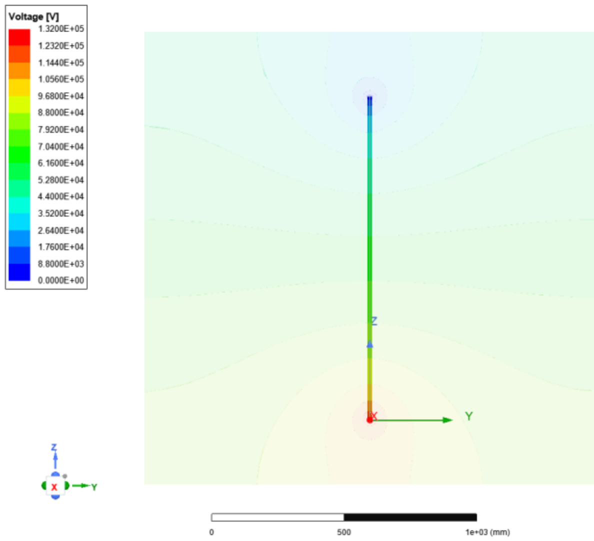


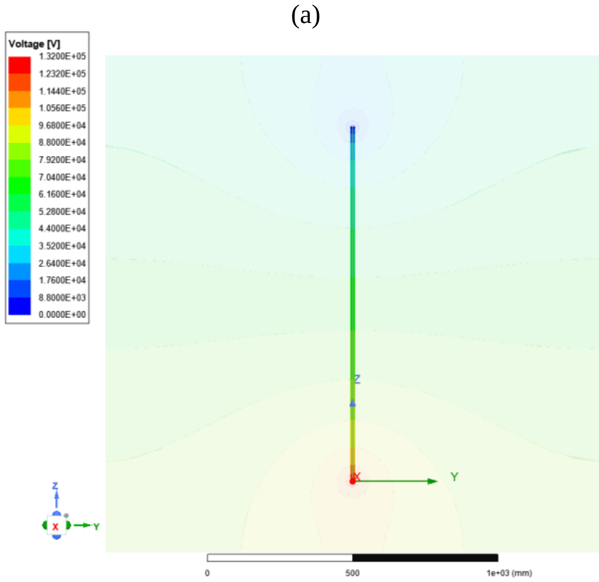
Fig. 7 Direct stress induced on kenaf fibre/epoxy insulator under bending load

3.4. *Insulator performance under electrical load*

The voltage distribution pattern in the rod is shown in Fig. 8 and it is observed that both glass FRP and kenaf FRP rods have uniform voltage distribution along the rods. The maximum voltage occurs at the high voltage end with 132 kV and keeps decreasing towards the ground end. As shown in Fig. 9, the voltage distribution at the surrounding of the rod is also affected by the high voltage load from the insulator. High potential voltage is observed at a certain area around the high voltage end, and it gradually decreased along the direction to the ground end. The distribution contours show that the kenaf FRP has a smaller area of the voltage spread to the surroundings compared to the glass FRP, which means that it is subjected to higher electrical stress inside the core.

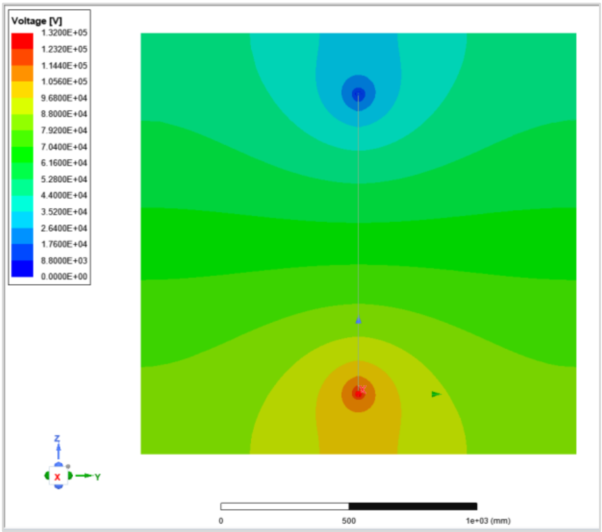


(a)

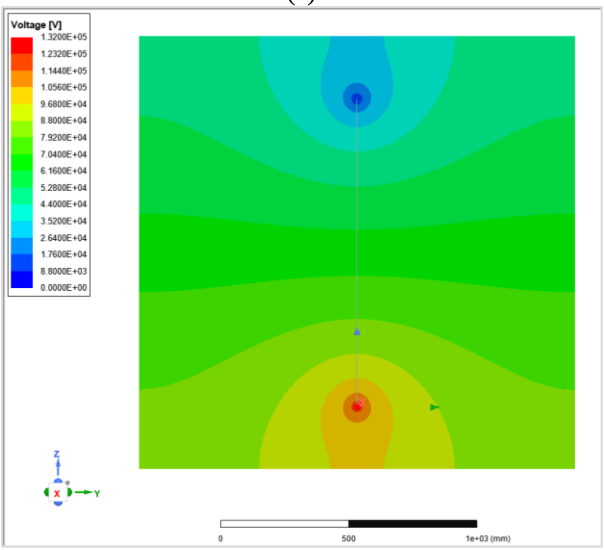


(b)

Fig. 8 Voltage distribution in the (a) glass fibre-epoxy insulator and (b) kenaf fibre-epoxy insulator



(a)



(b)

Fig. 9 Potential voltage distribution of the surrounding (a) glass fibre-epoxy insulator and (b) kenaf fibre-epoxy insulator

The voltage distribution along the distance for both materials is plotted on a graph in Fig. 10. According to the result, the kenaf FRP rod have slightly higher amount of voltage drop from both ends of the rod compared to glass FRP and drops smoothly at the middle range of the core. Fig. 11 represents the electrical stress plotted along the length of the rod. It can be seen that both materials have high similarity in the patterns since the glass FRP and kenaf FRP are both considered to have low conductivity. The greatest electrical stress obtained from the glass FRP is 2.55 kV/mm and kenaf FRP is 2.95 kV/mm. The result shows that the kenaf FRP will experience 13.6 % higher electrical stress compared to glass FRP. The result shows that the electrical performance of kenaf FRP is almost comparable to the glass FRP.

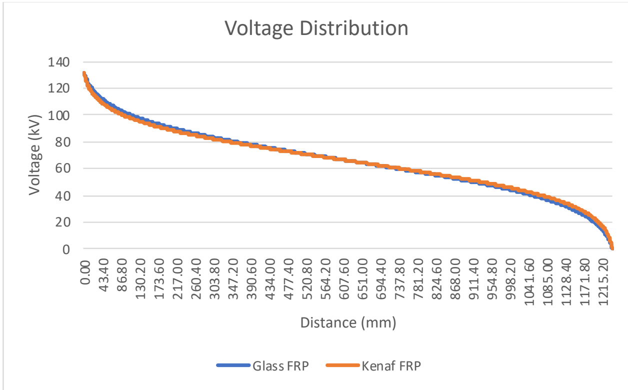


Fig. 10 Voltage distribution across insulator

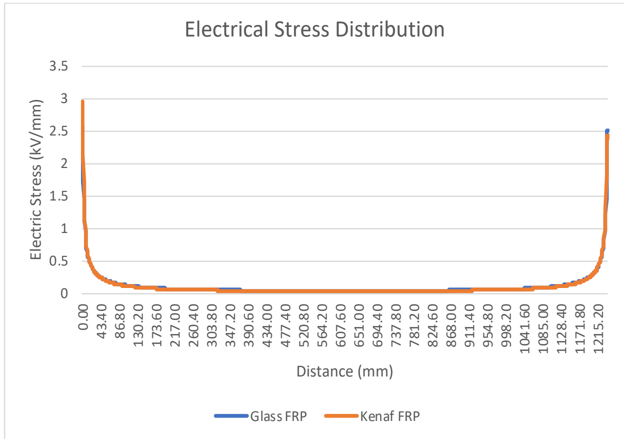


Fig. 11 Electrical stress distribution across insulator

4. Conclusion

In this analysis, the performance of biocomposites that is kenaf fiber reinforced epoxy composites has been evaluated. It was found that the biocomposite is capable to withstand different types of mechanical loads on the high voltage insulator due to the weight from the transmission conductor and wind loading. Under tensile load, although the elongation of kenaf FRP has increased around 2 times compared to glass FRP, it only experiences 34.789 MPa of maximum tensile stress which equals 21% of the tensile strength of the material. It shows that the kenaf FRP will not be fractured immediately by the given tensile load. Under torsion load, despite having a 2 times higher angle of rotation compared to glass FRP, the deflection of kenaf FRP at 103° is still 24 % below the typical damage limits of 135° for the twisting of a long rod insulator. Under bending load, even the kenaf FRP experiences slightly higher stress compared to the glass FRP, the maximum stress of 31.912 MPa is only 19% of its strength of 164 MPa, which means that the load is unlikely to cause a fracture to the kenaf FRP core and it will not break immediately. From the electrical simulation, the kenaf FRP is having similar behavior as the glass FRP under electrical load. Although the greatest electrical stress obtained from the kenaf FRP is 2.95 kV/mm, which is 13.6 % higher than glass FRP, the result shows that the electrical performance of kenaf FRP is almost comparable to the glass FRP. Even though the kenaf FRP is naturally having lower strength, the performance of kenaf FRP is almost comparable to glass FRP and the addition of kenaf fiber to the epoxy matrix has significantly improved the mechanical performance of plain epoxy.

Acknowledgements

The authors would like to express their gratitude towards the engineering faculty of UCSI University.

References

1. Cherney E.A. (2012) High voltage insulator mechanical load limitations to dielectric material damage. PES T&D 2012: 1-8.
2. Papailiou KO and F. Schmuck F (2013) Composite long rod insulators. In Silicone composite insulators, power systems. Springer-Verlag Berlin, Heidelberg.
3. Gorur RS (2005) High-voltage transmission. In Chen WK (ed) The electrical engineering handbook, Academic Press.
4. Hu Y and Liu K (2017) Transmission lines detection technology. In Hu Y and Liu K (eds) Inspection and monitoring technologies of transmission lines with remote sensing, Academic Press.
5. Ramesh M and Sudharsan P (2018) Experimental investigation of mechanical and morphological properties of flax-glass fiber reinforced hybrid composite using finite element analysis. Silicon 10:747–757.
6. Nimanpure S, Hashmi S, Kumar R, Bhargaw H, Kumar R, Nair P and Naik A (2019) Mechanical, electrical, and thermal analysis of sisal fibril/kenaf fiber hybrid polyester composites. Polym. Compos. 40: 664-676.
7. Macey RE, Vosloo WL and Tourreil C (2004) The practical guide to outdoor high voltage insulators. Crown Publications, Johannesburg.
8. Osman MR, Mazuki AA., Akil HM, Ishak ZA and Bakar AA (2013) Effect of Chemical Treatment on the Mechanical Properties of Pultruded Kenaf Fibre Reinforced Polyester Composites. Key Engineering Materials, 594-595: 691 - 695.
9. Mahjoub R, Yatim JM, Mohd Sam AR and Raftari M (2014) Characteristics of continuous unidirectional kenaf fiber reinforced epoxy composites. Materials and Design, 640-649.
10. Fairuz AM, Sapuan SM, Zainudin ES and Jaafar CNA (2015) The effect of gelation and curing temperatures on mechanical properties of pultruded kenaf fibre reinforced vinyl ester composites. Fibers Polym 16:2645–2651.
11. Fairuz AM, Sapuan SM, Zainudin ES and Jaafar CNA (2018), The effect of pulling speed on mechanical properties of pultruded kenaf fiber reinforced vinyl ester composites. J Vinyl Addit Technol, 24: E13-E20.
12. Zamri MH, Akil HM and MohdIshak Z (2016). Pultruded Kenaf Fibre Reinforced Composites: Effect of Different Kenaf Fibre Yarn Tex. Procedia Chemistry. 19:577-585.
13. Zhao X, Fu Z, Zhao M, Wang L and Wu Y (2015) Mechanics analysis of 110kV suspension-type composite insulators in ice coating and wind loads conditions. In: Proceedings of the 2015 international symposium on material, energy and environment engineering, Advances in engineering research. Atlantis Press.
14. External insulation characteristics of UHVDC lines (2018) UHV transmission technology. Academic Press.
15. Netravati, N, Sunitha, N, Shivakumara Aradhya, R and Ravi, K. (2016). Risk assessment of pollution on a 66kV composite insulator using ANSYS. International Journal of Science & Engineering Development Research, 1(11): 6-12.
16. Merizgui, T, Hadjadj, A, Kious, M, Prakash, V and Gaoui, B. (2019). Effect of magnetic iron(III) oxide particle addition with MWCNTs in kenaf fibre-reinforced epoxy composite shielding material in ‘E’, ‘F’, ‘I’ and ‘J’ band microwave frequencies. Materia

Authors Introduction

Mr. Tan Kang Rui

He received his Bachelor’s degree in Mechanical Engineering in 2022 from the Faculty of Engineering, Technology & Built Environment, UCSI University, Malaysia.

Dr. Cik Suhana Hassan

She received her bachelor's and master's degrees in 2009 and 2011, respectively, from Universiti Teknologi PETRONAS, and her PhD in 2019 from Universiti Putra Malaysia. She is currently an Assistant Professor at the Department of Mechanical and Mechatronics Engineering of UCSI University. Her research interests include the investigation of bio-composites characteristics for applications, particularly automotive.

Ms. Nor Fazilah Abdullah

She received her Bachelor’s degree in aerospace engineering (Hons) from IIUM, Gombak in 2010 and her Master’s degree in Mechanical Engineering from UKM, Bangi in 2015. Currently she is pursuing Doctoral of Philosophy programme at UCSI University, Kuala Lumpur. Her research interest in bio-based nanoparticles materials.

Dr. Farah Adilah Jamaludin

She received her PhD degree in Electrical Power Engineering from Universiti Putra Malaysia in 2020. She currently working as Assistant Professor at Electrical & Electronics Engineering Department, Faculty of Engineering, Technology & Built Environment, UCSI University since 2018. Her specialization is in high voltage insulation coordination, power system, and energy management system.

Dr. Chiong Meng Choung

He received his bachelor's degree from Swinburne University in 2008, master's degrees from Universiti Teknologi Malaysia in 2014, and PhD in 2019 from Universiti Teknologi Malaysia. He is currently an Assistant Professor at the Department of Mechanical and Mechatronics Engineering of UCSI University. His research is mainly on alternative fuel combustion

Ms. Eryana Eiyda Hussin

She received her Master's degree from the Faculty of Electronics and Computer System, Universiti Teknikal Malaysia Melaka. She is currently persuing her study under Doctor of Philisophy in Electrical and Electronic Engineering in Universiti Teknologi Petronas, Malaysia. She is also a lecturer from the Department of Electrical and Electronics of UCSI University, Malaysia.

Detection of Bullet Holes for Target Board in Malaysia Military (ATM) Shooting Exam Application

Jillian Hah Wai Yin, **Noor Idayu Mohd Tahir, Ammar A.M. Al-Talib**, **Osama Mohamed Magzoub**
Department of Mechanical and Mechatronics, Faculty of Engineering, Technology and Built Environment, UCSI University, 56000 Kuala Lumpur, Malaysia

Email: osam937@gmail.com, idayuketco@gmail.com, ammart@ucsiuniversity.edu.my

Abstract

This study focuses on designing and developing a bullet hole detection system for target boards in the Malaysia Army (ATM) shooting exercise environment. The deep learning algorithm is based on YOLO models, utilizing Raspberry Pi and IoT via Blynk for remote monitoring. The prototype includes a Raspberry Pi 4b, HQ Camera Module Lens, 35mm Telephoto Lens, and tripod stand, all at an affordable cost. The study demonstrates that the bullet hole detection system is accurate and effective for ATM shooting exams, meeting SDG 3, SDG 9, SDG 11, and SDG 12 goals.

Keywords: Bullet Hole Detection, Military Shooting Exams, Internet of Things (IoT), Machine Learning, YOLOv8

1. Introduction

The Malaysian Military (ATM) shooting exam assesses shooter proficiency by scoring shots on target boards under various conditions. Traditionally, this process has relied on manual review, requiring examiners to inspect target boards and tally scores after each session. This method is prone to human error, inconsistency, and delays. This project proposes a detection system to automate bullet hole identification and scoring to address these limitations [1].

Current automated systems, such as electronic target surfaces and photoelectric technologies, have limitations like high power consumption and reduced mobility. Advances in image processing now allow for reliable and real-time bullet hole detection. Using Raspberry Pi 4b, this project leverages YOLO models for object detection and image processing. Bullet holes are detected and processed in real-time, with data transmitted wirelessly to a dashboard for visualization.

By transmitting target board images with bullet hole locations over wired or wireless connections to a distant computer or display device, the bullet hole detection system significantly improves the convenience of the reviewing process. This enables examiners to assess target boards from a secure and comfortable location. For instance, [1] [2] described a system that transmitted target board images to a smartphone using image processing and wireless connectivity, allowing examiners to review data remotely.

The proposed prototype offers prompt and comprehensive feedback by presenting the overall bullet hole locations on a dashboard. It displays information such as the distribution of bullet holes across scoring zones, the holes' size and position, and the shots' precision and consistency. Additionally, the system provides advice and recommendations to enhance shooting performance, enabling shooters to track and evaluate their marksmanship improvements. Detection systems like these have been shown to deliver graphical or numerical feedback on display monitors, assisting shooters in identifying their strengths and weaknesses [3].

The technology analyses target board pictures using powerful algorithms, delivering precise, real-time feedback and decreasing scoring fluctuation. Using Raspberry Pi technology and YOLO deep learning models, the system provides efficient, accurate, and dependable detection. The study aims to develop a prototype integrating hardware like cameras and dashboards, evaluate the system's detection accuracy, and analyse AI's ability to recognize bullet holes and assign graphical representations based on shot locations. This innovation not only enhances objectivity and efficiency but also promotes fairness in the ATM shooting exam, ultimately modernizing military training processes [4].

2. Methodology and Experimental Setup

This project was designed to detect the bullet holes for the target board in the Malaysia Military (ATM) shooting exam application. Currently, the project has been improved to work better. Using Raspberry Pi 4 Model B to run several motors and sensors as shown in Table 1. The new design will smooth out feeding by image processing.

Table 1. System Components.

Components	Quantity
Raspberry Pi 4 Model B	1
Raspberry Pi High-Quality Camera Module Lens	1
Raspberry Pi 35mm Telephoto Lens (C Mount)	1
Tripod Stand	1
USB B Type Cable	1
1KG 1.75mm ABS Filament (Black)	1

The system is activated when the detection system is initialized, with the Raspberry Pi hardware booting up and preparing for operation. The Raspberry Pi establishes communication with the Blynk app, which serves to visualize the entire system for the user.

Once the system is configured, the YOLOv8 model is loaded, and the camera starts broadcasting. The system continually collects photos of the target board, either every second or based on human input. If a target is identified, the system determines whether a target board is visible in the picture [5] [6].

Collecting the dataset was the first step in the project. the researcher Handling this work, by setting up a good environment for data collection and creating the troughs using a 3D printer. To get a large enough number of images as shown in Table 2, pictures of the troughs were shot empty, partially empty, and full.

A total of 1069 images were collected via the internet and using the camera on the iPhone 12 Pro Max with an image size of 3024 × 4032. The breakdown number of images for each class is shown in Table 2. The images of objects were captured with conditions such as lighting variations and reused target boards with stickers and captured images of the target boards from various distances.

Table 2. Number of Images for Each Class.

Sources	Number of Images
Dashboards	332
Cars	274
Wood	134
Glass	164
Walls	165
Total	1069

The images of the object were captured with different orientations and angles. This can enhance the generalization of the model to understand the structure and appearance of the object so that it can identify the object with a different viewpoint and orientation in the real world. Table 3 shows the example images of the three classes for this research.

Table 3. Sample Images of the troughs after annotation.

	Sample images with different angels
Bluet Hols.	
Military scoring board	

Once the images are gathered, the images are imported into Roboflow for data annotations, data pre-processing, and data augmentation. The dataset first undergoes the data annotation process whereby the object in the image is labeled. Then, data pre-processing methods were used to reduce the size of the image to 640 × 640 and auto-orient the annotation when there were changes to the image. The size of the dataset is 550 images only. Increasing the size of the training dataset can increase the performance of a deep-learning model. However, a large dataset requires time to capture images, label images, and prepare the dataset. Therefore, data augmentation techniques are used to artificially increase the size of the dataset by creating a modified version of the existing dataset. The dataset size for each class after data augmentation. The size of the dataset after data augmentation is 8453 images. Lastly, the dataset was split into the train set, validation set, and test set with a train:val: test ratio of 87:13:0 to be 792:117:0 of images. The dataset is converted into YOLOv8 PyTorch format for use in the training process [6] [7].

3. Results and Discussion

3.1. *Prototype*

The prototype was created using the items listed above, and Fig. 1 depicts the physical prototype. Provides an overview of the full prototype, including the Raspberry Pi, camera, and tripod, as well as the 3D-printed enclosure built for protection and stability. The 3D-printed shell secures the Raspberry Pi and camera module, protecting it from the elements and making it simple to install on the tripod. The tripod improves the stability and placement precision of the camera, which is critical for collecting exact photos of the target board. The finished prototype flawlessly incorporates all components, exhibiting the capability and potential of the bullet hole detecting system.

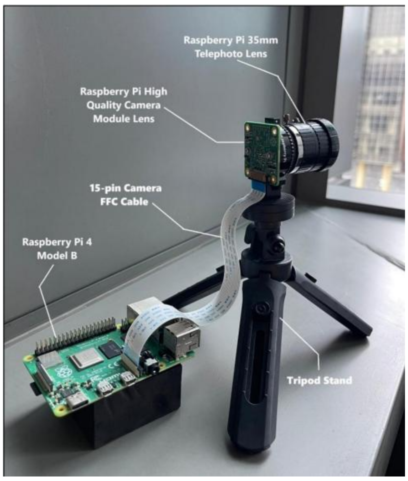


Fig. 1. Robot Farm.

The training results for the YOLOv8n model, illustrated in Fig. 2, show significant improvements in performance metrics. The train/box loss graph indicates a steady decrease from around 0.8 to 0.2 over 100 epochs, signifying enhanced accuracy in predicting trough locations. The train/classification loss reduced substantially from approximately 2.0 to nearly 0.0, highlighting the model's effectiveness in distinguishing between empty, partially empty, and full troughs, leading to high classification accuracy. Additionally, the train/distribution focal loss (DFL) decreased from about 1.3 to below 1.0, reflecting the model's increased precision in locating the troughs. Validation metrics, including Val/box loss and Val/classification loss, also improved, suggesting that the model generalizes well to new data, minimizing overfitting. Precision and recall metrics provide further insights, with high precision close to 1.0 indicating a low false positive rate and recall values approaching 1.0 demonstrating effective capture of all relevant trough statuses. The mean Average Precision at IoU=0.50 (mAP50) and mAP50-95 remained high, with mAP50 near 1.0 and mAP50-95 stabilizing around 0.9, confirming the model's robustness and accuracy across various detection challenges.

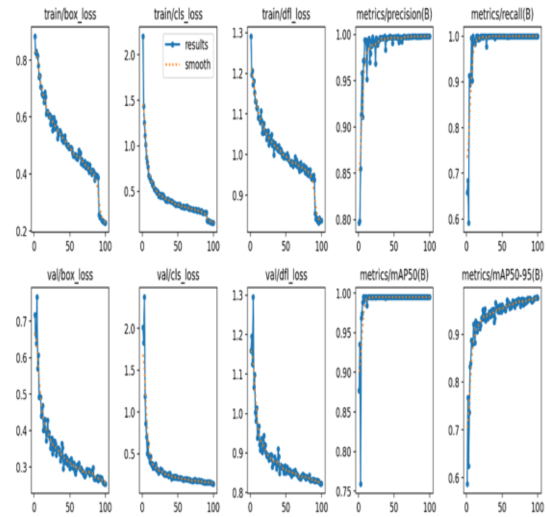


Fig. 2. Training and Validation Metrics for YOLOv8n Model on Trough Status Classification.

The comparison between YOLOv5 and YOLOv8, as illustrated in Fig. 3 and Fig. 4, highlights significant differences in their performance for bullet-hole detection. YOLOv5 achieves high precision (85.1%) and recall (80%), with a True Positive (TP) rate of 0.80 for bullet holes and a False Positive (FP) rate of 0.14. Its mean Average Precision (mAP) scores are moderate, with mAP@50 at 42.3% and mAP@50:95 at 29% or visualization.

In contrast, YOLOv8 demonstrates improved overall detection capabilities. It maintains similar precision (86%) and recall (80%) but achieves slightly reduced FP rates (0.13 for bullet holes) and higher mAP scores, with mAP@50 at 44.1% and mAP@50:95 at 32%. The normalized confusion matrix of YOLOv8 reflects better differentiation of bullet-holes, fewer misclassifications, and enhanced true positive rates.

As shown in the figure, YOLOv8's advancements make it the superior model for bullet-hole detection, with better learning, generalization, and performance across varying IoU thresholds compared to YOLOv5.

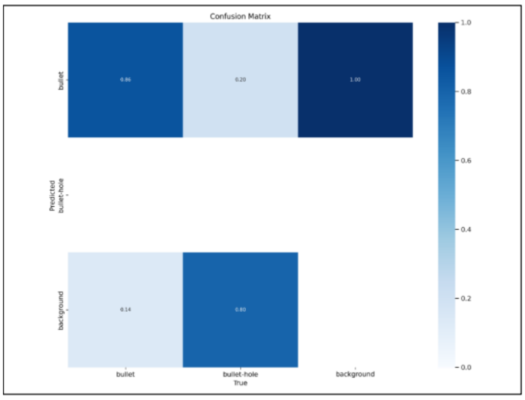


Fig. 3. Yolov5 Confusion Matrix of the trained model.

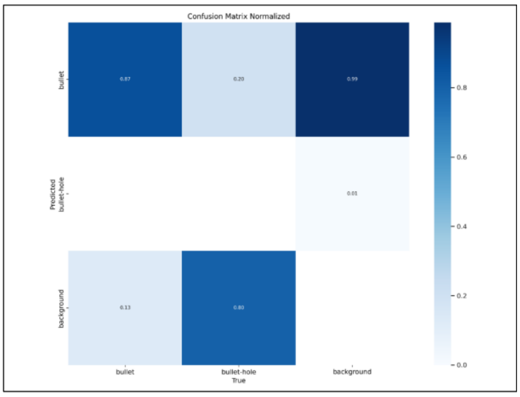


Fig. 4. Yolov8 Confusion Matrix Normalized of the trained model.

Fig. 5 shows the setup from the military base, which was crucial for gathering real-world data and insights necessary for developing and refining the detection of bullet holes system. The primary goal of the site visit was to collect high-quality images of target boards used in actual military shooting exercises. These images are vital for training the YOLOv8 model to accurately detect bullet holes. Additionally, the visit allowed for an environmental assessment, and an understanding of the conditions under which the shooting exercises take place. Observations included lighting variations, weather conditions, and the distances between shooters and target boards, which were essential for designing a strong system.

During the site visit, significant observations about traditional methods were made. In the traditional setup, around 120 participants undergo shooting exams over three days. This method not only leads to considerable bullet wastage but also poses safety risks as personnel are required to collect data from the target boards, sometimes while gunfire is ongoing. The manual collection and scoring of targets are time-consuming and prone to errors, impacting the efficiency and safety of the examination

Fig. 5. An example picture of the setup from the military base.

process. Furthermore, the site visit provided an opportunity to validate the prototype system in a real-world setting. This involved setting up the Raspberry Pi with a high-quality camera module and testing its functionality in capturing and processing images of the target boards. Direct interaction with military personnel also facilitated valuable feedback on the system's usability and accuracy, helping to identify areas for improvement to ensure the system meets practical needs. Fig. 6 shows the example system logbook of bullet holes detected. The system logbook can be used to track the current and past events of the system. In this example, 4 bullets were detected.

Fig. 6. Bullet holes system logbook.

Fig. 7 shows that although the YOLOv8 can detect bullet holes on the dashboard, it is not mature enough to detect all bullet holes consistently, and some may still be missed. Additionally, due to limitations in the Blynk app, the view of the live stream results is not stable. This instability can affect the reliability of real-time detection and needs to be addressed in future improvements. Enhancing the YOLOv8 model's accuracy and resolving the live stream issues in the Blynk app will be crucial steps in refining the system for practical use.

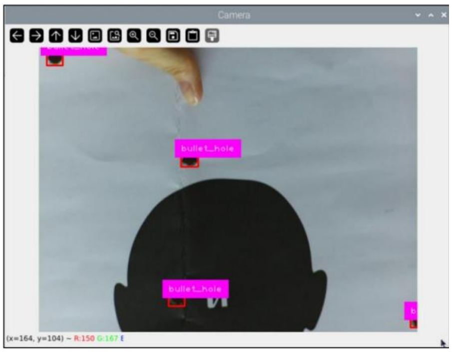


Fig. 7. Bullet holes system logbook.

The system leverages the Blynk app to visualize and manage the bullet-hole detection process, providing an interactive platform for user input and output. However, the app presents certain limitations, particularly in handling video live streaming. The quality and stability of the live video stream depend heavily on the available internet bandwidth. In environments with limited or unstable connections, the stream may experience lags, buffering, or dropped frames, potentially causing delays and inaccuracies in real-time monitoring. These challenges highlight the importance of considering network stability in deploying such systems.

4. Discussion

The project developed an automated system to monitor and manage goat feeding troughs, effectively addressing traditional method issues. Integrating a YOLOv8 model with hardware such as the Raspberry Pi 4, ultrasonic sensors, and a camera module accurately identified troughs statuses (bullet holes), significantly reducing the need for manual checks through image processing and machine learning. Key steps included collecting and enhancing a large set of trough images, using a deep learning model for object detection, and implementing real-time monitoring with the Blynk app, resulting in timely alerts for farmers. The model demonstrated high precision, recall, and F1 scores, with confusion matrices confirming accurate classification and performance curves indicating reliability. The project's success, driven by careful planning, effective methodology, and advanced technology, reduces farmers' workload, enhances resource management and goat care, and sets the stage for further innovations in automated farming, illustrating the potential of AI and IoT to improve military shooting exams.

5. Conclusion

In conclusion, the project successfully developed a bullet-hole detection system that addresses the limitations of traditional manual review methods, which often suffer from human error, subjectivity, and inconsistencies. Utilizing advanced technologies such as YOLOv8 for object detection, image processing algorithms, IoT integration via the Blynk app, and the Raspberry Pi platform, the system delivers high accuracy and real-time feedback. This enhances the fairness and efficiency of shooting exams and improves military training programs by providing instant feedback, allowing shooters to adjust and refine their skills more effectively.

The system also offers significant economic and safety benefits, reducing ammunition wastage, operational costs, and the risks associated with manual scoring near active shooting ranges. By automating data collection, it minimizes the need for human intervention, improving safety and sustainability.

Additionally, the project aligns with several United Nations Sustainable Development Goals (SDGs). It contributes to SDG 3 (Good Health and Well-being) by enhancing safety in shooting activities, SDG 9 (Industry, Innovation, and Infrastructure) by leveraging advanced technologies, SDG 11 (Sustainable Cities and Communities) through improved safety and resource efficiency, and SDG 12 (Responsible Consumption and Production) by reducing ammunition waste. With ongoing refinements and validation, the system is poised for broader application in military and sports contexts, emphasizing sustainability, scalability, and minimal environmental impact.

References

1. S. Dhamad, Automation of Real-time Target Scoring System Based on Image Processing Technique, 2021, [Online], Available: https://www.researchgate.net/publication/356593652_Automation_of_Real-time_Target_Scoring_System_Based_on_Image_Processing_Technique
2. A. R. Syafeeza, M. K. Mohd Fitri Alif, Y. Nursyifaa Athirah, A. S. Jaafar, A. H. Norihan, and M. S. Saleha, IoT based facial recognition door access control home security system using raspberry pi, International Journal of Power Electronics and Drive Systems, Vol. 11, No. 1, pp. 417–424, Mar. 2020, https://ijpeds.iaescore.com/index.php/IJPEDS/article/view/20490/0
3. S. K. Sanapannavar, C. M. Lakshmanagowda, and G. Sundararajan, A deep learning-based surveillance system for enhancing public safety through internet of things and digital technology using Raspberry Pi, *International Journal of Electrical and Computer Engineering*, Vol. 14, No. 6, pp. 7198–7210, Dec. 2024, https://ijece.iaescore.com/index.php/IJECE/article/view/36588
4. J. Terven and D. Cordova-Esparza, A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS, Apr. 2023, https://www.mdpi.com/2504-4990/5/4/83
5. A. John and D. Meva, A Comparative Study of Various Object Detection Algorithms and Performance Analysis, *International Journal of Computer Sciences and Engineering Open Access Research Paper*, Vol. 8, No. 10, 2020, https://www.researchgate.net/publication/346346964_A_Comparative_Study_of_Various_Object_Detection_Algorithms_and_Performance_Analysis
6. Intelligent Target-scoring System based on image processing Applicant: Rongxin Du Intelligent Target scoring System based on image processing, https://www.columbia.edu/~rd2537/docu/Intelligent%20Target-scoring.pdf
7. P. Shrivastava, V. K. Jain, and S. Nagpal, Gunshot residue detection technologies—a review, Dec. 01, 2021, *Springer Science and Business Media Deutschland GmbH*, https://ejfs.springeropen.com/articles/10.1186/s41935-021-00223-9

Authors Introduction

ENG. Jillian Hah Wai Yin

She received his Bachelor's degree in Mechatronics Engineering with Honors in 2024 from the Faculty of Engineering, UCSI University in Kuala Lumpur, Malaysia.

IR NOOR IDAYU BINTI MOHD TAHIR

Noor Idayu Mohd Tahir graduated from University Teknologi MARA, Malaysia in Mechanical Engineering. She was design engineer in the oil and gas industries for 7 years. Currently, she works as a lecturer under the Department of Mechanical and Mechatronics, Faculty of Technology and Built Environment, UCSI Kuala Lumpur. Her research interests include design and autonomous robotics.

Ammar Abdulaziz Al Talib

He received his B.Sc and M.Sc degrees in Mechanical Engineering from the University of Mosul Iraq. He has finished his PhD degree from UPM University, Malaysia. Member of the Institute of Mechanical Engineers UK (CEng. MIMechE). He has developed all the Postgraduate Programs at the Faculty of Engineering at UCSI University / Malaysia and worked as the Head of Postgraduate and Research department at the same faculty for the years 2010-2018.

ENG. Osama Mohamed Magzoub

He received his Bachelor's degree in Mechatronics Engineering with Honors in 2024 from the Faculty of Engineering, UCSI University in Kuala Lumpur, Malaysia.

Mobile App Development for Monitoring Goat Activities

Samy M.S. Elmasri, Noor Idayu Mohd Tahir, Ammar A.M. Al-Talib
Department of Mechanical and Mechatronics, Faculty of Engineering, Technology and Built Environment, UCSI University, 56000 Kuala Lumpur, Malaysia

Email: Samy.elmasri454@gmail.com, idayuketco@gmail.com, ammart@ucsiuniversity.edu.my

Abstract

"Mobile App Development for Monitoring Goat Activities" aims to create an automated and efficient system for managing goat feeding and water consumption. The primary objectives are to build a prototype using a camera and sensors for data collection, analyze a data logging system to record feeding habits and water consumption, and develop a mobile application to allow farmers to monitor the feeding system in real-time. The prototype, constructed using a Raspberry Pi 4 equipped with a camera module and ultrasonic sensors, collects real-time data on the status of feeding troughs, detecting whether they are empty, partially empty, or full. A YOLOv8n deep learning model was trained on a comprehensive dataset of images to classify these statuses accurately. The data logging system records and analyzes feeding habits and water consumption patterns, providing valuable insights into the goats' behavior and ensuring they have consistent access to food and water. The mobile application developed for this project, integrated with the Blynk app, provides farmers with real-time notifications and detailed reports on the feeding system's status. This allows for timely interventions, better resource management, and improved animal welfare. The project successfully combines hardware and software solutions to enhance the efficiency of goat farming, reducing manual labor and setting a foundation for future innovations in automated livestock management. By demonstrating the potential of integrating AI and IoT technologies in agriculture, this project highlights the benefits of continuous monitoring and timely alerts in maintaining a well-managed and sustainable farming operation.

Keywords: Automated goat feeding system, Machine Learning, Sustainable farming, YOLOv8, AI in agriculture

1. Introduction

Precision Livestock Farming (PLF) revolutionizes goat feeding management by integrating advanced technology for enhanced efficiency and sustainability. Utilizing computer vision and strategically placed cameras, our system accurately captures and logs goat feeding behavior, eliminating errors from manual observations. Smart sensors monitor water consumption, feed disappearance, and environmental factors, providing comprehensive data. This empowers farmers with actionable insights to improve animal welfare and productivity. By combining cutting-edge technology with an intuitive mobile app, this project aims to create a sustainable, data-driven approach to goat farming, transforming traditional practices into precise and informed management strategies [1] [3]. To dynamics analysis with moving bodies, the selection of the numerical integration method is crucial for the realization of the actual dynamics occurring in the real world. It cannot be solved by a simple way to chip the time step of the integration in the explicit numerical method. In the implicit numerical integration, it will help to refine the time step adaptively [2].

Humans first domesticated sheep and goats, with goats being domesticated between 6000 and 7000 BC and sheep even earlier, between 11000 and 9000 BC [4]. Both species are crucial to livestock, providing food such as meat and milk, fleece, and other products. Economically, they play significant roles in agriculture and biomedical research, with a global population of 2.2 billion. The global consumption of sheep meat is around 2.5 kg per person annually, out of a total meat consumption of 41.6 kg per person [5]. According to the Food and Agriculture Organization (FAO) statistics (2018) [6], the population of sheep and goats is increasing, as illustrated in Fig. 1 and Fig. 2.

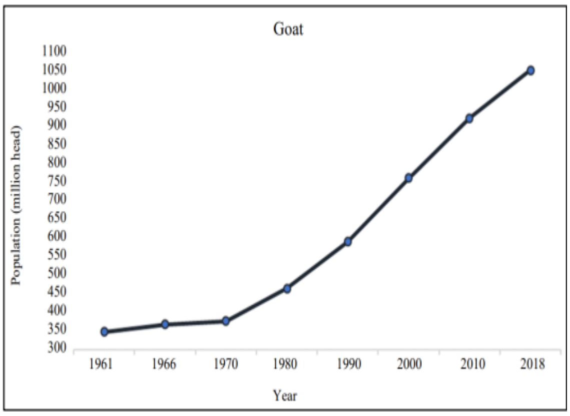


Fig. 1. The population of goats in the world was from 1961 to 2018. [Adapted from FAOSTAT]

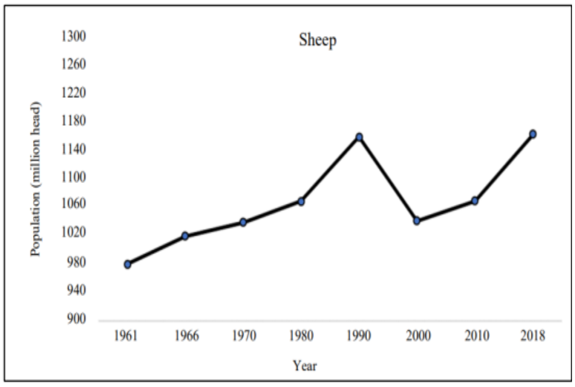


Fig. 2. The population of sheep in the world was from 1961 to 2018. [Adapted from FAOSTAT]

2. Methodology and Experimental Setup

Started by Engineer AbdiRizak, this project was designed to automatically feed animals. Currently, the project has been improved to work better. Using Raspberry Pi 4 Model B to run several motors and sensors as shown in Table 1. The new design will smooth out feeding by image processing.

Table 1. The planning and control components.

Components	Quantity
Raspberry Pi 4 Model B	1
Raspberry Pi 4 Heatsink	1
RPi Micro-HDMI to Standard HDMI	1
USB2.0 Camera Module OV3660	1
HC-SR04 Ultrasonic Sensor	2
USB Type C to Type C	1
Bracket for Ultrasonic HC-SR04	1
A4988 Stepper Motor Driver Module	2
3.7V 2000mAh Li-Ion Battery	5
3.7V 3350mAh 18650 Li-Ion Battery	1

The system will start with an ultrasonic sensor that detects the status of the Feeder Box if it is full or empty. When the Feeder Box is empty, then the system will send a message through the Blynk app to the farmer that you have to fill the Feeder Box up. This notification will help to keep the Feeder Box full. If the farmer fills the Feeder Box, then the system proceeds with the rest of the process.

The system will start with an ultrasonic sensor that detects the status of the Feeder Box if it is full or empty. When the Feeder Box is empty, then the system will send a message through the Blynk app to the farmer that you must fill the Feeder Box up. This notification will help to keep the Feeder Box full. If the farmer fills the Feeder Box, then the system proceeds with the rest of the process.

Water level monitoring is Another important part of the system to make sure goats always have enough water. An ultrasonic sensor will check the trough water level. The relay will turn on the water pump to fill the trough when the water level is low. The relay will turn off the pump to save water when the level required is reached.

Collecting the dataset was the first step in the project. the researcher Handling this work, by setting up a good environment for data collection and creating the troughs using a 3D printer. To get a large enough number of images as shown in Table 2, pictures of the troughs were shot empty, partially empty, and full.

Table 2. Number of Images for Each Class.

Classes	**Number of Images**
Empty	195
Partially Empty	112
Full	206
Total	513

The images of the object were captured with different orientations and angles. This can enhance the generalization of the model to understand the structure and appearance of the object so that it can identify the object with a different viewpoint and orientation in the real world. Table 3 shows the example images of the three classes for this research.

Table 3. T Sample Images of the troughs after annotation.

	Sample images with different angels	
Empty		
Partially Empty		
Full		

Once the images are gathered, the images are imported into Roboflow for data annotations, data pre-processing, and data augmentation. The dataset first undergoes the data annotation process whereby the object in the image is labelled. Then, data pre-processing methods were used to reduce the size of the image to 640 × 640 and auto orient the annotation when there were changes to the image. The size of the dataset is 550 images only. Increasing the size of the training dataset can increase the performance of a deep-learning model. However, a large dataset requires time to capture images, label images, and prepare the dataset. Therefore, the data augmentation techniques are used to artificially increase the size of the dataset by creating a modified version of the existing dataset. Table 4 shows the dataset size for each class after data augmentation. The size of the dataset after data augmentation is 8453 images. Lastly, the dataset was split into the train set, validation set, and test set with a train:val:test ratio of 87:13:0 to be 792:117:0 of images. The dataset is converted into YOLOv8 PyTorch format for use in the training process.

Table 4. Number of images for each class after data augmentation.

Classes	Number of Images
Empty	304
Partially Empty	190
Full	298
Total	792

3. Results and Discussion

3.1. *Robot Farm*

As a result of the project, the robot farm automates the feeding and monitoring water for the goats. A feeder box, feeding troughs, water pump, and water trough are the main elements of the system. Two stepper motors and a servo motor help control the feeder box movement and operation. An ultrasonic sensor for monitoring the food and water level in the booth trough and Feeder Box to ensure it is always full. The Camera is used to control the parts with image processing as shown in Fig. 3. All these components work together to make sure the goats get food and water automatically.

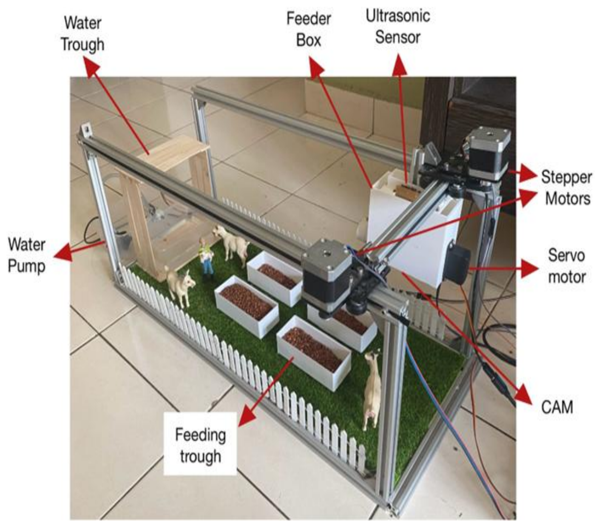


Fig. 3. Robot Farm.

The training results for the YOLOv8n model, illustrated in Fig. 4, show significant improvements in performance metrics. The train/box loss graph indicates a steady decrease from around 0.8 to 0.2 over 100 epochs, signifying enhanced accuracy in predicting trough locations. The train/classification loss reduced substantially from approximately 2.0 to nearly 0.0, highlighting the model's effectiveness in distinguishing between empty, partially empty, and full troughs, leading to high classification accuracy. Additionally, the train/distribution focal loss (DFL) decreased from about 1.3 to below 1.0, reflecting the model's increased precision in locating the troughs. Validation metrics, including Val/box loss and Val/classification loss, also improved, suggesting that the model generalizes well to new data, minimizing overfitting. Precision and recall metrics provide further insights, with high precision close to 1.0 indicating a low false positive rate and recall values approaching 1.0 demonstrating effective capture of all relevant trough statuses. The mean Average Precision at IoU=0.50 (mAP50) and mAP50-95 remained high, with mAP50 near 1.0 and mAP50-95 stabilizing around 0.9, confirming the model's robustness and accuracy across various detection challenges.

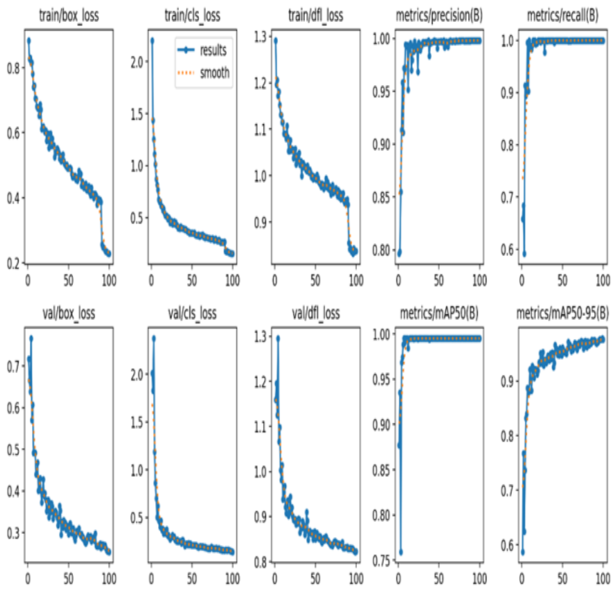


Fig. 4. Training and Validation Metrics for YOLOv8n Model on Trough Status Classification.

The Precision-Confidence Curve in Fig. 5 shows how the accuracy of the model's predictions (precision) changes as the confidence level varies. A higher confidence level means the model is more certain about its predictions. In this case, the precision remains very high, close to 1.0, across almost all confidence levels for all trough statuses (empty, partially empty, and full). This means the model is very accurate, even when the confidence level is low, ensuring that most predictions it makes are correct.

The Recall-Confidence Curve in Fig. 6 shows how the model's ability to find all true instances changes with different confidence levels. Over many confidence levels, the recall stays almost high at 1.0. This shows that even with a low confidence threshold, the model is rather good in identifying all possible occurrences of the trough

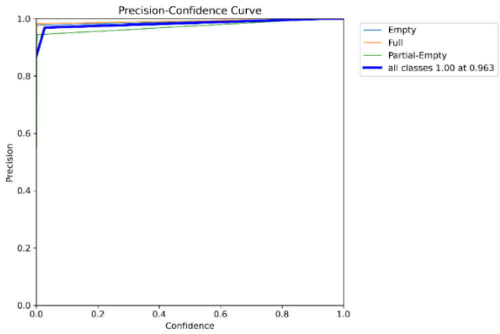


Fig. 5. Precision-Confidence Curve.

statuses. As the model becomes more selective with its predictions, the slight drop in recall at higher confidence levels is normal.

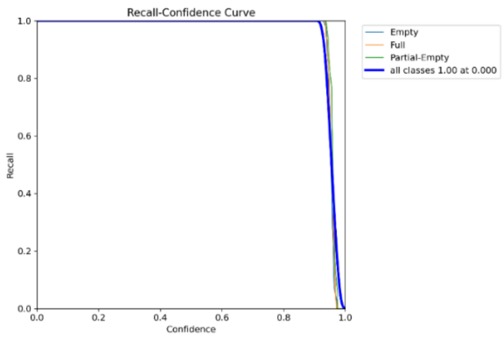


Fig. 6. Recall-Confidence Curve.

The Precision-Recall Curve in Fig. 7 shows how can the model identify and classify the trough statuses. The curve is near the top-right corner, showing that both precision and recall are close to 1.0 for all classes, which means the model is excellent at finding all instances of each class and making accurate predictions. The total precision (mAP @ 0 5) is very high at 0.995 which means the model has very high accuracy and reliability.

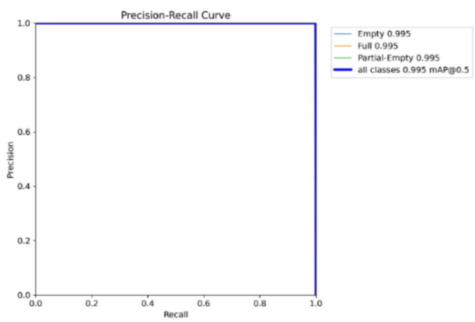


Fig. 7. Precision-Recall Curve.

The above results can be verified using Fig. 8 (Normalized Confusion Matrix). This proportion matrix corrects predictions for each class in the cells. The diagonal elements are all 1.00, indicating that the model correctly classified each trough status with an accuracy of 100%, predicting the entire class without a single mistake. The normalized confusion matrix provides a more interpretable visualization of the model's performance, especially useful for classes with imbalanced distribution. However, in this instance, both the raw counts and the normalized matrix demonstrate perfect classification performance.

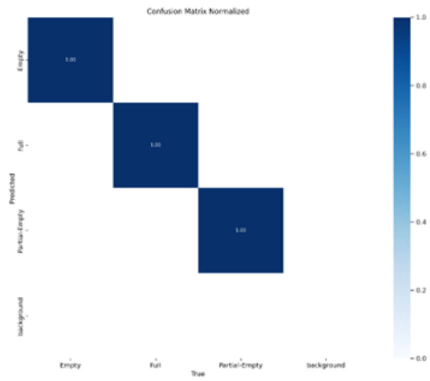


Fig. 8. Normalized Confusion Matrix.

4. Discussion

The project developed an automated system to monitor and manage goat feeding troughs, effectively addressing traditional method issues. By integrating a YOLOv8n model with hardware such as the Raspberry Pi 4, ultrasonic sensors, and a camera module, the system accurately identified troughs statuses (empty, partially empty, or full), significantly reducing the need for manual checks through image processing and machine learning. Key steps included collecting and enhancing a large set of trough images, using a deep learning model for object detection, and implementing real-time monitoring with the Blynk app, resulting in timely alerts for farmers. The model demonstrated high precision, recall, and F1 scores, with confusion matrices confirming accurate classification and performance curves indicating reliability. The project's success, driven by careful planning, effective methodology, and advanced technology, reduces farmers' workload, enhances resource management and goat care, and sets the stage for further innovations in automated farming, illustrating the potential of AI and IoT to improve agriculture.

5. Conclusion

The project successfully created an advanced system for automatically monitoring and managing goat-feeding troughs and solved a key problem in traditional farming methods. By using modern technology like the Raspberry Pi 4, ultrasonic sensors, a USB camera, and the YOLOv8n deep learning model, the system can identify whether the troughs are empty, partially empty, or full. This accuracy is crucial for making sure goats are fed properly and consistently, which directly affects their health and well-being. The project highlighted the importance of having accurate, real-time information, allowing farmers to quickly respond to their animals' needs. During the project, there were several challenges, especially with managing the different power needs of the components and making sure they worked well together. These issues were solved with careful planning and creative solutions. For example, using a voltage regulator to ensure the components receive the right amount of power, to keep the system stable and functioning properly. The inclusion of a user-friendly mobile app for real-time alerts made the system practical, enabling farmers to monitor the Feeder Box and take action when needed. The successful operation of this system showed how AI and IoT technologies can be applied in agriculture, demonstrating their potential to improve traditional farming practices. Not only did the project achieve its goals, but it also provided important lessons for future projects looking to use technology in agriculture. This includes expanding the system's abilities to monitor other conditions or using it for different types of farm management. The project's comprehensive approach, from planning to implementation and evaluation, serves as a good example for future projects. By improving the efficiency and accuracy of farm management, this project helps towards the larger goal of enhancing food security and sustainability in agriculture. Overall, the project demonstrates how modern technology can effectively address challenges in farming, leading to more advanced and sustainable agricultural practices.

References

1. Norton, T., Chen, C., Larsen, M. L. V., & Berckmans, D. (2019) Review: Precision livestock farming: Building 'digital representations' to bring the animals closer to the farmer, Animal, 13 (12), pp. 3009–3017, https://www.sciencedirect.com/science/article/pii/S1751731111900199X
2. Saar, M., Edan, Y., Godo, A., Lepar, J., Parmet, Y., & Halachmi, I. (2022a) A machine vision system to predict individual cow feed intake of different feeds in a cowshed, Animal, 16 (1), 100432, https://www.sciencedirect.com/science/article/pii/S1751731121002780
3. Borges Oliveira, D. A., Ribeiro Pereira, L. G., Bresolin, T., Pontes Ferreira, R. E., & Reboucas Dorea, J. R. (2021) A review of deep learning algorithms for computer vision systems in Livestock. Livestock Science, 253, 104700, https://doi.org/10.1016/j.livsci.2021.104700
4. Teletchea, F. (2019) 'Animal Domestication: A Brief Overview', in Animal Domestication. IntechOpen. Available at: https://doi.org/10.5772/intechopen.86783
5. Morris, S.T. (2017) Overview of sheep production systems, in Advances in Sheep Welfare. Elsevier, pp. 19–35. Available at: https://doi.org/10.1016/B978-0-08-100718-1.00002-9
6. Food and Agriculture Organization (FAO) Statistics Food. FAOSTAT. (n.d.). https://www.fao.org/faostat/en/#home

Authors Introduction

ENG. SAMY ELMASRI

He received his Bachelor's degree in Mechatronics Engineering with Honors in 2024 from the Faculty of Engineering, UCSI University in Kuala Lumpur, Malaysia.

IR NOOR IDAYU BINTI MOHD TAHIR

Noor Idayu Mohd Tahir graduated from University Teknologi MARA, Malaysia in Mechanical Engineering. She was design engineer in the oil and gas industries for 7 years. Currently, she works as a lecturer under the Department of Mechanical and Mechatronics, Faculty of Technology and Built Environment, UCSI Kuala Lumpur. Her research interests include design and autonomous robotics.

AMMAR ABDULAZIZ AL TALIB

He received his B.Sc and M.Sc degrees in Mechanical Engineering from the University of Mosul Iraq. He has finished his PhD degree from UPM University, Malaysia. Member of the Institute of Mechanical Engineers UK (CEng. MIMechE). He has developed all the Postgraduate Programs at the Faculty of Engineering at UCSI University / Malaysia and worked as the Head of Postgraduate and Research department at the same faculty for the years 2010-2018.

Surface Stiffness Estimation using Active Strobe Imager

Taiki Yamaguchi
Graduate School of Engineering Science, Osaka University, 1-3, Machikaneyama-cho, Osaka, Japan

Kensuke Harada
Graduate School of Engineering Science, Osaka University, 1-3, Machikaneyama-cho, Osaka, Japan
Email: {yamaguchi@hlab, harada@}.sys.es.osaka-u.ac.jp

Kouji Mizoue
Mizoue Project Japan, 305-1 Takagi-cho, Fuchu-city,726-0013, Hiroshima, Japan

Makoto Kaneko
Graduate School of Engineering Science, Osaka University, 1-3, Machikaneyama-cho, Osaka, Japan
Email: {president, mkaneko}@mizoueproject.com

Abstract

In this paper, we propose a method for estimating mechanical impedance of surface like skin using an Active Strobe Imager (ASI). ASI has the capability to non-contactly excite the target surface using an air jet flow, and to qualitatively observe traveling waves through strobe illumination. On the other hand, this paper shows that the surface impedance parameter can be estimated without contacting the target surface along with the visualization through ASI. We solve two problems: one is the inverse problem and the other is the forward problem. In the inverse problem, we estimate the spring constant of the target using information obtained from the measurement data. In the forward problem, we determine the surface displacement from the applied force, we compared the obtained displacement from the measurement data, demonstrating that the spring constants were correctly estimated.

Keywords: Spring-constant, Non-contact, Estimation.

1. Introduction

The mechanical impedance of materials is recognized as a crucial parameter characterizing the subject. For instance, in the case of soft biological tissues, such as internal organs, skin, and muscles, the mechanical impedance serves as essential biomarkers in medical diagnostics.[1][2][3][4][5] Given this context, numerous foundational studies propose the mechanization of palpation perfor-med by physicians for medical diagnostics on soft matter.[6][7][8][9][10][11][12] However, most of these approaches predominantly rely on contact-based estimations, rendering them unsuitable for subjects where direct contact is undesirable, due to concerns such as infection or tissue damage.

In response to this issue, Kaneko et al. have proposed the Active Strobe Imager (ASI) as a non-contact method for estimating the mechanical impedance. [13] The ASI incorporates periodically controlled air jets and a stroboscope, as shown in Figure 1. It allows for the observation of the propagating waves induced by the air jet on the target, enabling the qualitative assessment of its stiffness without direct contact. Given these capabilities, qualitative estimation of stiffness in soft biological tissues has been widely conducted. Tanaka et al.[14] then irradiated the skin surface with a slid laser beam, measuring the observed waves generated on the skin surface using a high-speed camera, providing finer observations

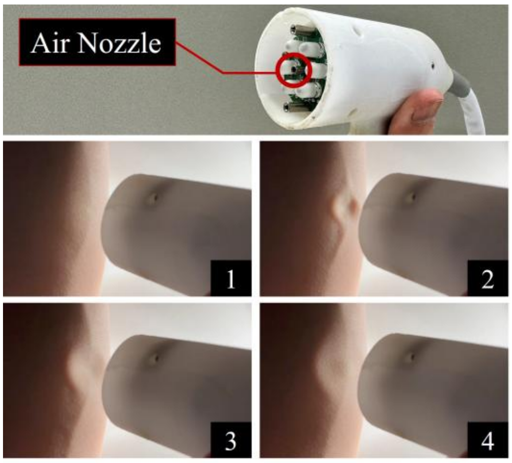


Fig 1. Waves propagating on the skin surface created by periodic air jet flows.

and assessments when air jetting was performed. In addition, Kawahara et al.[15] proposed a method to visually detect relatively rigid areas compared to the surrounding tissues by applying fluidic forces to the measurement target. Funai et al.[16] to proposed design parameters clearly visualing the movements of rapidly vibrating objects by appropriately setting stroboscopic illumination.Thus, attempts to use ASI for visually understanding the motion of the target and qualitatively assessing stiffness have been widely made. However, quantitative evaluations of measurement targets using ASI have not been investigated so far.[17]

In this paper, we propose a method to utilize strobe illumination to quantitatively estimating the surface mechanical impedance.We solve two problems.i.e., the inverse and forward problems-,i.e., We first model the dynamics of elastic membrane by using a series of linear springs. In the inverse problem, by estimating the force applied by the air jet by identifying the relation between the pressure of the air jet and force obtained from FSR, we probe that the spring constant can be identified. On the other hand, in the forward problem, we confirm that the estimated spring constant well matches the experimental result. In the forward problem, the analytical solution can be obtained in the explicit form.

In the subsequent chapters, we will summarize the measurement instruments and visualization techniques in Chapter 2, discuss the estimation methods for spring constants in Chapter 3, compile the measurement data in Chapter 4, and conclude in Chapter 5, 6.

2. Measurement Environment & Visualization Principles

To estimate the spring constant of the rubber membrane using ASI, it is necessary to have data on the applied force and displacement. Here, we will explain the equipment used to measure the force of the air jet generated from ASI and the displacement of the rubber membrane. Additionally, we will demonstrate the changes in the rubber membrane by visualizing the characterristics of the measured vibrations using stroboscopic illumination. Major headings should be typeset in sboldface with the first letter of important words capitalized.

2.1. *Force measurement device*

Figure 2 depicts the equipment used to measure the force of the air jet generated by ASI. A Force Sensing Resistor (FSR) is utilized as the force sensor, which can measure the periodic force of the air jet by applying it to the sensor.

2.2. *Instrumentation for Displacement Measurement*

Next, figure 3 illustrates a displacement measurem -ent apparatus. This apparatus is equipped with ASI, a rubber membrane, a laser displacement sensor, a camera, and a stroboscope. This device serves two functions: firstly, it measures the displacement of the rubber membrane, and secondly, it visualizes the vibrations of the rubber membrane induced by periodic air jetting.

2.3. *Visualization Principles*

We employed strobe effect for visualization. This method allows us to visualize the vibrations of the rubber membrane, which are too rapid to be observed by the naked eye due to ASI air jet, by precisely timing the

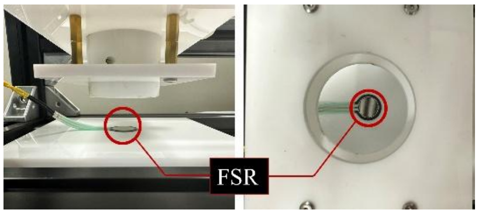


Fig 2. Force measurement instrument utilizing a force sensing register.

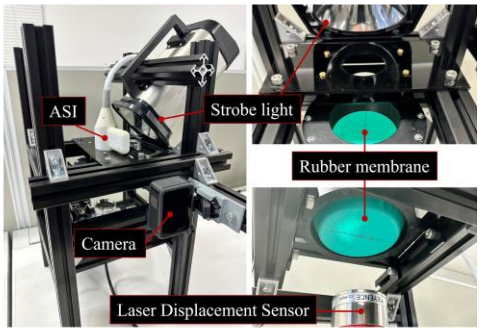


Fig 3. Displacement measurement device capable of measuring the displacement of a rubber membrane and visualizing progressive waves.

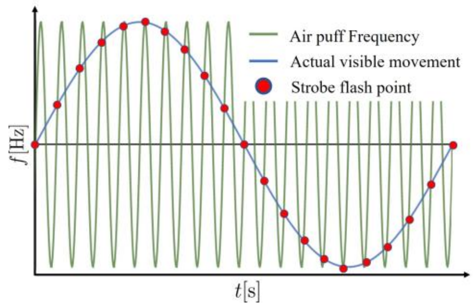


Fig 4. Waves propagating on the skin surface created by periodic air jet flows.

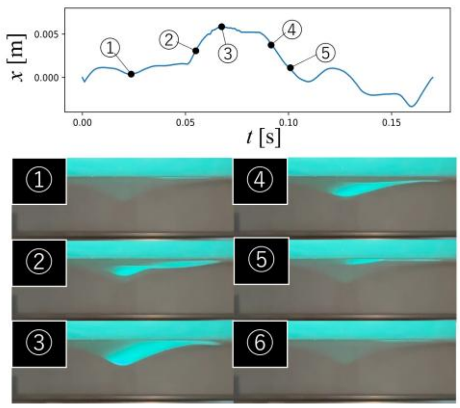


Fig 5. Visualization principle generated by periodic air jetting and stroboscope illumination.

strobe illumination. figure 4 depicts a graph illustrating the timing of air jet and strobe illumination. The green curve represents the frequency of the air jet, while the red dots indicate the strobe illumination points. By approspriately adjusting the strobe illumination frequency relative to the frequency of the air jet, we utilize a mechanism that allows us to observe the solid blue line enabling visualization.

Figure 5 illustrates a graph depicting the displacement of the rubber membrane subjected to excitation by ASI, as measured by a laser displacement sensor, along with its visualization of vibrations. In this experiment, the frequency of air jetting was set to 60 Hz with a 50% duty factor, and the stroboscope illumination frequency was set to 61 Hz. The displacement information measured by the laser displacement sensor is represented by the solid blue line. As observed in the visualized images, the rubber membrane initially undergoes depression, followed by a vibration pattern where the membrane returns when the electromagnetic valve closes.

3. Estimation of spring constant

In this section, we will elucidate a model for estimating spring constants based on the data of force and displacement measured using the instruments discussed in Chapter 2. Specifically, we will consider models for a thin rubber membrane, one employing two springs and another employing three springs and we will estimate the spring constants for each model using the ASI and compare the results.

3.1. *Two-Spring Model*

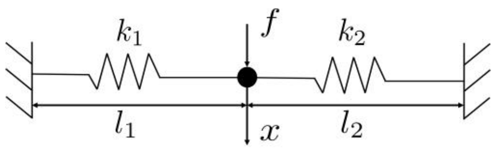


Fig 6. Overview diagram of the Two-Spring Model.

Estimate the spring constants of the rubber membrane, we considered the system shown in figure 6 and formulated the motion equation, represented as equation 1, to describe the response to the applied forces. Here, k_1 and k_2 represent the spring constants, l_1 and l_2 denote the distances from the fixed ends of the springs to the measurement points, f represents the force of the air jet, and x represents the displacement of the rubber membrane. Additionally, both the force f and the displacement x are continuously measured until the end of the measurement period, denoted as t_{end}.Furthermore, we express equation 1 in the form of a state equation using equation 2.

$$f(t) = 2k_1\frac{x^3(t)}{l_1^2} + 2k_2\frac{x^3(t)}{l_2^2} \quad , \quad \forall t \in [0, t_{end}) \tag{1}$$

$$\begin{bmatrix} f(t) \\ \vdots \\ f(t_{end}) \end{bmatrix} = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} \begin{bmatrix} \frac{2x^3(t)}{l_1^2} & \frac{2x^3(t)}{l_2^2} \\ \vdots & \vdots \\ \frac{2x^3(t_{end})}{l_1^2} & \frac{2x^3(t_{end})}{l_2^2} \end{bmatrix} \tag{2}$$

Next here, we consider an inverse problem to estimate the spring constants. If the time-series force $f(t)$ and displacement $x(t)$ as depicted in the equation are known, the spring constants k_1 and k_2 can be expressed using the generalized inverse of the displacement matrix, as shown in the following equation 3 (where "#" denotes the generalized inverse). Additionally, in the current measurement environment where the laser displacement sensor provides measurements at only one point, we assume that the force f applied by air injection is ideally applied from the injection nozzle to the rubber membrane. Therefore, it can be represented as the product of the nozzle cross-sectional area S and air pressure P, as given in equation 4.

$$\begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \frac{2x^3(t)}{l_1^2} & \cdots & \frac{2x^3(t_{end})}{l_1^2} \\ \frac{2x^3(t)}{l_2^2} & \cdots & \frac{2x^3(t_{end})}{l_2^2} \end{bmatrix}^{\#} \begin{bmatrix} f(t) \\ \vdots \\ f(t_{end}) \end{bmatrix} \tag{3}$$

$$f(t) = SP \tag{4}$$

On the other hand, in order to compare the displacement with the actual measured value, it is necessary to find an equation that explains the displacement based on the relationship between the estimated spring constants k_1 and k_2 obtained in Equation 3 and the force. Here, the comparison is made from Equation 5, where Equation 1 is obtained analytically for displacement x.

$$x = \left(\frac{f}{K}\right)^{\frac{1}{3}} \tag{5}$$

$$K = 2\left(\frac{k_1}{l_1^2} + \frac{k_2}{l_2^2}\right)$$

3.2. *Three-Spring Model*

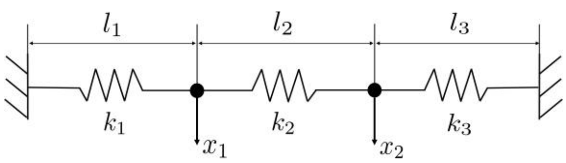


Fig 7. Overview diagram of the Three-Spring Model.

Furthermore, consider a model in which the rubber membrane is described by three springs. As shown in figure 7, unlike the previously mentioned two-spring model, adding more springs results in two points where forces are applied. To simplify the problem, we will proceed with estimating the spring constants under the condition that the force is applied to either the x_1 or x_2 side only. Therefore, we provide the equations of motion for the model where force is applied to the x_1 side, as shown in Figure. 8a), in equations 6a and 6b, and the equations of motion for the model where force is applied to the x_2 side, as shown in figure 8b, in equations 7a and 7b. Furthermore, the force f here is given in the same manner as in equation 4 for the two-spring model.

$$\begin{cases} 2k_{11}\dfrac{x_1^3(t)}{l_1^2} + 2k_{12}\dfrac{(x_1(t)-x_2(t))^3}{l_2^2} = f(t) & \text{(6a)} \\ -2k_{21}\dfrac{(x_1(t)-x_2(t))^3}{l_2^2} + 2k_{22}\dfrac{x_2^3(t)}{l_3^2} = 0 & \text{(6b)} \end{cases}$$

$$\begin{cases} 2k_{11}\dfrac{x_1^3(t)}{l_1^2} + 2k_{12}\dfrac{(x_1(t)-x_2(t))^3}{l_2^2} = 0 & \text{(7a)} \\ -2k_{21}\dfrac{(x_1(t)-x_2(t))^3}{l_2^2} + 2k_{22}\dfrac{x_2^3(t)}{l_3^2} = f(t) & \text{(7b)} \end{cases}$$

Here, k represent spring constants, while l_1, l_2, and l_3 denote the distances to each measurement point. We have set up the force $f(t)$ and displacements $x_1(t)$ and $x_2(t)$ as measurement data obtained over time. Consider the problem of estimating the spring constant by inverse problem of Equations 6 and 7 in conjunction. As mentioned in the introduction, when a force is applied to one side, no force is applied to the other side. If we wish to estimate the spring constants k_{12}, k_{22} using only the two equations in Equation 6, there is a problem that k_{22} cannot be obtained because no force is applied in Equation 6b. Therefore, we focus on Equations 6a and 7b, in which a force is applied, and estimate the respective spring constants from the equations of state shown in Equation 8.

$$\boldsymbol{f} = \boldsymbol{k}\boldsymbol{x} \tag{8}$$

$$\boldsymbol{f} = \begin{cases} \begin{bmatrix} f \\ 0 \end{bmatrix}, & \text{if a force is applied to } x_1, \text{ and this results in } f_2 = 0. \\ \begin{bmatrix} 0 \\ f \end{bmatrix}, & \text{if a force is applied to } x_2, \text{ and this results in } f_1 = 0. \end{cases}$$

$$\boldsymbol{k} = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix} \quad \boldsymbol{x} = \begin{bmatrix} \frac{2x_1^3(t)}{l_1^2} & \frac{-2(x_1(t)-x_2(t))^3}{l_2^2} \\ \frac{2(x_1(t)-x_2(t))^3}{l_2^2} & \frac{2x_2^3(t)}{l_3^2} \end{bmatrix}$$

Therefore, from Equation 8 we obtain Equation 9 for estimating the spring constant.

$$\boldsymbol{k} = \boldsymbol{x}^{\#}\boldsymbol{f} \tag{9}$$

Next, as in the two-spring model, consider the forward problem of finding the displacements x_1 and x_2 from the estimated spring constants $\boldsymbol{k}$, and force $\boldsymbol{f}$.
First, consider the forward problem shown in figure 8a. Here, the analytical displacements x_1 and x_2 are calculated by solving the simultaneous equations in Equation 6. First, Equation 6b is computed for x_2 to obtain Equation 10.

$$x_2 = Qx_1 \tag{10}$$

$$Q = \frac{1}{1+A} \qquad A = \left(\frac{k_{21}l_3^2}{k_{22}\ l_2^2}\right)^{\frac{1}{3}}$$

Next, substitute Equation 10 into x_2 in Equation 6a to obtain the analytical solution for x_1.

$$x_1 = \left(\frac{f}{W}\right)^{\frac{1}{3}} \tag{11}$$

$$W = 2\left(\frac{k_{11}}{l_1^2} + \frac{k_{12}}{l_2^2}(1-Q)^3\right)$$

We will similarly seek an analytical solution for x_2. Since we already have the expression for x_1 from Equation 11, we can substitute it into Equation 6b and solve for x_2 as a solution to a cubic equation. For brevity, we will only present the coefficients of the cubic equation here.

$$\alpha x_2^3 - 6\beta\left(\frac{f}{W}\right)^{\frac{1}{3}}x_2^2 + 6\beta\left(\frac{f}{W}\right)^{\frac{2}{3}}x_2 - 2\beta\left(\frac{f}{W}\right) = 0 \tag{12}$$

$$\alpha = 2\left(\frac{k_{21}}{l_2^2} + \frac{k_{22}}{l_3^2}\right) \qquad \beta = \frac{k_{21}}{l_2^2}$$

Similarly, we seek to obtain the analytical solution for the case depicted in figure 8b. We begin by focusing on Equation 7a and solving for x_1, thereby deriving Equation 13.

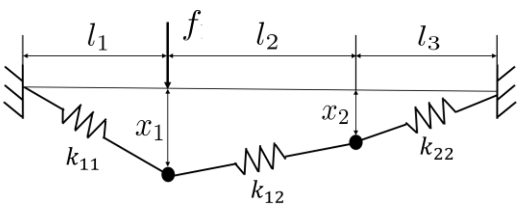
$$x_1 = Sx_2 \tag{13}$$

$$P = \frac{S}{1+S} \qquad S = \left(\frac{k_{12}l_1^2}{k_{11}\ l_3^2}\right)^{\frac{1}{3}}$$

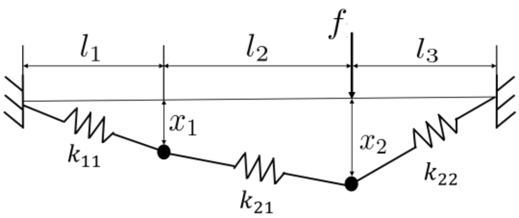
We will now substitute equation 13 into equation 7b to obtain the expression for x_2.

$$x_2 = \left(\frac{f}{T}\right)^{\frac{1}{3}} \tag{14}$$

$$T = 2\left(-\frac{k_{21}}{l_2^2}(P-1)^3 + \frac{k_{22}}{l_3^2}\right)$$



(a) When force is applied to x_1 side.



(b) When force is applied to x_2 side.

Fig 8. Overview diagram of the Three-Spring Model.

Thereafter, similar to Equation 12, Equation 14 is substituted into Equation 7a, and the coefficients of the cubic equation are summarized.

$$\gamma x_1^3 - 6\delta\left(\frac{f}{T}\right)^{\frac{1}{3}} x_1^2 + 6\delta\left(\frac{f}{T}\right)^{\frac{2}{3}} x_1 - 2\delta\left(\frac{f}{T}\right) = 0 \quad (15)$$

$$\gamma = 2\left(\frac{k_{11}}{l_1^2} + \frac{k_{12}}{l_2^2}\right) \qquad \delta = \frac{k_{12}}{l_2^2}$$

4. Experimental Measurement

In this section, we conducted experiments to measure force and displacement to estimate the spring constants for both the two-spring and three-spring models. We will summarize the results for each of them.

4.1. *Measurement of the Three-Spring Model*

The displacement and force time series data needed to estimate the two-spring model are shown in figure 9. The distances shown in figure 6, l_1 and l_2, were both set to l_1 and l_2 as 35×10^{-3} [m]. And ASI duty factor set to 100% for this data, and the magnitude of the force was ideally given as 1.57[N] by Equation 4.

4.2. ***Measurement of the Three-Spring Model***

Figure 10 depicts the waveforms of force and displacement when forces are applied to x_1, x_2 respectively, as shown in figure. 8a and figure 8b. The distance between measurement points is set as follows: $l_1 = l_3 = 27.5 \times 10^{-3}$ [m] and $l_2 = 15 \times 10^{-3}$ [m]. These waveforms were generated by directing ASI air jet at a duty factor of 100%.

On the other hand, ASI originally measures the target using periodic air jet. Therefore, in figure 11 (Displacement & Force), we measured the displacements, x_1 and x_2, by subjecting the rubber membrane to forced vibrations using a periodic air jet at 60 Hz with a 50% duty factor. However, there is an issue where the actual waveform of the force applied to the rubber membrane does not drop to zero at a 50% duty factor due to the physical delay in the opening and closing of the electromagnetic valve. To address this issue, we used the force measuring device shown in figure. 2 to measure the actual waveform and based on the duty factor indicated in figure 11 (FSR), we created an ideal rectangular waveform for the force, represented by the orange line.

5. Estimated Results

We will present the results of the spring constants obtained from the inverse problem in figure 12, 13, and 14, as well as the results obtained by solving the forward problem. First, figure. 12 shows the estimation results for a model with two springs, with spring constants estimated as $k_1 = k_2 = 8429.25$ [N/m]. On the other hand, we depict the forward problem solved based on the applied force and estimated spring constants with a red line. At this point, the displacement obtained from the analysis was $x = 3.84 \times 10^{-3}$ [m]. In contrast, we plotted the results of the measured data with a blue dashed line for comparison. As can be seen from the figure, the estimated displacement value and the measured displacement value match very closely, confirming that the spring constants were correctly estimated.

Next, focus on the results obtained for the model with three springs. In figure 13, these results were obtained by solving the inverse problem using the displacement and force data shown in figure 10. Here, because x_1 and x_2 interact with each other, we did not obtain results similar to the two-spring model. However, it is worth noting that the spring constants at both ends estimated for the two-spring model and the three-spring model when subjected to static forces show very close values.

Furthermore, when comparing the data obtained through the forward problem with the measurement data obtained using a laser displacement sensor, we confirmed that these results also closely match.

Finally, summarize the results obtained from the dynamic force application in figure 14. First, it is evident that the estimated spring constants are considerably lower when compared to the results obtained from the three-spring model subjected to static forces.

Additionally, a significant discrepancy can be observed between the analytically solved displacement values depicted by the red line and the measured displacement values. We attribute this difference to the influence of inertia caused by the periodic application of force, which was not accurately estimated.

However, based on the results from figure 12 and 13, which were obtained under static conditions, it has been demonstrated that in this model, the spring constants can be accurately estimated when subjected to static force application

6. Summary

In this study, we used ASI to measure and quantitatively evaluate a rubber membrane non-contact. Here, we estimated the spring constant from two models for thin rubber membranes and compared them by solving an forward problem using the obtained spring constants. As a result, the proposed model demonstrated accurate estimation under conditions with applied static forces. In the future, we plan to construct even more complex models and continue with the estimation process.

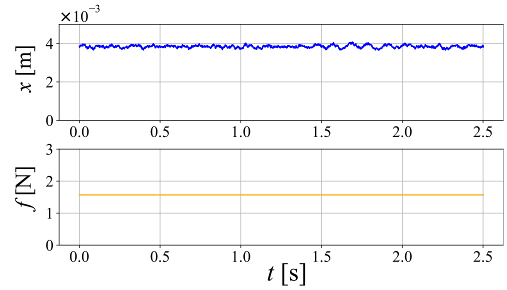


Fig 9. Static measurement data of displacement (blue line) and force

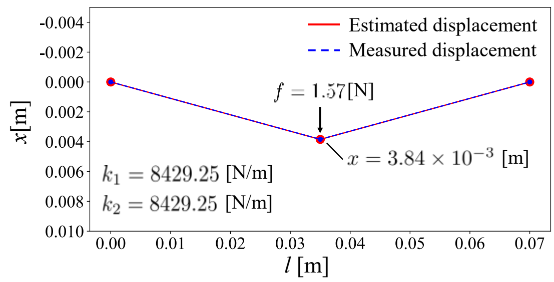


Fig12. Estimation results for a two-spring model. (when the force is static)

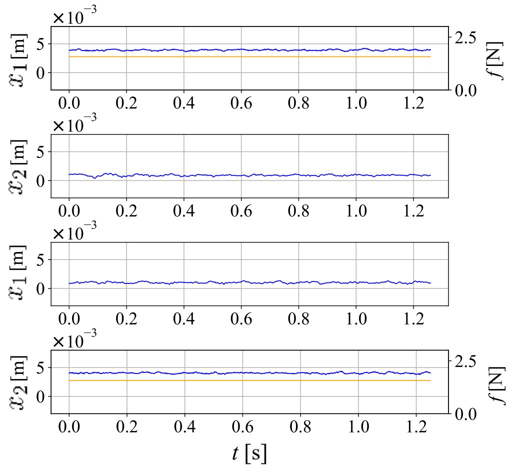


Fig10. Static measurement data of displacement (blue line) and force(orange line) for three spring model.

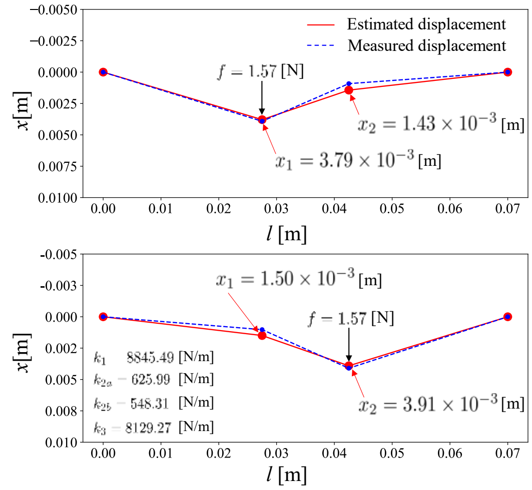


Fig13. Estimation results for three-spring models. The upper and lower figures show the static case with force applied to x_1 and x_2, respectively

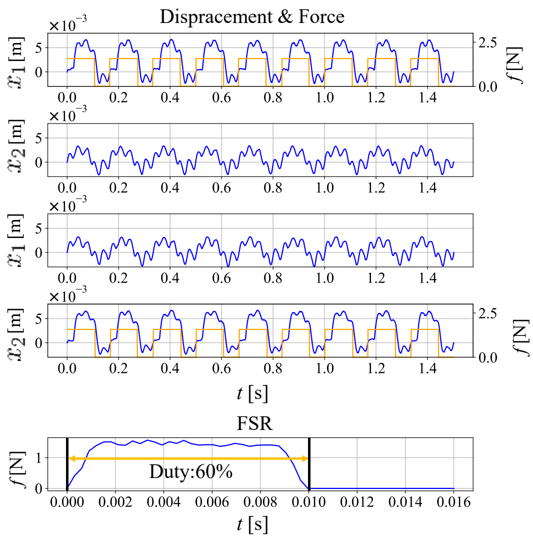


Fig 11. Dynamic measurement data of displacement (blue line) and force(orange line) in a two-spring model.

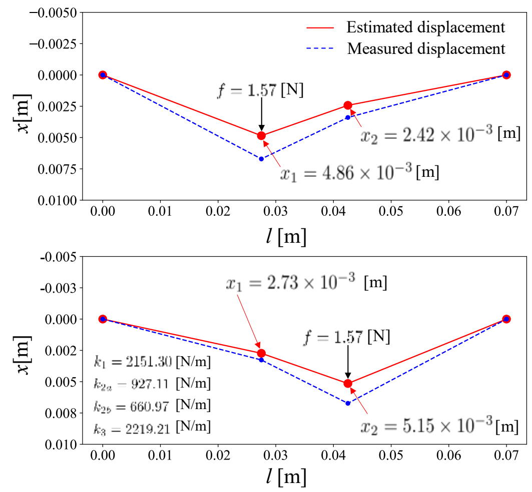


Fig14. Estimation results for the three springs model. The upper and lower figures show the case where the force is applied to x_1 and x_2, respectively, where the force is dynamic.

References

1. I. Kato, K. Koganezawa, and A. Takanishi. Automatic breast cancer palpation robot:Wapro-4. Advanced Robotics, Vol. 3, No. 4, pp. 251–261, 1988.
2. P. S. Wellman and R. D. Howe. Modeling probe and tissue interaction for tumor feature extraction. In 1997 ASME Summer Bioengineering Conf., Vol. 35, pp. 237–238, 1997.
3. A. M. Galea and R. D. Howe. Mammography registered tactile imaging. Surgery Simulation and Soft Tissue Modeling, pp. 183–193, 2003.
4. M. B. Shields. Diagnostic and surgical techniques. Survey of Ophthalmology, Vol. 24, No. 4, pp. 211–219, 1980.
5. T. Kawahara, C. Toya, N. Tanaka, M. Kaneko, Y. Miyata, M. Okajima, and T. Asahara. Non-contact impedance imager with phase differentiator. In Proc. of the 1st IEEE/RAS-EMBS Int. Conf. on Biomedical Robotics and Biomechatronics (BioRob2006), No. 159, 2006.
6. Y. Murayamaa, M. Haruta, Y. Hatakeyamab, T. Shiina, H. Sakumab, S. Takenoshita,S. Omataa, and C. E. Constantinou. Development of a new instrument for examination of stiffness in the breast using haptic sensor technology. Sensors and Actuators, A: Physical, Vol.143, No. 2, pp. 430–438, 2008.
7. C. Kleesattel and G. M. L. Gladwell. The contact-impedance meter-1. ULTRASONICS, Vol.6, pp.175–180, 7 1968.
8. H. L. Oestreicher. Field and impedance of an oscillating sphere in a viscoelastic medium with an application to biophysics. The Journal of Acoustical Society of America, Vol.23, No.6, pp.707–714, 1951.
9. G. M. L. Gladwell and C. Kleesattel. The contact-impedance meter-2. ULTRASONICS, Vol.6, No.4, pp.244–251, 10, 1968.
10. S. Omata and Y. Terunuma. New tactile sensor like the human hand and its applications. Sensors and Actuators A, Vol.35, pp.9–15,1992.
11. S. Chen, M. Fatemi, J. F. Greenleaf, and Mayo Clinic. Shear property characterization of viscoelastic media using vibrations induced by ultrasound radiation force. Proc IEEE Ultrason Symp, Vol.2, pp.1871–1875, 2002.
12. M. Ochi, J. Iwasa, M. Tobita, M. Katoh, T. Yamamoto, R. Fukushima, and T. Kurushima. Method for measuring stiffness of a cultured tissue, using a stiffness measurement device having a vibration detective unit. US Patent 7,198,908 B2, 2007.
13. M. Kaneko, C. Toya and M. Okajima: Active Strobe Imager for Visualizing Dynamic Behavior of Tumors, Proc.of the 2007 IEEE International Conf. on Robotics and Automation, 3009/3014 (2007)
14. N. Tanaka and M. Kaneko: Skin Surface Shock Wave,Proc. of Engineering in Medicine and Biology Society,2006. EMBS '06. 28th Annual International Conferenceof the IEEE, 4123/4126 (2006)
15. T. Kawahara, S. Matsunaga, S. Tanaka and M. kaneko: Non-contact Stiffness Imager, Robotics Society of Japan,Vol.24 No.3, pp.363-369, 2006.
16. K. Funai, K. Mizoue, M. Higashimori and M. Kaneko. Optimal Parameter Determination of Active Strobe Imager, T. SICE Vol.46 No.12 December (2010)
17. N. Tanaka, M. Higashimori, M. Kaneko, " Active Sensing for Viscoelastic Tissue with Coupling Effect",SICE Journal of Control, Measurement, and System Integration, Vol.44, No.10, 779-785, 2008.

Authors Introduction

Mr. Taiki Yamaguchi

He received his Master's degree from the Graduate School of Science and Engineering, Saga University, Japan, in 2022. He is currently a Doctoral course student at the Graduate School of Engi-neering Science, Osaka University, Japan.

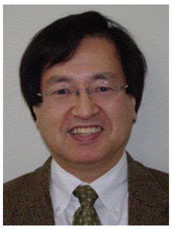
Prof. Kensuke Harada

He is a Professor at the Graduate School of Engineering Science, Osaka University, Japan, a Fellow of the IEEE, and a Cross-Appoint-ment Fellow at the National-Institute of Advanced Industrial Science and Technology (AIST), Japan.

Mr.Kouji Mizoue

He completed his graduate studies in the Department of Electronic and Electrical Engineering at the Graduate School of Engineering, Fukuyama University, in 2001. He is the President of Mizoue Project Japan and is engaged in the production and development of medical measurement devices.

Prof.Makoto Kaneko



He is a Professor Emeritus at the Graduate School of Engineering, Osaka University, Japan. Currently, he serves as a Visiting Professor and is recognized as a Fellow of the IEEE.

Painting Task Planning for Large Structure using a Mobile Manipulator

Hiroshi Tanaka
Graduate School of Engineering Science, Osaka University, Japan

Masato Tsuru
Graduate School of Engineering Science, Osaka University, Japan

Takuya Kiyokawa
Graduate School of Engineering Science, Osaka University, Japan

Kensuke Harada
Graduate School of Engineering Science, Osaka University, Japan
Email: {h.tanaka@hlab., tsuru@hlab., kiyokawa@, harada@}sys.es.osaka-u.ac.jp

Abstract

Painting a large structure with a robotic arm that is fixed to the ground is difficult due to its limited reachable range. To plan the robotic painting motion of such a large structure, we develop a robotic software system assuming a mobile manipulator to explore the environment using SLAM. Our software system includes both detection of AR markers and construction of the environmental map to determine the painting location. It can measure the error in self-position estimation that occurs during the movement. It can also generate spray trajectories for the recognized painting location and control the whole body using Model Predictive Control (MPC) to perform painting over a wide area.

Keywords: Painting, Large Structure, Mobile Manipulator, Model Predictive Control, SLAM

1. Introduction

While the manufacturing processes for a wide variety of industrial products have been automated in recent years, it is difficult to automate the painting of large structures. Large structures require the use of mobile manipulators because the painting area is so large that it exceeds the reach of the arm. Unlike a manipulator fixed to a pedestal, a mobile robot is subject to dynamic friction between its wheels and the ground, which causes errors in the position of the tip of the manipulator.

Figure 1 shows the large structure addressed in this study. It shows a bridge with several bridge girders placed on a platform. Although the shapes of the bridge girders themselves are known, it is difficult to accurately place them on the table, so the robot must perform the painting operation while recognizing in real time which position on the bridge girder is to be painted. In addition, such bridge girders often have few visual features, and it is often difficult for the robot to identify its own position by camera.

To address these problems, this study proposes a software system painting large structures using mobile manipulators. The proposed software system consists of three modules, 3D-SLAM[2], Model Predictive Controller (MPC)[1], and AR marker recognition[3]. The use of MPC enables a mobile manipulator to follow a target trajectory without being aware of the position of the dolly. In order to paint large objects with few visual features, the combination of AR markers and SLAM makes it possible to perform accurate painting operations while recognizing its own position.

In this paper, Section 2 describes related studies, followed by a description of the proposed method in Section 3. Finally, Section 4 describes the results of the evaluation of paint trajectories using physical simulations.

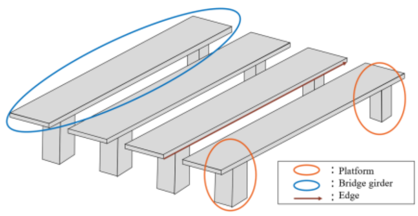


Figure 1 Large structures addressed in this study

2. Related Works

Researchers have used mobile manipulators to achieve various tasks [4]. Lin et al. researched spray painting using a robot manipulator [5]. Chen et al. also painted arbitrary free-form surfaces by approximating the shape of the object with a mesh model [6]. However, these studies on painting used a fixed robot arm, and did not consider using a mobile manipulator.

Dhanaraj et al. studied the spray painting using a mobile manipulator by fixing the cart section [7]. However, in this research, the position of the trolley had to be specified in advance, so it was not possible to paint while moving the trolley, and the painting took longer than when the trolley and arm were controlled simultaneously. Another study of painting using a mobile manipulator is the use of MPC in the agricultural field [8]. However, in a vineyard, the carts can only move in one direction, and the distance between the arm and the vines is always the same.

On the other hand, in this study, we propose a framework on spray painting using a mobile manipulator which actively explores the large structure by combining MPC and vision system.

3. Proposed Method

A schematic diagram of this research is shown in Figure 2. In this study, the following three ROS-based systems were created and combined for painting large structures, especially bridge girders, using mobile manipulators. The first is 3D-SLAM, which creates a map of the surrounding three-dimensional environment and estimates the robot's self-position based on this map. The second is an AR marker recognition system, which is necessary to identify the painting area. These systems enable the robot to obtain the position and orientation of AR markers on a 3D map of the environment generated by Simultaneous Localization and Mapping (SLAM), and to determine the position and orientation of multiple painted surfaces in the environment in 3D space. The third is a whole-body control system for mobile manipulators using Model Predictive Control (MPC). This allows the manipulator to follow the planned trajectory of the spray without explicitly providing the trajectory of the dolly, making it possible to easily paint over a large area beyond the reach of the arm. We propose a method to search for painting surfaces on multiple bridge girders by combining SLAM and AR marker recognition, and a method to generate spray launches for the shortest possible painting time by considering the next direction of movement for the recognized painted surfaces.

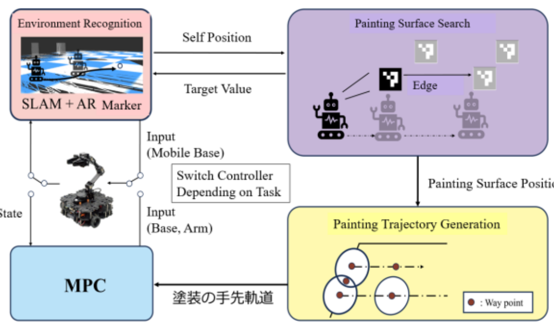


Figure 2 System overview

3.1. *Environment*

In this study, painting task is performed in the environment shown in Figure 1 where several bridge girders to be painted are placed on fixed temporary platforms. In such an environment, it is difficult to identify the pose of each bridge girder by aligning the point cloud using the CAD model of the bridge girders because only a portion of the bridge girders can be observed by the vision sensor. Therefore, this study uses the information on the edge of the bridge girder to identify the location of the bridge girder. The layout of the bridge girders is difficult to completely fix due to the huge size of the structure, and each bridge girder has multiple painted areas on its underside. For this reason, AR markers are placed at the four corners of each painted surface to recognize the painted surface. In this study, a mobile manipulator is used to paint over a wide range of painting trajectories beyond its own size without being aware of the position of the dolly.

3.1.1. *Combined System of SLAM and Marker Recognition*

By using a wheeled mobile robot, the robot's current position can be obtained using wheel odometry. However, errors occur in the self-position due to the effects of wheel slippage. SLAM can be used for compensating the error in self-position using a vision sensor. In particular, we use RTabMap[2] as a 3D-SLAM. RTabMap uses both wheel odometry and an RGB-D camera to obtain point cloud data of the map.

Figure 3 shows a system that combines 3D-SLAM and AR marker recognition. In this study, two types of RGB-D cameras are used: a SLAM camera (Camera 1) installed so that it faces horizontally and a camera (Camera 2) installed so that it faces vertically to recognize AR markers. With this camera configuration, we aim to avoid the large error in the self-position estimation because the accuracy of estimating the camera position based on the feature values is reduced when the camera faces the ceiling.

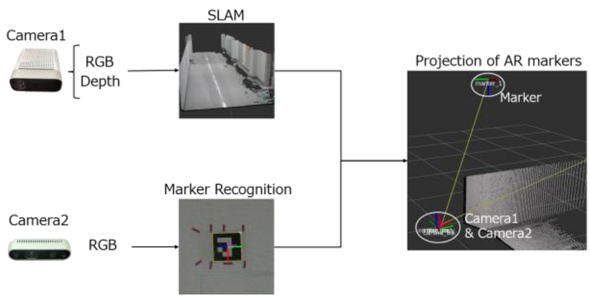


Figure 3 System diagram of combined 3D-SLAM and AR marker recognition

3.1.2. *Motion of Mobile Manipulator during Painting Action*

In this study, OCS2[9] is used as the MPC, and the MPC is used to provide state feedback and robot control. The robot was modified to send input commands to the end-effector. The trajectory to be followed by the end-effector and the target time are taken as inputs. The inputs are calculated by using the internal kinematics model and optimizer in MPC, and convergence to the target position is measured by receiving feedback from the robot. MPC is a control method that calculates the optimal input by solving optimization problems for the predictive model at each time to predict future responses.

OCS2 automatically computes the system dynamics from the geometric model of the mobile manipulator and calculates the inputs satisfying the constraints. The state variables of the system are the position and posture of the cart and the joint angles of the manipulator, as well as the translational and rotational speeds of the cart as input, and the manipulator.

3.2. *Painting Trajectory Generation*

This section describes a method for generating the shortest possible spray trajectory for a recognized painting location. Figure 4 shows an overview of the proposed method. The MPC takes as inputs a list of the coordinates of the four corners of the recognized painting area, the average Z-coordinates (height) of the corners, and a list of candidate points to be searched for next, and generates a set of square-wave like trajectories with minimum travel distance to paint the rectangular area. When we plan the square-wave like trajectory, we consider the spray width for overlapping with neighbor spray trajectory with the predefined amount. Finally, the tool tip trajectory can be obtained from the painting trajectory. The tool-tip trajectory becomes the input to the MPC.

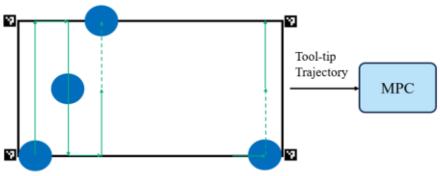
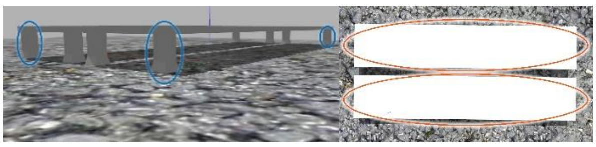


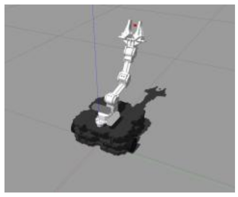
Figure 4 Paint Trajectory Generation for Recognized Paint Locations

4. Simulation

In this study, Gazebo was used as a physics simulator to evaluate the proposed method. Figures 5 shows the painting environment in Gazebo. The red and blue circles indicate bridge girders and temporary abutments, respectively. The dimensions of the bridge girder is 0.7 m high, 2.35 *m* deep, and 12 *m* wide. In this study, forward and vertical upward facing RGB-D cameras were added to the model. The control horizon of MPC was set to 15.0[*s*]. Figure 6 shows the painting environment used in Experiment 1. Markers are placed in the four corners of the painted area with red circles, and the number of the painted area inside the marker is defined as 1 to 6, as in "Pait Area 1". Similarly, Figure 7 shows the painting environment shown in Experiment 2. The difference between these experiments is the position of AR markers.



(a) Bird's-eye view (b) Top view



(c) Mobile manipulator

Figure 5 Painting environment in the simulator

4.1. *Results*

Figure 8 shows a scene from Experiment 1. The red line in the figure shows the ideal trajectory of MPC, and it can be seen that the coordinate system, which is the result of marker recognition, has increased. These results show that the painting locations have been searched based on the markers and edges in the environment, and that the mobile manipulator can appropriately control the whole body in relation to the ideal trajectory of the tip of the hand using MPC. Figure 9 shows the ideal and actual tip trajectories. The gray rectangles represent the painting area calculated from the placement of markers, and the red and orange dots represent the start and end points of each painting activation, respectively. This figure shows that this method can be used to adjust the painting path appropriately to the starting position of the mobile manipulator. This figure also shows that each coating point is able to follow the track with little error.

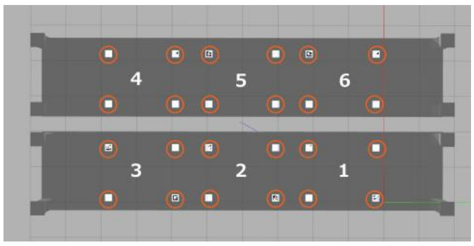


Figure 6 Painting environment in Experiment 1

Figure 10 shows a scene from Experiment 2. From these figures, it can be confirmed that the system can search and paint appropriately even when the size of the painted area changes in the horizontal direction. Figure 11 shows the position of the painting tip in the entire environment, the ideal tip trajectory used as input to MPC, and the painting location using the marker positions in the simulation. From this figure, it can be seen that the method is able to calculate trajectories appropriately for the size of the coating area and that it is able to follow the ideal trajectory at each coating area. These results confirm that the method successfully searches for paint locations by using the edges of the environment and performs painting operations on the recognized paint locations.

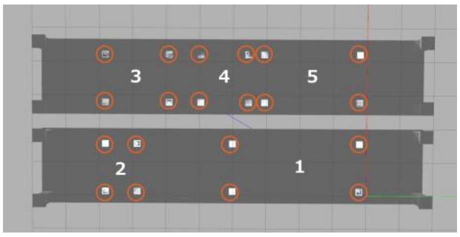
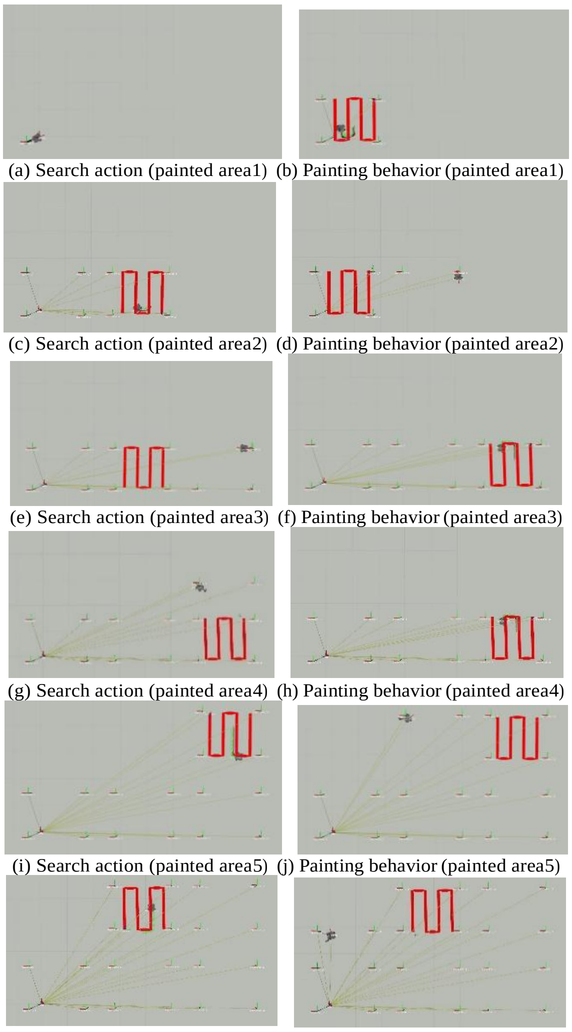


Figure 7 Painting environment in Experiment 2


(a) Search action (painted area1) (b) Painting behavior (painted area1)
(c) Search action (painted area2) (d) Painting behavior (painted area2)
(e) Search action (painted area3) (f) Painting behavior (painted area3)
(g) Search action (painted area4) (h) Painting behavior (painted area4)
(i) Search action (painted area5) (j) Painting behavior (painted area5)
(k) Search action (painted area6) (l) Painting behavior (painted area6)
Figure 8 Result of Experiment 1

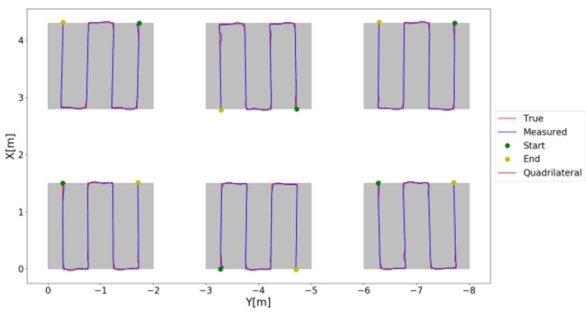
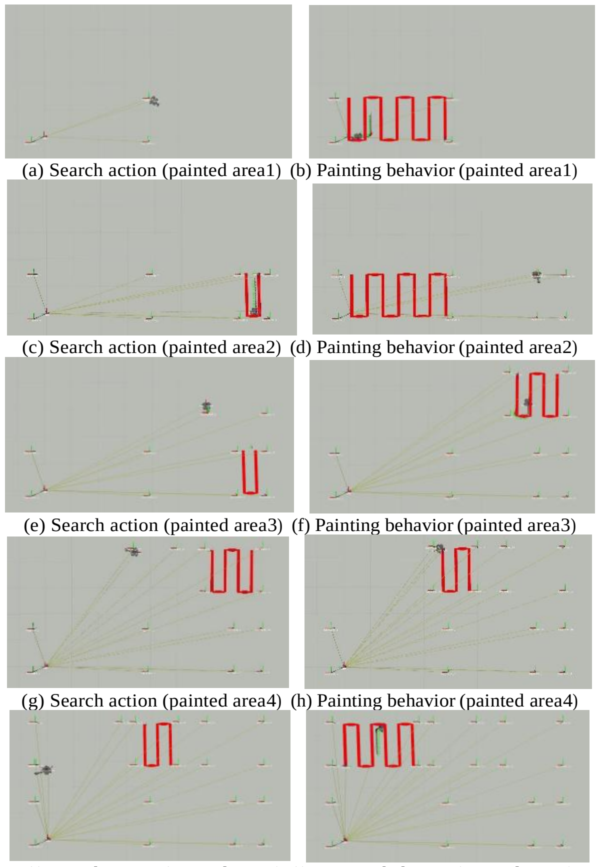


Figure 9 Relationship between the ideal trajectory and the position of the tool-tip during painting in the entire environment of Experiment 1


(a) Search action (painted area1) (b) Painting behavior (painted area1)
(c) Search action (painted area2) (d) Painting behavior (painted area2)
(e) Search action (painted area3) (f) Painting behavior (painted area3)
(g) Search action (painted area4) (h) Painting behavior (painted area4)
(i) Search action (painted area5) (j) Painting behavior (painted area5)
Figure 10 Result of Experiment 2

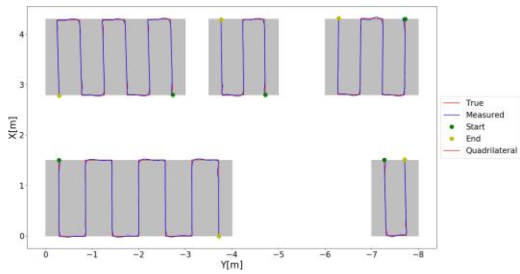


Figure 11 Relationship between tool-tip position and ideal trajectory during painting in the entire environment of Experiment 2

5. Conclusions

In this study, a software system was developed to plan painting operations on large structures using mobile manipulators. The proposed software system consists of three modules, 3D-SLAM, Model Predictive Controller (MPC), and AR marker recognition. In the action planning, the robot recognizes AR markers with its vision sensor and can plan the trajectory of the manipulator to paint within the range indicated by the markers. Future work includes testing the proposed method on actual bridge girders to verify its effectiveness.

References

1. J. Pankert and M. Hutter. Perceptive Model Predictive Control for Continuous Mobile Manipulation, Robotics and Automation Letters, 5(4), 2020.
2. M. Labb´e and F. Michaud. Rtab-map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation. *J. Field Robotics*, 36(2):416-446, 2019.
3. R. Munoz-Salinas and R. Medina-Carnicer. Ucoslam: Simultaneous local- ization and mapping by fusion of keypoints and squared planar markers. *Pattern Recognition*, 101:107193, 2020.
4. S. Chitta, B. Cohen, and M. Likhachev. Planning for autonomous door opening with a mobile manipulator. *IEEE Int. Conf. on Robotics and Automation*, pp. 1799-1806, 2010.
5. W. Lin, A. Anwar, Z. Li, M. Tong, J. Qiu, and H. Gao. Recognition and pose estimation of auto parts for an autonomous spray painting robot. *IEEE Trans. on Industrial Informatics*, 2019.
6. H. Chen et al., Automated robot trajectory planning for spray painting of free-form surfaces in automotive manufacturing. In *Proc. of IEEE Int. Conf. on Robotics and Automation*, vol. 1, pp. 450 -455 vol.1, 2002.
7. N. Dhanaraj, Y. Yoon, R. Malhan, P. Bhatt, S. Thakar, and S. Gupta. A mobile manipulator system for accurate and efficient spraying on large surfaces. *Procedia Computer Science*, 200:1528-1539, 2022.
8. V. Ivo, V. Goran, and K. Zdenko. task space model predictive control for vineyard spraying with a mobile manipulator. *Agriculture*, 12(3), 2022.
9. https://github.com/leggedrobotics/ocs2

Authors Introduction

Mr. Hiroshi Tanaka

He received his Bachelor's and Master's degree in Engineering in 2022 and 2024, respectively, from Osaka University in Japan.

Dr. Masato Tsuru

He received his M.S., and Ph.D. degrees from Osaka University, Japan, in 2021 and 2024, respectively.

Prof. Takuya Kiyokawa

He received his Ph.D degree from Nara Institute of Advanced Science and Technology in 2021. Currently, he is an Associate Professor at the Graduate School of Engineering Science, Osaka University, Japan.

Prof. Kensuke Harada

He is a Professor at the Graduate School of Engineering Science, Osaka University, Japan, a Fellow of the IEEE, and a Cross-Appoint-ment Fellow at the National- Institute of Advanced Industrial Science and Technology (AIST), Japan.

Real-time cable tracking by wire segmentation and Coherent Point Drift

Ryunosuke Yamada [1 *], Tokuo Tsuji [1 **], Takahiro Shimizu [2], Shota Ishikawa [1 2],
Tomoaki Ozaki [2], Yusuke Sakamoto [1], Tatsuhiro Hiramitsu [1], Hiroaki Seki [1]
[1] *Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan*
[2] *DENSO CORPORATION Advanced Research and Innovation Center, 500-1, Minamiyama, komenoki-cho, Nisshin-shi, Aichi-ken, 470-0111, Japan*

*Email: *noboriryuu0623@stu.kanazawa-u.ac.jp, **tokuo-tsuji@se.kanazawa-u.ac.jp*

Abstract

In this paper, a real-time cable tracking system by fast segmentation method and Coherent Point Drift (CPD) is proposed. Fast cable segmentation based on color space is inaccurate because of background contrast. Therefore, this technique uses edge information from the image to address this problem. The method consists of three processes: threshold processing in the Luv color space, edge processing using a Laplacian filter, and processing for extracting the common region of the binary images generated by each process. In the experiments, the accuracy of the segmentation region and the processing time required for each process of the tracking system are shown.

Keywords: Deformable Linear Objects, Segmentation, Three-dimensional Tracking, Real-time systems

1. Introduction

Our research focuses on automating cable routing tasks that are common in the manufacturing process. Cable routing tasks mainly require the operation of hanging the cable on the routing target. However, cables easily change state during manipulation, so it is difficult to hang them on the target with pre-planned movements. Therefore, it is necessary to track the cable to modify the movement depending on the deformation. In our experimental environment, a 6-DOF DENSO COBOTTA robot, cables, and pins to be placed are set on a base plate as shown in Fig. 1. The cables are recognized using RGB and depth images captured by the environmental camera. In the process, the robot and the pins prevent recognition. Our main research strategy is feedback control that includes a predictive model of cable behavior. If we can acquire cable status in real time, the accuracy of cable behavior prediction will be improved. Based on the above considerations, a high-speed cable tracking system that is robust to occlusion is required. There is a tracking method using markers [1], but it is not suitable because it is poor at handling occlusions. Therefore, we implement a tracking system with a function to acquire a cable point cloud and a function to place cable nodes in the point cloud.

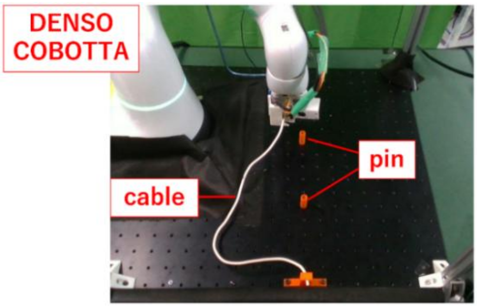


Fig. 1. Our experimental environment. There is DENSO COBOTTA robot and a white cable, and a pin to hang the cable.

The cable point cloud is obtained by overlaying a depth image and a cable mask image. The basic method for rapid processing of cable mask images is the color space-based method. Using this method, a segmentation method that uses a chain model to deal with occlusion [2] has been proposed. However, this method works only if there are no objects in the environment that are the same color as the cables. There is also an instance segmentation method [3] that uses superpixelization and graphs to deal with complex environments, but its processing is expensive. Learning-based methods have a trade-off between accuracy and inference time. As a fast model, Caporali et al. (2022) proposed a method for instance segmentation of cables [4]. The method uses deep learning to obtain cable regions, and subsequently superpixelizes them and solves the graph connection. Cable tracking is performed by placing cable nodes on the point cloud. Kicki (2023) et al. proposed a fast tracking [5] that applies β-spline functions to a thin cable point cloud. However, it was not tested in the case of global occlusion. On the other hand, Xiang (2023) et al. proposed a fast and robust tracking method [6] using CPD to place nodes stochastically in a point cloud even in the presence of global occlusions.

In this study, we propose a system that acquires a cable point cloud from a masked image by fast segmentation without learning, and tracks it by CPD. The segmentation method is using Laplacian filter and Luv color space so that the accuracy is not affected by the background contrast. In experiments, we checked the accuracy of

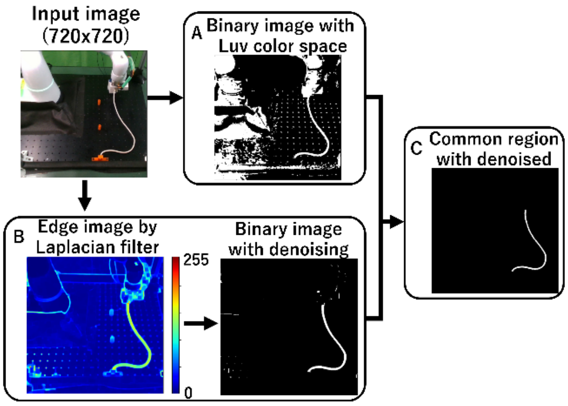


Fig.2 Flow of the cable mask image generation.

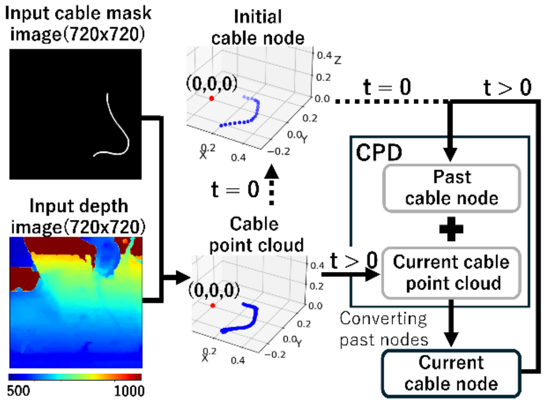


Fig.3 Node placement process for cable point clouds by CPD (D).

segmentation and the processing time of each tracking process.

2. Methodology

The system consists of cable mask image generation and cable node placement by CPD including point cloud processing. The flow of the cable mask image generation is shown in Fig. 2. The input image is an image cropped to 720x720. The method includes three steps: (A) threshold processing in the Luv color space, (B) edge processing using Laplacian filters, and (C) extracting common regions in the mask images generated by (A) and (B). After that, a node placement process (D) is applied to the cable point cloud, as shown in Fig. 3. The following is a description of each of these parts.

2.1. Binarization using Luv color space

For brightness-aware color detection, the RGB images are converted into the Luv color space with a brightness component (L) and a chromatic component (u, v). The value range of the target cable was determined by overlaying the evaluation cable mask image and the Luv color space image. The output binary image is shown in Fig. 2-A. If the value is within the threshold range, it is 255; otherwise, it is 0.

2.2. Binarization using Laplacian filter

The process generates a cable mask image of a region of edges, paying attention to the long and narrow shape of cables. The Laplacian filter used in the process is a filter that performs second-order differentiation in the image. There are Canny and Sobel filters for acquiring image edges, but Laplacian filter is used because it is fast and isotropic. In this method, a Laplacian filter with a ddepth of CV_16U and a kernel size of 7 is applied to the input image. The edge image is then grayscaled and filtered with a Blur filter. The resulting heat map is the image on the left of Fig. 2-B. After that, the greyscale image is denoised by a closing process and a thresholding process produces the binary image on the right of Fig. 2-B.

2.3. Common region extraction

In this section, a cable mask image is generated by taking the common region of two binary images and removing the noise. First, a bitwise AND operation is performed on the binary image in the LUV color space (A) and the binary image using Laplacian filter (B) to extract their common region. Then, a cable mask image shown in Fig. 2-C is generated by morphological transformation and denoising based on area.

2.4. Cable node placement

The tracking is performed by repeatedly placing the previous step's cable nodes on the current cable point cloud, as shown in Fig. 3. The cable point cloud is acquired by overlaying the depth image and the cable mask image. Before processing, the point cloud is transformed into the world coordinate system and downsampled using the Voxel Grid. As an initial process, 10 nodes are placed at constant intervals in the cable point cloud by convergence calculations. After that, the robot grasps at the 9th node coordinate of the cable and performs a tracking movement. The employed CPD method is the Bayesian coherent point drift [7] proposed by Hirose (2021). Cable fixed point is added as a constraint condition to improve tracking stability.

Table 1. Evaluation of segmentation accuracy

	Luv (A)	Laplacian (B)	Common (C)
Recall	0.981	0.918	0.891
Dice	0.076	0.550	0.813

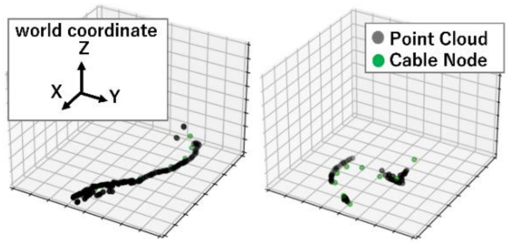


Fig. 4. Results of tracking by CPD. Left: node placement with no missing parts in the point cloud; right: node placement with missing parts in the point cloud.

Table 2. Processing time (msec)

Luv (A)	Laplacian (B)	Common (C)	CPD (D)
3.68	5.62	1.28	41.51

3. Evaluation experiments and results

3.1. Segmentation evaluation data preparation

In the evaluation experiments, 20 annotated images of cables following specific rules are used. The reason is that the method uses edge images, so evaluation with clear boundary data is suitable. The creation of evaluation data involves four steps and exception processing. First, images are taken with and without the robot holding the cable. Next, it takes the difference between the two images. Then, the Canny edge of the image with the cable is taken, and the edge is overlaid on the difference image to create an outline of the cable. Finally, the outlines are filled in. The exception is an image whose cables make a loop. In this image, the inside of the loop is also filled. Therefore, this issue is dealt with by taking a common region with the white region.

3.2. Evaluation of segmentation accuracy

The accuracy of the segmentation was evaluated by Recall: how well the correct region was covered, and Dice coefficient: how appropriate the generated region was. When evaluating accuracy, it is necessary to determine the threshold for converting the edge image to the binary image. In this study, the Dice coefficient was taken by changing the threshold value by one, and the threshold value of 113 was adopted as the maximum Dice coefficient. Table 1 shows the evaluation results of the binary images generated by each process. Compared to processes (A) and (B), the binary image from common region extraction (C) has a lower Recall score, indicating a larger unsegmented region, but a higher Dice coefficient, which means less noise and better segmentation.

3.3. Evaluation of the acceptability of tracking

Tracking is evaluated based on whether the cable node can be updated to the end for one robot movement. It was performed 20 times on a specific tracking route and in various cable conditions. The results were 95% tracking success rate. The one failed tracking attempt was the case of a large cable deformation and a large part of the cable point cloud missing. Fig. 4 shows the node placements at a certain time when the tracking was continued, respectively without and with missing parts in the point cloud. This explains that tracking is robust to deficiencies caused by occlusions and other factors.

3.4. Evaluation of processing times

Processing was implemented in Python and run on an Intel Core i7-8700K, 3.70 GHz. The results of the processing times for binary image generation (A), (B), (C) and tracking (D) without point cloud processing are shown in Table 2. All binary image generation is an average of 1000 processing times. The average processing time of the tracking process (D) by the CPD is the time taken to update the nodes in 20 tracking operations. All segmentation processes are mainly composed of thresholding and filtering, so the processing times did not widely fluctuate. On the other hand, tracking fluctuated significantly, with an average of 10.03 iterations inside the CPD and a maximum of 63 iterations.

4. Conclusion

A fast and robust for occlusion cable tracking system was proposed. In our experimental environment, there were objects with the same color as the cables, but our segmentation technique was able to extract the cable regions. In accuracy evaluation experiments, we confirmed that the accuracy was improved by overlaying the binary image by Luv and the binary image by edges. In processing time evaluation, we confirmed that all processes, including tracking by CPD, were fast. Future work is robot motion using real-time cable behavior prediction.

References

1. A. Rastegarpanah, R. Howard, and R. Stolkin, “Tracking linear deformable objects using slicing method,” Robotica, vol. 40, no. 4, 2022, pp. 1188–1206.
2. A. Keipour, M. Bandari, and S. Schaal, "Deformable One-Dimensional Object Detection for Routing and Manipulation," in IEEE Robotics and Automation Letters, vol. 7, no. 2, 2022, pp. 4329-4336.
3. D. D. Gregorio, G. Palli, and L. D. Stefano, “Let’s take a walk on superpixels graphs: Deformable linear objects segmentation and model estimation,” in Proc. Asian Conf. Comput. Vis., Springer, 2018, pp. 662–677.

4. A. Caporali, K. Galassi, R. Zanella, and G. Palli, "FASTDLO: Fast Deformable Linear Objects Instance Segmentation," in IEEE Robotics and Automation Letters, vol. 7, no. 4, 2022, pp. 9075-9082.
5. P. Kicki, A. Szymko, and K. Walas, "DLOFTBs – Fast Tracking of Deformable Linear Objects with B-splines," IEEE International Conference on Robotics and Automation (ICRA), 2023, pp. 7104-7110.
6. J. Xiang, H. Dinkel, H. Zhao, N. Gao, B. Coltin, T. Smith, and T. Bretl, "TrackDLO: Tracking Deformable Linear Objects Under Occlusion With Motion Coherence," in IEEE Robotics and Automation Letters, vol. 8, no. 10, 2023, pp. 6179-6186.
7. O. Hirose, "A Bayesian Formulation of Coherent Point Drift," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 7, 2021, pp. 2269-2286.

Authors Introduction

Mr. Ryunosuke Yamada

He received his B.S. degree in engineering from Kanazawa University, Japan, in 2023. He is currently a master's degree student in the Division of Frontier Engineering, Kanazawa University. His research interest includes simulation and robotics.

Dr. Tokuo Tsuji

He received his BS, MS, and doctoral degrees from Kyushu University in 2000, 2002, and 2005, respectively. He worked as a research fellow of Graduate School of Engineering, Hiroshima University, from 2005 to 2008. He worked as a research fellow of Intelligent Systems Research Institute of National Institute of Advanced Industrial Science and Technology (AIST) from 2008 to 2011. From 2011 to 2016, he worked as a research associate at Kyushu University. From 2016, he has been working as an associate professor at Institute of Science and Engineering, Kanazawa University. His research interest includes multifingered hand, machine vision, and software platform of robotic systems.

Mr. Takahiro Shimizu

He is a member of Intelligent Robotics R & I Division at DENSO CORPORATION. He graduated from the Graduate School of Informatics, Nagoya University in 2020. He is currently working on research and development related to the application of AI in factory automation, with a specific focus on automation of deformable object manipulation by robot arms.

Mr. Shota Ishikawa

He is the R&I Engineer of Intelligent Robotics R & I Section in AI R & I Division at DENSO CORPORATION. He received Master of Engineering degree from the Mechanical and Control Engineering, Kyushu Institue of Technology in 2013. He is currently a Doctoral course student in Kanazawa University, Japan.

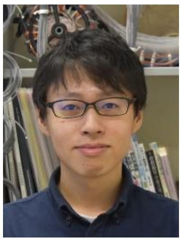
Mr. Tomoaki Ozaki

He is the Chief of Intelligent Robotics R & I Section in AI R & I Division at DENSO CORPORATION. He graduated from the Graduate School of Information Technology, Kyushu Institute of Technology in 2001. He is currently working on research and development related to the application of AI in factory automation, with a specific focus on autonomous control of robot arms.

Mr. Yusuke Sakamoto

He is an undergraduate majoring in the Division of Frontier Engineering, Kanazawa University, Japan. His research interest includes robotics.

Dr. Tatsuhiro Hiramitsu



He is assistant professor of Institute of Science and Engineering, Kanazawa University. He received Dr E. degrees from school of engineering, Tokyo Institute of Technology, Japan, in 2019. His research interest is in the soft structure mechanisms for robotic systems. He is a member of the Japan Society of Mechanical Engineers (JSME), the Robotics Society of Japan (RSJ), and Institute of Electrical and Electronics Engineers (IEEE).

Dr. Hiroaki Seki

He received his Ph.D. in precision machinery engineering from the University of Tokyo in 1996. He is currently a professor of Institute of Science and Technology in Kanazawa University. His research interests include novel mechanism and sensor system in robotics and mechatronics.

Motion Prediction for Human-Robot Collaborative Tasks Using LSTM

Kaihei Okada [1 *], Tokuo Tsuji [1 **], Tatsuhiro Hiramitsu [1], Hiroaki Seki [1], Toshihiro Nishimura [1], Yosuke Suzuki [1] and Tetsuyou Watanabe [1]

[1] Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan

*Email: *kaihei2112@stu.kanazawa-u.ac.jp, **tokuo-tsuji@se.kanazawa-u.ac.jp*

Abstract

This study proposes an assistive robot system to reduce caregiving burdens in an aging society by supporting impaired body movements. The system focuses on bimanual tasks, such as pouring a drink from a bottle into a cup. Using 3D skeletal data excluding the impaired left hand, a deep learning model (LSTM) predicts the motion stages and 3D positions of the left hand, and the robot performs the substitute motions. The system uses data from multiple users to show its potential for improving patient independence and reducing caregiver workload.

Keywords: Human motion prediction, LSTM, Three-dimensional human skeleton, Caregiving robots

1. Introduction

1.1. Background and Objective

Japan is facing a serious issue of labor shortages in caregiving and nursing due to its rapidly aging society. To address this problem, the development of assistive robots has garnered significant attention. The objective of this study is to support daily life activities for patients whose body parts are paralyzed or missing due to aging, accidents, or illnesses by utilizing assistive robots. Specifically, the focus is on tasks that require the use of both hands. For instance, in the case of a patient whose left hand is paralyzed, this study proposes a mechanism to predict the movements of the left hand using a deep learning model based on the movements of other body parts and to replicate these movements with a robot. This approach is generalizable, aiming to construct a system that can predict the movement of specific body parts (e.g., right hand, left hand, right foot, left foot) from the movements of other parts and supplement these movements with a robot. This system is expected to expand the range of possible activities for patients and reduce the burden on caregivers.

1.2. Research Targets and Challenges

This study focuses on an everyday task that requires the use of both hands: "pouring a drink from a bottle into a cup and drinking it." This task includes a sequence of motions such as opening the bottle cap, pouring the drink into the cup, drinking from the cup, and closing the cap. The movements of the left hand during these motions are classified into seven stages: stopping, approaching, grasping, holding, pouring, releasing, and leaving. However, since the motions of stopping and holding are treated as the same stage, the movements of the left hand are ultimately classified into six stages. These stages are linked to the movements of the right hand, and their cooperation is an important feature. Furthermore, as shown in Fig. 1, the task and the movement stages of the left hand are visually explained. The motions of grasping and releasing are momentary and therefore difficult to predict. To address this challenge, these motions are reformulated as a binary classification problem regarding whether the hand is open or closed. Consequently, the movement stages are treated as four multi-class classification problems (stopping, approaching, pouring, leaving), while the hand state is treated as a binary classification problem (open, close).

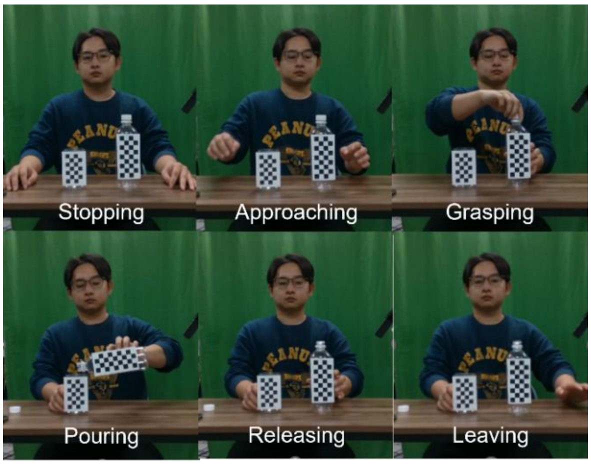


Fig. 1 Example of motion stages of left hand

1.3. Research Approach

The goal of this study is to classify the movement stages and hand states of the left hand and predict its three-dimensional movement. To achieve this, data obtained from the movements of the right hand and other body parts are utilized, and a deep learning model capable of time-series prediction is employed to estimate the movements of the left hand. Based on these estimations, the robot generates movements that replicate those of the left hand. The proposed system is applicable to many tasks required in daily life and aims to support the lives of the elderly and those with physical limitations. Additionally, the system is expected to reduce the workload of caregivers and contribute to the efficiency of caregiving settings.

2. Related Work

As the background of this research, we review technologies related to prosthetics and assistive devices, human keypoint detection technologies, and time-series prediction techniques. By organizing these topics, we aim to clarify the positioning and novelty of this study.

2.1. Prosthetics for Replacing Missing Limbs

Prosthetics come in various types, including cosmetic prostheses, work-oriented prostheses, active prostheses, and myoelectric prostheses. Among these, active prostheses and myoelectric prostheses are primarily used to assist with daily tasks.

Active prostheses use a harness attached to the shoulder or body to control movement, enabling motions such as elbow flexion and hand opening and closing. This technology allows for simple movements as well as motions requiring both hands. In contrast, myoelectric prostheses detect the weak electrical currents (myoelectric signals) generated by muscle contractions to control the opening and closing of the hand. The primary advantages of myoelectric prostheses include high gripping strength, the ability to open and close the hand without changing the prosthesis' position, and reduced discomfort due to the absence of a harness.

In recent years, advanced myoelectric prostheses have been developed. For example, the i-Limb Ultra [3], developed by Össur, features individually actuated fingers, and supports various grip patterns. However, myoelectric prostheses also face challenges such as high costs and the difficulty of learning how to operate them. This study draws inspiration from such technologies to propose a novel support method for addressing physical impairments such as limb loss or paralysis.

2.2. Assistive Devices for Supporting Paralyzed Limbs

Assistive devices for patients with paralyzed limbs have been extensively studied. For instance, assistive hands are devices designed to support the movement of paralyzed hands or arms by leveraging the patient's residual muscle strength. These devices are also used in rehabilitation programs aimed at restoring muscle strength [4].

This study aims to develop a system that can support both patients with limb loss and those with paralysis. This feature stems from the proposed hardware's independence from the extent of physical damage. In this respect, our approach differs from conventional assistive devices and prosthetics.

2.3. Keypoint Detection for Human Motion Analysis

Keypoint detection has garnered attention as a technology for analyzing human motion. OpenPose [1] is widely used for real-time detection of 2D keypoints. By detecting keypoints for the entire body, including hands, face, and torso, OpenPose has been applied in motion analysis and rehabilitation support systems.

This study utilizes OpenPose to detect the patient's body movements and predict the motion of missing or paralyzed limbs based on this data. By integrating keypoint detection technology into a novel life-support system, we aim to enhance assistive capabilities.

2.4. Time-Series Prediction with Deep Learning

In the field of motion prediction, deep learning methods for handling time-series data have been extensively researched. Notably, Long Short-Term Memory (LSTM) networks excel in learning long-term dependencies and have demonstrated high performance in time-series prediction tasks [2]. LSTM has been applied in various domains, including motion data analysis and action recognition.

This study employs LSTM to predict the motion of missing or paralyzed limbs based on movement data from healthy body parts. By doing so, we aim to accurately reproduce the patient's physical movements and enable seamless assistance by life-support robots.

3. Proposed Alternative System for Human Motion

3.1. Overview of the Proposed System

The proposed system consists of three cameras and a robot, arranged as illustrated in Fig. 2. A participant performs a task while being recorded by the cameras. Each camera detects 2D skeletal keypoints of the human. Using stereo disparity, the 3D skeletal structure is reconstructed. A deep learning model, specifically an LSTM, is employed to predict the 3D positions of the left hand and left elbow, as well as the motion stages, based on the 3D positions of other skeletal keypoints. The robot's motions are then generated based on the predictions. Fig. 3 provides an overview of the prediction method.

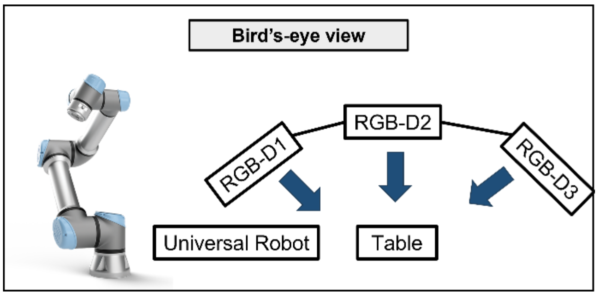


Fig. 2 Overview of the system

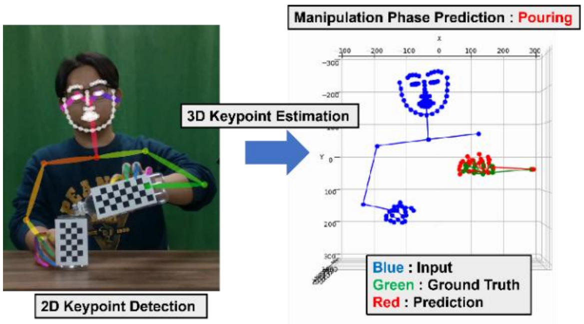


Fig. 3 Overview of the method

3.2. 2D Skeletal Detection

For videos captured by the cameras, OpenPose is used to detect human keypoints. The keypoints include 25 points for the body, 21 points for each hand, and 70 points for the face, totaling 137 keypoints. OpenPose is used because eye and hand movements are important for motion prediction. Additionally, the positions and orientations of objects are critical for the task. Checkerboard patterns are attached to objects for detection. Fig. 4 shows the detected keypoints of the human and objects during the task.

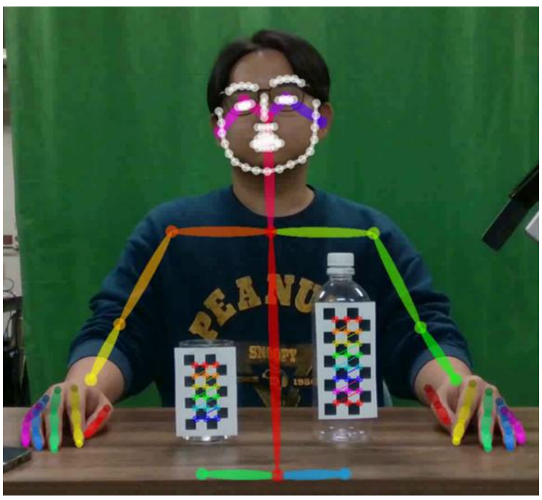


Fig. 4 Detection of keypoints and objects

3.3. 3D Skeletal Estimation

3.3.1. Stereo Calibration and Homogeneous Transformation Matrix

Camera calibration is performed to determine the projection matrices P_1 and P_2 for each camera. These matrices represent the relationship between points in 3D space and their corresponding positions in the camera image planes. In addition, the homogeneous transformation matrix H_{12} is obtained by stereo calibration. The homogeneous transformation matrix represents the relationship between cameras.

3.3.2. 3D Keypoint Estimation using DLT Method

Assume that a 3D point $X = (x, y, z, 1)^T$ is observed on the image planes of two cameras. The corresponding 2D projections are $U_1 = (u_1, v_1, 1)^T$ and $U_2 = (u_2, v_2, 1)^T$. Using the projection matrices P_1 and P_2, the relationship between these 2D points and the 3D point can be expressed as:

$$U_1 = \alpha P_1 X$$
$$U_2 = \beta P_2 X$$

where α and β are scale factors. Since the 2D points and their corresponding 3D projections are collinear, the cross product between U_1 and $P_1 X$ must be zero, which leads to a set of linear equations. Similarly, a similar set of equations is derived for the second camera. By combining these equations from both cameras, we obtain a system of linear equations. To solve for the 3D point X, Singular Value Decomposition (SVD) is used. The SVD decomposition of the matrix provides the optimal solution that minimizes the error due to observation noise. The smallest singular value corresponds to the column vector in V that represents the 3D point X. By applying this method to all keypoints across all frames, time-series data is obtained.

3.4. Time Series Prediction Using LSTM

In this study, we employ the deep learning model Long Short-Term Memory (LSTM) to accomplish the following tasks. First, we predict the left hand and left elbow keypoints and the position and posture of an object from the keypoints of the left hand and left elbow excluding those keypoints. Then, we predict the motion stage and the state label of the left hand. The following symbols are defined for this study. L_t represents the left hand and left elbow keypoints at frame t, R_t represents the keypoints of the left hand and left elbow excluding these at frame t, O_t represents the position and posture of an object, P_t represents the motion stage label of the left hand, and H_t represents the state label of the left hand.

Using information from up to k frames before the current frame, the left hand and left elbow keypoints, L_t can be predicted using the following model:

Tabel 1 Metrics for validation data

Metrix	RMSE[mm]	Acc(Motion)[%]	F1(Motion)	Acc(Hand)[%]	F1(Hand)
Keypoint Model	52.3	N/A	N/A	N/A	N/A
Label Model	N/A	0.759	0.697	0.820	0.817
Keypoint Label Model	**40.5**	**0.906**	**0.819**	**0.959**	**0.958**

$$L_t = LSTM(R_{t-k+1}, \dots, R_t)$$

Furthermore, if the initial position and posture of the object, O_0, is known, we predict the values for the next frame, as follows:

$$L_t, O_{t+1} = LSTM(R_{t-k+1}, \dots, R_t, O_{t-k+1}, \dots, O_t)$$

Additionally, we consider a model that predicts P_t and H_t. This model can be expressed as follows:

$$L_t, O_{t+1}, P_t, H_t = LSTM(R_{t-k+1}, \dots, R_t, O_{t-k+1}, \dots, O_t)$$

Fig. 5 shows the structure of the LSTM model.

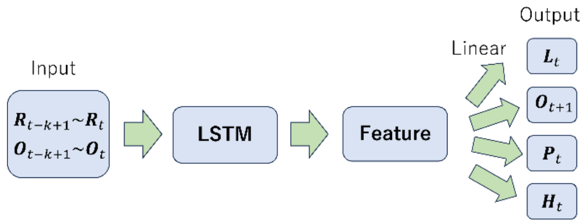


Fig. 5 Structure of our model

Furthermore, models that predict only L_t and O_{t+1} are referred to as the "**Keypoint Model**," those that predict only P_t and H_t are the "**Label Model**," and models that predict all outputs are referred to as the "**Keypoint Label Model**." Results from these models will be compared.

4. Experiment and evaluation

4.1. Experimental Setup and Dataset Creation

In this study, three RGB-D cameras (RealSense D435) were utilized, focusing exclusively on color images. Additionally, a robot arm (Universal Robot e5) was employed to assist with tasks. During data collection, the participants were seated on a chair and instructed to perform a series of tasks on a table: opening a bottle cap, pouring water into a cup, drinking water from the cup, and closing the cap. Each sequence of tasks generated one set of time-series data. A total of 70 datasets were recorded from five participants.

The collected data was split into 90% for training and 10% for validation. The time-series model was trained using the training data, and its performance was evaluated using the validation data. This evaluation aimed to quantitatively demonstrate the effectiveness of the proposed system.

4.2. Model Training

For model training, the loss function for 3D keypoint position estimation was defined as Mean Squared Error (MSE) Loss, while Multi-Class Cross Entropy Loss was used for task stage classification. The total loss function was weighted as follows:

$$Loss = 100 \times MSE\ Loss + Cross\ Entropy\ Loss$$

The Adam optimizer was employed for training, with the number of epochs set to 100. This configuration enabled efficient and effective model training.

4.3. Evaluation Metrics for Validation Data

Table 1 summarizes the average results for all validation data. The comparison includes the model predicting only the 3D positions of keypoints, the model predicting only motion stage and hand state labels, and the model predicting both (referred to as the **Keypoint Label Model**). The 3D positions of keypoints are evaluated using RMSE, while motion stage and hand state labels are evaluated using accuracy and F1 Score.

As shown in Table 1, the **Keypoint Label Model** achieved the best performance across all metrics. This result suggests that predicting labels enhances the accuracy of keypoint predictions and improving keypoint predictions also benefits label prediction accuracy. This synergy is attributed to the ability of the LSTM to effectively learn features from sequential data that are relevant for predicting both keypoints and labels. Since keypoints and labels are interrelated, they provide valuable information for each other’s predictions. Moreover, the multitask learning approach enhanced the generalization capability of the model, mitigating the risk of overfitting. These findings align with previous research on multitask learning, demonstrating its effectiveness in improving overall model performance.

4.4. Confusion Matrices for Validation Data

Fig. 6 presents the confusion matrices for hand state and motion stage label predictions, respectively, generated by the **Keypoint Label Model**. The hand state labels are predicted with high accuracy. Similarly, the motion stage labels are also predicted with high accuracy overall, although some labels exhibit lower performance. Specifically, the model tends to misclassify the "approaching" stage as the "stopping" stage and the "leaving" stage as the "stopping" stage.

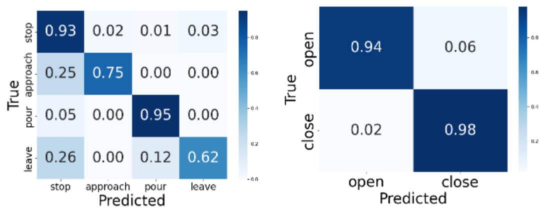


Fig. 6 The confusion matrix of Motion stages and hand state

To investigate these misclassifications, the temporal flow of labels was visualized. Fig. 7 shows the ground truth labels and predictions of motion stages over time for Validation Data 2. Around frame 350, it can be observed that the transition to the "approaching" stage occurs prematurely, leading to the misclassification of the "approaching" stage as the "stopping" stage. Similarly, around frame 200, the preceding "stopping" stage causes misclassifications in the subsequent "leaving" stage. These observations suggest that improving the timing of stage transitions could enhance prediction accuracy.

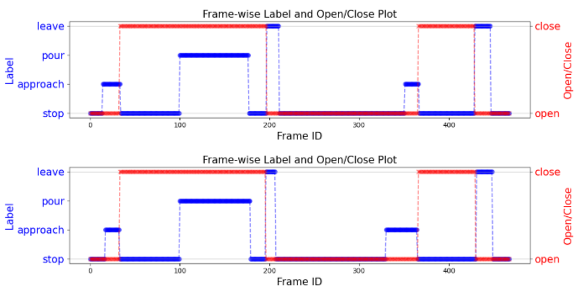


Fig. 7 The top figure shows the ground truth, while the bottom figure represents the predictions, each plotting the flow of the motion stages.

4.5. 3D Keypoint Prediction for Validation Data

Fig. 8 visualizes sampled predictions of 3D keypoints over time made by the **Keypoint Label Model**. The blue points represent the input, green points represent the ground truth, and red points represent the predictions. Although the actual predictions involve 3D positions, they are projected to 2D for visualization purposes.

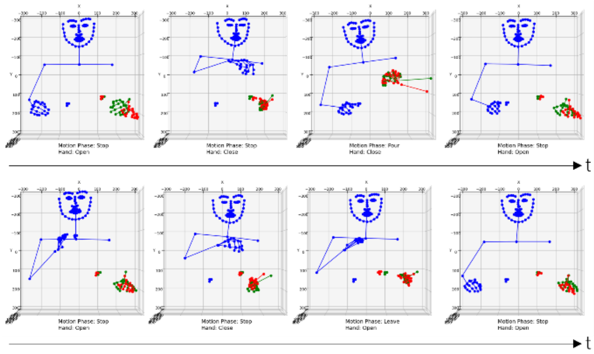


Fig. 8 The predicted 3D human skeleton

As shown in Fig. 8, the predicted keypoints closely match the ground truth, demonstrating the high accuracy achieved by the **Keypoint Label Model** in 3D keypoint prediction. However, certain keypoints, such as the initial position of the left hand or the elbow position during pouring motions, are inherently difficult to predict accurately, leading to lower performance for these specific cases.

These results highlight the strengths of the **Keypoint Label Model** in simultaneously predicting keypoints and labels with high accuracy, while also identifying potential areas for further improvement in challenging scenarios.

5. Conclusion

In this study, we utilized an LSTM-based model to predict the 3D positions of the left hand and elbow, as well as the motion stages, using data from regions other than the left hand and elbow. For the experiments, a dataset of 70 samples was created using five participants. Evaluation results on the validation data demonstrated that the model predicting both the 3D positions and motion stages simultaneously achieved higher accuracy compared to models predicting them separately. In future work, we aim to investigate the model's capability to handle an increased variety of motion types. Additionally, we plan to evaluate the generalization performance of the model on data from individuals outside the current dataset.

References

1. Hochreiter, S. (1997). Long Short-term Memory. Neural Computation MIT-Press.
2. Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7291-7299).
3. Ossur, "i-Limb Ultra." Accessed: Dec. 26, 2024. [Online]. Available: https://www.ossur.com/en-gb/prosthetics/arms/i-limb-ultra

4. H. Ishisaki, K. Tabata, T. Tsuji and T. Watanabe, "Hand Grasping Assist Glove Combining Exoskeleton Structure and Pneumatically Driven Actuator," 2023 IEEE/SICE International Symposium on System Integration (SII), Atlanta, GA, USA, 2023, pp. 1-5, doi: 10.1109/SII55687.2023.10039272.

Authors Introduction

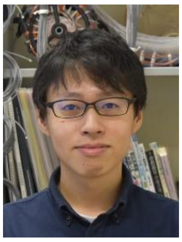
Mr. Kaihei Okada

He received his B.S. degree in engineering from Kanazawa University, Japan, in 2023. He is currently a master's degree student in the Division of Frontier Engineering, Kanazawa University. His research interest includes robot vision and life support robots.

Dr. Tokuo Tsuji

He received his BS, MS, and doctoral degrees from Kyushu University in 2000, 2002, and 2005, respectively. He worked as a research fellow of Graduate School of Engineering, Hiroshima University, from 2005 to 2008. He worked as a research fellow of Intelligent Systems Research Institute of National Institute of Advanced Industrial Science and Technology (AIST) from 2008 to 2011. From 2011 to 2016, he worked as a research associate at Kyushu University. From 2016, he has been working as an associate professor at Institute of Science and Engineering, Kanazawa University. His research interest includes multifingered hand, machine vision, and software platform of robotic systems.

Dr. Tatsuhiro Hiramitsu



He is assistant professor of Institute of Science and Engineering, Kanazawa University. He received Dr E. degrees from school of engineering, Tokyo Institute of Technology, Japan, in 2019. His research interest is in the soft structure mechanisms for robotic systems. He is a member of the Japan Society of Mechanical Engineers (JSME), the Robotics Society of Japan (RSJ), and Institute of Electrical and Electronics Engineers (IEEE).

Dr. Hiroaki Seki

He received his Ph.D. in precision machinery engineering from the University of Tokyo in 1996. He is currently a professor of Institute of Science and Technology in Kanazawa University. His research interests include novel mechanism and sensor system in robotics and mechatronics.

Dr. Toshihiro Nishimura

He received the B.S., M.S., and Ph.D. degrees in mechanical engineering from Kanazawa University, Kanazawa, Japan, in 2016, 2018, and 2019, respectively. He was a Researcher of industrial robots with FANUC Corporation, from 2018 to 2021. He is currently an Assistant Professor with the Faculty of Frontier Engineering, Institute of Science and Engineering, Kanazawa University. His research interests include robotic hands, mechanism design, soft robotics, and 3D printer.

Dr. Yosuke Suzuki

He received the B.Eng., M.Eng., and Ph.D. degrees in engineering from the Tokyo Institute of Technology, Tokyo, Japan, in 2005, 2007, and 2010, respectively. He is currently an Associate Professor with the Faculty of Mechanical Engineering, Institute of Science and Engineering, Kanazawa University, Kanazawa, Japan. His research interests include tactile and proximity sensors, robotic grasping, and distributed autonomous systems.

Dr. Tetsuyou Watanabe

He is a professor with Kanazawa University. He received the B.S., M.S., and Dr.Eng. degrees in mechanical engineering from Kyoto University, Kyoto, Japan, in 1997, 1999, and 2003, respectively. From 2003 to 2007, he was a Research Associate with the Department of mechanical Engineering, Yamaguchi University, Japan. From 2007 to 2011, he was an assistant professor with Division of Human and Mechanical Science and Engineering, Kanazawa University. From 2011 to 2018, he was an associate professor with Faculty of Mechanical Engineering, Institute of Science and Engineering, Kanazawa University. Since 2018, he has been a professor with Kanazawa University. From 2008 to 2009, he was a visiting researcher at Munich University of Technology. His current research interests include robotic hand, grasping, object manipulation, medical and welfare sensors, surgical robots, and user interface. He got several awards including best paper award at Transactions of the Society of Instrument and Control Engineers and World Robot Summit Second Prize of World Robot Challenge Industrial Robotics Category Second Prize of World Robot Challenge Industrial Robotics Category.

Individual recognition of food in bulk by using 3D model of food

Yuya Otsu [*], Tokuo Tsuji [**], Tatsuhiro Hiramitsu, Hiroaki Seki
Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan

*Email: *otsuy0525@stu.kanazawa-u.ac.jp, **tokuo-tsuji@se.kanazawa-u.ac.jp*

Abstract

In this paper, we propose a method of individual recognition of food in bulk by using 3D model of food. First, color images and depth images of them are generated by using 3D model of food and physics engine of simulator. Then, color and depth composite images are created by converting two channels from color images and one channel from depth images. In the experiments, the accuracy of individual recognition of food in bulk with color and depth composite images are shown to compare the accuracy with only color images.

Keywords: Instance segmentation, Color space, 3D model

1. Introduction

In the food manufacturing industry, individual recognition of food in bulk on an image plays a crucial role in quality control and determining the gripping positions for topping handlers. Instance segmentation has been proposed as a method of image recognition and it generally uses supervised learning. In supervised learning, it is necessary to prepare many training data to improve learning accuracy. And, since training data is usually prepared manually, it is burdensome to prepare many training data. Therefore, as an efficient training data generation method for food recognition, such as generating composite images of food by combining multiple images of individual food [1][2] have been proposed. However, accurate segmentation of individual's region with only RGB image is difficult in the case of single type food because the boundaries between individual food become ambiguous when it is stacked in bulk.

To solve this issue, we propose a method to automatically generate many color images and depth images of food in bulk in a short time using 3D models of food and physics engine of simulator, and to generate composite image that include both color and depth information for individual recognition of food in bulk. Firstly, food models captured by a 3D scanner are placed on a virtual space and obtain color image and depth image by shooting from a virtual viewpoint. Then, RGB channels of the color image are converted to the Luv color space, and the two channels (u and v) are reduced to a single channel using principal component analysis (PCA). Finally, color and depth composite images are created by combining L channel, the new channel obtained PCA, and the single channel of depth image.

In this study, we propose the individual recognition of food in bulk with color and depth composite images created using 3D model of food. In experiments, we confirmed the accuracy of individual recognition of food in bulk with color and depth composite images to compare the accuracy with only color images.

2. Methodology

The flow of the color and depth composite image generation is shown in Fig. 1. The RGB channels of the color image generated by using 3D models of food in virtual space are converted to the Luv color space. The channel obtained by reducing the dimensionality of the two channels, u and v, is referred to as C, the depth image generated by using the distance information from the camera to each food is referred to as D, and the combined channels of color and depth are referred to as LCD.

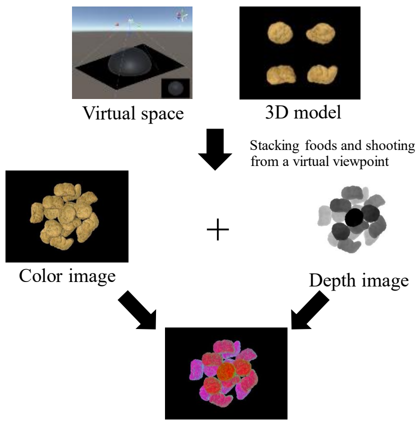


Fig. 1 Flow of color and depth composite image generation.

2.1. Generating 3D model of food by 3D scanner

In this study, we assume a chicken nugget stacked and create 3D models of four different shapes of chicken nuggets. A SCANDIMENTION SOL 3D scanner [3] (Fig. 2) is used for this purpose. Each chicken nugget is placed on the scanner's rotating platform and rotated inside a light-shielding tent while being scanned with a laser to capture the mesh and texture. The 3D models of the chicken nugget obtained are shown in Fig. 3.

Fig. 2 SOL 3D scanner [3]

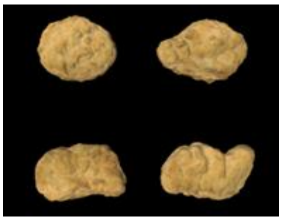


Fig. 3 3D models of chicken nuggets

2.2. Generating color image using virtual space

Color image of food in bulk is generated by arranging the 3D models of food in a virtual space using a physics engine of simulator and capturing them from virtual viewpoints. In this study, Unity [4] is used as a virtual space. A transparent hemispherical mold is placed in the Unity space. 3D models of food are randomly generated within the mold and stacked in bulk along the mold. An example of the Unity space used for stacking food and the color image generated is shown in Fig. 4.

2.3. Generating depth image

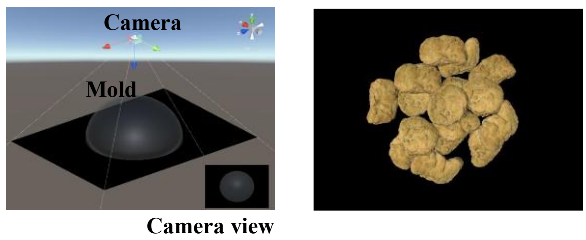
After stacking the food in the Unity space, the distance information from the camera to each food is obtained. The acquired values are normalized by setting the distance from the camera to the top of the mold as 0 and the distance from the camera to the ground as 1. Then the normalized values are applied to the materials of each food and the scene is captured to generate a depth image. An example of depth image generated is shown in Fig. 5.

2.4. Dimensionality reduction method for color channel

The RGB channels of the color image generated in section 2.2 are converted to the Luv color space. The Luv color space is defined to equalize color differences and is expressed using a three-dimensional orthogonal coordinate system with L (lightness), u, and v (chromaticity) as its axes [5]. After converting to Luv, the two channels, u and v, are reduced to a single channel using principal component analysis, resulting in the C channel. Fig. 6 shows the uv color map of the color image and the results of dimensionality reduction. Also, Fig. 7 shows the grayscale image with the L channel and the C channel obtained by dimensionality reduction.

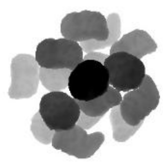
2.5. Compositing color and depth channels

By combining the L and C channels generated in Section 2.4 with the D channel generated in Section 2.3, a composite image of color and depth (LCD image) is obtained. An example of the generated LCD image is shown in Fig. 8.



Unity space　　　　　Color image

Fig. 4 Generate method of color image



Depth image

Fig. 5 Example of depth image of food in bulk

3. Food recognition using composite images as training data

This section describes the specific details of the instance segmentation used for food recognition and the generation of training data using composite images.

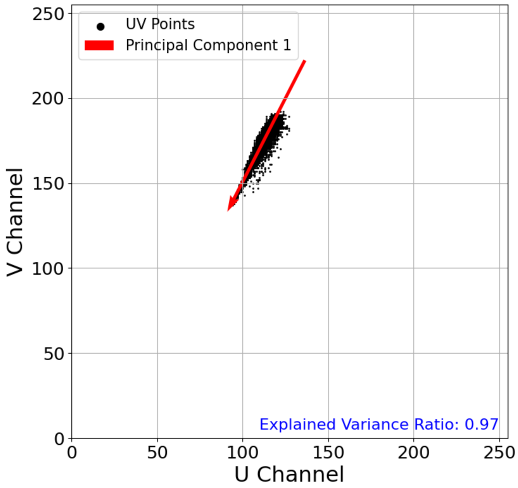
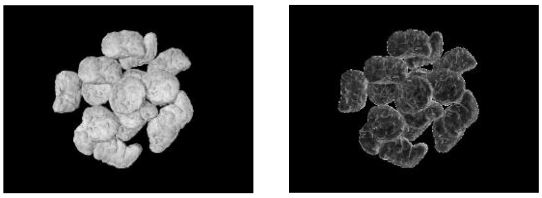
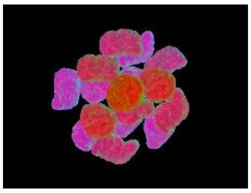


Fig.6 Example of uv color map and result of principal component analysis



L channel image C channel image

Fig.7 Example of L channel image and C channel image (gray scale)



LCD image

Fig.8 Example of color and depth composite image

3.1. Instance segmentation

Instance segmentation is a process of detecting the class of each object and extracting its region pixel by pixel. The machine learning used in this study is supervised learning. It is a method that uses pre-labeled input data for training and builds a model that can predict labels for unknown input data. In this study, we use YOLO-v7[6], a deep learning model for object detection and segmentation.

3.2. Generation of training datasets using composite images

In the dataset for instance segmentation used in this study, the area information of each object in the image called the COCO format [7] is often represented by the contour information of them. In this method, 3D models of food are stacked in the Unity space, and a mask image is generated for each food models by changing the texture of them to white and that of the background to black. An example of a mask image corresponding to the color image is shown in Fig. 9.

Color image Mask image

Fig.9 Example of color image and mask image

4. Evaluation experiments and results

4.1. Experiment Preparation

4.1.1 Types of food for experiment

Assuming a stacking of chicken nuggets in bulk, four chicken nuggets of four different shapes each are prepared.

4.1.2 Capture of validation images (color images and depth images)

The camera was an Intel RealSense D435, 50 cm away from the ground, under fluorescent light, and with forward light. 10 images were taken as validation images. As in section 2.3, the depth images are normalized by setting the distance from the camera to the top of the stack to 0 and the distance from the camera to the ground to 1, and the depth images are generated in grayscale. An example of the image of a chicken nugget stacked and the normalized depth image are shown in Fig. 10.

4.1.3 Generation of training images (color images)

The same number of 3D models of food used to take the validation images are randomly arranged in a virtual space and the color image is generated by shooting from the same point of view as the location where the validation image was taken.

4.1.4 Creation of COCO-format dataset for validation

Since the validation images are photographs and cannot be used with the method in 3.2, the annotation tool Roboflow [8] is used to create the dataset.

Color image Depth image

Fig.10 Example of color image and depth image for test data

4.2. Experimental conditions

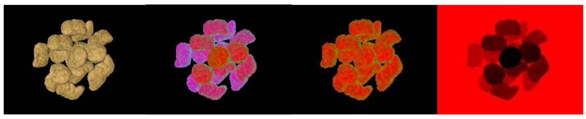
For the stacked chicken nuggets, learning is performed using the following four types of composite images as the training data. The learning conditions are summarized in Table 1. For the LC images and D images, a 3-channel image with 0 in the empty channels is used as the training data. An example of the datasets used in each learning for the experiment are shown in Fig. 11 and Fig. 12. The common training conditions are: model: YOLO-v7[6], number of training data: 500, number of validating data: 10, number of epochs: 300, batch size: 16

4.3. Evaluation method

The value of AP (Average Precision) when IoU (Intersection over Union) is set as the threshold is used to evaluate the performance of each learning model. IoU is a value between 0 and 1 that represents the degree of overlap between the predicted region of an object and the ground truth region. And, AP is the integral value of the P-R curve, which plots Precision against Recall.

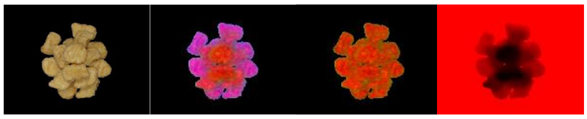
Table 1. Each learning condition

	Training data and validation data
Learning A	RGB images
Learning B	LCD images
Learning C	LC images
Learning D	D images



RGB LCD LC D

Fig.11 Example of training data



RGB LCD LC D

Fig.12 Example of validation data

4.4. Experimental results and discussion

The results of AP for each learning are shown in Fig. 13. AP (50) represents the AP at an IoU threshold of 0.5, AP (50-95) shows the average of AP when the IoU threshold is varied from 0.5 to 0.95. Also, for learning A and B, a part of the model's predictions on the validation images is shown in Fig. 14. Fig. 13 shows that the model learned on the LCD image is 8.8% more accurate than that on the RGB image for AP (50), confirming that the model learned with the LCD image is more accurate than that with the RGB image. Similarly, for AP (50-95), the LCD image was 15.4% more accurate than the RGB image, confirming that the model learned on the LCD image was more accurate than the model learned on the RGB image, even when the IoU threshold was increased.

Also, Fig. 14 shows that the output from the model learned with the RGB images segmented the overlapping parts as the same individual, whereas the output from the model learned with the LCD images improved the segmentation accuracy. It is thought that the difference in depth caused by the overlapping of foods was learned by the LCD image, which improved the recognition accuracy. These results confirm the effectiveness of using color and depth composite images as training data for individual recognition of a single type of food in bulk.

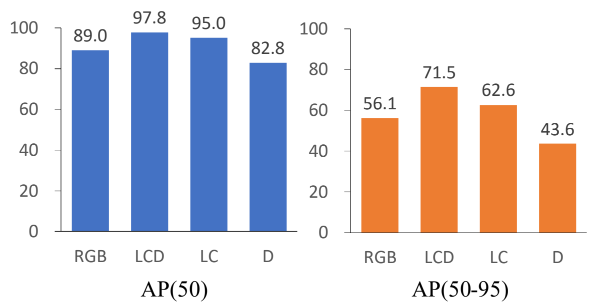


Fig.13 Result of AP for each learning condition

Fig.14 Result of segmentation for learning condition A and B (part of food)

5. Conclusion

In this study, we proposed a method of generating a lot of color image and depth image of food in bulk automatically by using 3D models of food and generating color and depth composite image. And the effectiveness of using the composite images as training data was demonstrated by using them as training data for instance segmentation.

References

1. T. Yamabe, T. Ishichi, T. Tsuji, T. Hiramitsu, and H. Seki, "Training Data Augmentation for Semantic Segmentation of Food Images Using Deep Learning," International Conference on Artificial Life and Robotics, 2022
2. T. Ishichi, T. Yamabe, T. Tsuji, T. Hiramitsu, and H. Seki, "Ingredient segmentation with transparency," IEEE/SICE International Symposium on System Integration (SII), 2023, pp. 1-5.
3. "SCANDIMENSION,SOL3DScanner,"https://scandimension.com/products/sol-3d-scanner
4. "Unity," https://unity.com
5. Y. Ohno, "CIE fundamentals for color measurements," In NIP & Digital Fabrication Conference, Society of Imaging Science and Technology, vol. 16, 2000, pp. 540-545.
6. C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 7464-7475.
7. T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, ... and C. L. Zitnick, "Microsoft coco: Common objects in context," In Computer Vision–ECCV 13th European Conference, vol. 5, 2014, pp. 740-755
8. "Roboflow,"https://roboflow.com

Authors Introduction

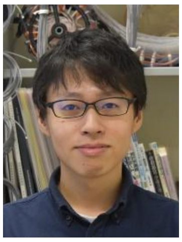
Mr. Yuya Otsu

He received his B.S. degree in engineering from Kanazawa University, Japan, in 2023. He is currently a master's degree student in the Division of Frontier Engineering, Kanazawa University. His research interest includes food recognition with machine learning

Dr. Tokuo Tsuji

He received his BS, MS, and doctoral degrees from Kyushu University in 2000, 2002, and 2005, respectively. He worked as a research fellow of Graduate School of Engineering, Hiroshima University, from 2005 to 2008. He worked as a research fellow of Intelligent Systems Research Institute of National Institute of Advanced Industrial Science and Technology (AIST) from 2008 to 2011. From 2011 to 2016, he worked as a research associate at Kyushu University. From 2016, he has been working as an associate professor at Institute of Science and Engineering, Kanazawa University. His research interest includes multifingered hand, machine vision, and software platform of robotic systems.

Dr. Tatsuhiro Hiramitsu



He is assistant professor of Institute of Science and Engineering, Kanazawa University. He received Dr E. degrees from school of engineering, Tokyo Institute of Technology, Japan, in 2019. His research interest is in the soft structure mechanisms for robotic systems. He is a member of the Japan Society of Mechanical Engineers (JSME), the Robotics Society of Japan (RSJ), and Institute of Electrical and Electronics Engineers (IEEE).

Dr. Hiroaki Seki

He received his Ph.D. in precision machinery engineering from the University of Tokyo in 1996. He is currently a professor of Institute of Science and Technology in Kanazawa University. His research interests include novel mechanism and sensor system in robotics and mechatronics.

Modeling Yawning Contagion as a Reaction-Diffusion System: Emergence of Turing Patterns in Behavioral Contagion

Yasuhiro Suzuki
Graduate School of Informatics, Nagoya University, Furocho,
Chikusa, Nagoya 464-8601, Japan
E-mail: ysuzuki@i.nagoya-u.ac.jp
https://ysuzuki.info

Abstract

When we see someone yawning, we often feel compelled to yawn ourselves - a phenomenon known as behavioral contagion in psychology. While one person's yawn acts as an activator that triggers yawns in others, we sometimes suppress the urge to yawn in situations like meetings, representing an inhibitor of this behavior. We formulated this yawning contagion as a reaction-diffusion phenomenon in an activator-inhibitor system and confirmed the emergence of Turing patterns. Our findings provide a theoretical framework for understanding and potentially controlling the spread of social behaviors in human populations.

Keywords: Reaction-Diffusion, Behavioral contagion, Yawning

1. Introduction

In this study, we use the "contagious yawning phenomenon [1]" as a starting point to analyze the phenomenon using a mathematical framework known as the reaction-diffusion phenomenon. The reaction-diffusion phenomenon is a model that forms the basis of a wide range of natural phenomena, such as chemical reactions and biological morphogenesis, and its importance is particularly demonstrated in Alan Turing's "Theory of Morphogenesis [2]." We uses the reaction-diffusion phenomenon to elucidate the mechanism of contagious yawning using a mathematical model and to open the possibility of applying it to natural and social phenomena.

When you see someone yawning, you are likely to yawn as well. In psychology, this reaction-diffusion phenomenon is called behavioral contagion [1]. Someone is yawning, which spreads and induces someone else to yawn. Factors that activate things like this are called activators.

On the other hand, when someone yawns in a meeting, and you are compelled to yawn, you might think, "I should not yawn here..." and hold back from yawning. Things that inhibit actions or reactions are called inhibitors. We describe the behavioral contagion of yawning as an activator and an inhibitor.

First, when someone is yawning, behavioral contagion makes you want to yawn, too. In other words, someone's yawning becomes an activator, making you want to yawn. Yawning induces yawning through behavioral contagion, so this is written as "yawn -> (+) yawn." Here, -> (+) indicates the reaction of an activator (a reaction that activates the behavioral contagion of yawning).

On the other hand, if someone is yawning, a reaction occurs to try to hold back to avoid being influenced. In this reaction, the "yawn" first activates the inhibitory factor "hold back a yawn." This is written as "yawn -> (+) hold back a yawn." When the inhibitory factor "hold back a yawn" is activated, it suppresses yawning, so this is written as "hold back a yawn -> (-) yawn." The greater the influence of this reaction, the more yawning is suppressed, and the less likely behavioral contagion will occur.

Alan Turing was one of the pioneers who focused on reaction-diffusion phenomena to understand natural phenomena. One of the biochemical reactions that interested him was the morphogenesis of living organisms. When an egg is fertilized, an organism takes shape through repeated cell division from a single fertilized egg. This process is called morphogenesis.

The shape of a fertilized egg is initially symmetrical. Still, after about a week, when the internal organs begin to form, the position and shape of the internal organs become asymmetrical. Turing was intrigued by the fact that the symmetry of the fertilized egg eventually breaks down, and various shapes emerge.

He formulated this phenomenon by considering it as a reaction-diffusion system. He then used the reaction-diffusion system to develop the change in the shape of a zygote, which has a symmetrical shape, to an asymmetrical shape as a change from a symmetric state, where the state is the same everywhere, to an asymmetric state, where the state differs depending on the location.

2. method

A reaction rule models the activation/inhibition system. The activation factor increases by X an autocatalytic reaction; that is, the amount of the activation factor X increases from X;

$$r_1: X \underset{k_1}{\rightarrow} X, X.$$

Below we denote $k_1, \ldots, k_4$ as reaction coefficient. The inhibitory factor Y is produced from an activator X as follows:

$$r_2: X \underset{k_2}{\rightarrow} Y, X.$$

The inhibition of X by inhibitor Y is:

$$r_3: X, Y \underset{k_3}{\rightarrow} Y,$$

Note that in this reaction, Xonly decreases, $Y$and does not decrease. Finally Y, the disappearance is

$$r_4: Y \underset{k_4}{\rightarrow} \emptyset,$$

where $\emptyset$ denotes the empty set. To summarize the above, the activation/inhibition system is based on the rewriting system [3], as follows:

$$r_1: X \underset{k_1}{\rightarrow} X, X,$$
$$r_2: X \underset{k_2}{\rightarrow} Y,$$
$$r_3: X, Y \underset{k_3}{\rightarrow} Y, X$$
$$r_4: Y \underset{k_4}{\rightarrow} \emptyset.$$

2.1. Diffusion model

The above system does not include diffusion. Therefore, diffusion needs to be considered separately. Why does diffusion occur? If you drop a drop of ink onto clean water and stare at it for a long time, the ink that spreads across the surface will eventually dissolve into the water. No matter how long you stare at the water's surface with the ink dissolved, the ink will never return to its state when it is dropped back onto the water. In other words, diffusion is the process of homogenizing things.

The diffusion occurs between the cells x_i and x_j is $d = (x_i - x_j) \times D$, where D is the diffusion coefficient.

3. Result

Here, we consider a two-variable system consisting of 10 cells. However, the right side of the rightmost cell is connected to the leftmost cell. In other words, we consider the 10 cells connected in a ring shape. Then, we define $X$the initial state of and as follows:Y

$$X = [0, 10, 0, 10, 0, 10, 0, 10, 0, 10]$$
$$Y = [0, 10, 0, 10, 0, 10, 0, 10, 0, 10]$$

In each cell, the reaction proceeds according to the reaction rules the reaction coefficients are $k_1 = k_2 = k_4 = 0.001$and $k_3 = 0.005$. In the simulation, the state quantity is first updated by performing diffusion. Next, the state quantity is updated by applying the reaction rules in parallel. Then, the state quantity is updated by diffusion and then by the reaction rules, and this process is repeated.

The diffusion coefficient with Y, $D\_X = D\_y = 0.0$, there will be no change from the initial state. Such a state is called an equilibrium state (equilibrium means "balanced"); in this reaction system, if the state $X = Y$ quantities of Y(and X) are equal, then neither $X$the state quantities nor $Y$the state quantities change.

Therefore, since the initial state holds in the corresponding cells $X = Y$, if there is no diffusion, it will remain in the initial state.

What happens if diffusion occurs when $D_x = D_y = 0.01$, the reaction is in equilibrium (state of equilibrium)? First, $X = Y$if we Xassume that the diffusion coefficients of $X$and are the same, then while $Y$and Ychange, they always $X = Y$ remain the same. In other words, the reaction is always in equilibrium. Eventually, diffusion Xwill Yhomogenize and the whole will reach equilibrium. This means that the diffusion coefficient is $D_x = D_y = 0.2, 0.3, \cdots$. Next $D_x = 1.0$ and $D_y = 0.001$, the diffusion coefficient Y for X, is more significant than $X$and Y are homogenized. On the other hand, if $D_x = 0.01$ and $D_y = 1.0$, then Turing Pattern like behavior appeared (Figure 1).

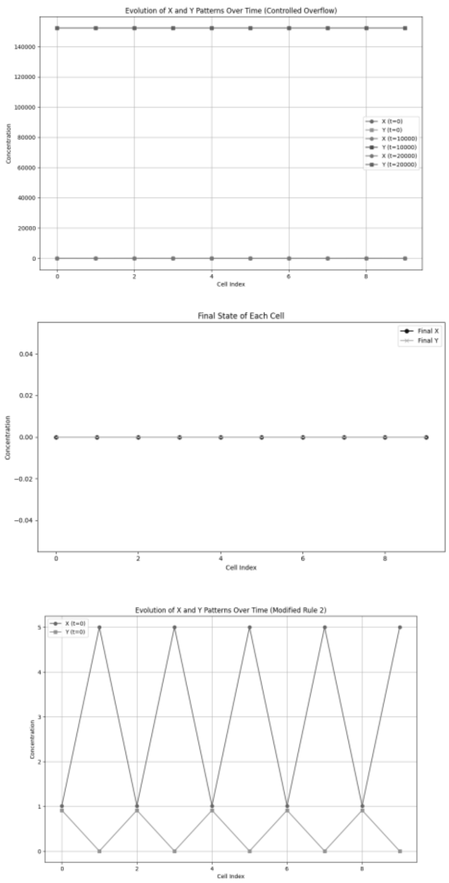


Figure 1 (Top) $Dx = Dy = 0.01$ (Middle) *Dx=1.0 > Dy=0.01* (Bottom) $Dy = 1.0 > Dx = 0.001$.

4. Discussion

Based on the preparations made in the previous Section, we can now specifically consider why a system's behavior in equilibrium changes due to diffusion.

$X$the diffusion coefficient of is Ylarger, the increase of is suppressed in the cell with the initial state of $(10{,}10)$, $X$and tends to increase in the adjacent cells $Y > X$due to diffusion. Eventually, every cell $X > Y$ becomes, $Y < X$and Xif there is no outflow of the state quantity from the cell, Xincreases while maintaining a homogeneous state.

X the diffusion of is Y slower than that of, then $(10{,}10)$ the amount of diffusion of $X$is greater $Y < X$ in the cell with the initial state of, $Y$and X increases. Since the diffusion of Y is faster, increases in the adjacent cells $Y > X$, and decreases. As a result, increases Xin the cell with X the initial state of, and decreases because the $(10{,}10)$of the adjacent cells X is suppressed. As a result, pattern of Y become larger and small quantities of X is formed.

This way, patterns can emerge by changing the diffusion rate even in a relatively simple reaction-diffusion system. In general, Turing pattern is stable for both time and space, however this model exhibits unstable for both (Figure 2).

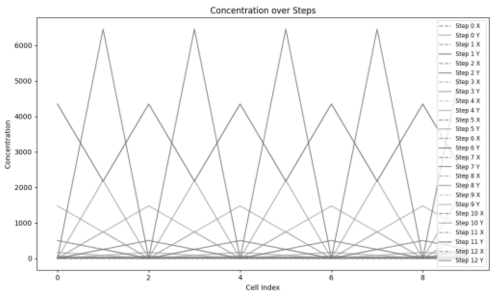


Figure 2. When Dy >> Dx, Turing pattern like behaviors appeared, however, the patterns are unstable and expanding as the time step increased.

As mentioned above, Turing discovered this mechanism and demonstrated it mathematically. The patterns that emerge from this mechanism are called Turing patterns.

After the proposal of Turing patterns, this mechanism was thought to be abstract, but Shigeru Kondo showed that the patterns on the surface of animals' bodies are Turing patterns.

The stripes of zebras and giraffes are fixed on the surface of their bodies. However, the spacing between the patterns on their body surface does not increase as they grow. Kondo wondered, "Perhaps the number of stripes increases as they grow?" Then, by observing a tropical fish (the imperial angelfish) that fulfills this condition, he showed that the stripes form a Turing pattern (Figure 3).

Figure 3. The Turing pattern on the surface of a zebra's body (Photograph by author)

Acknowledgements

This work was supported in part by KAKENHI 21K12108, 22H01444 and 23K06883.

References

1. PLATEK, Steven M.; MOHAMED, Feroze B.; GALLUP JR, Gordon G. Contagious yawning and the brain. Cognitive brain research, 2005, 23.2-3: 448-452.
2. Turing, Alan Mathison. "The chemical basis of morphogenesis." Bulletin of mathematical biology 52 (1990): 153-197.
3. Suzuki, Yasuhiro, and Hiroshi Tanaka. "Symbolic chemical system based on abstract rewriting system and its behavior pattern." Artificial Life and Robotics 1 (1997): 211-219.

Authors Introduction

Dr. Yasuhiro Suzuki

He received his Doctor degree in Informatics in 2001 from Graduate School of Informatics, Kyoto University, JAPAN. He is currently an associate professor of Graduate School of Informatics, Nagoya University, JAPAN

Dominant Region Analysis: A Novel Framework for Quantifying Competitive Reactions Based on the Gillespie Algorithm

Yasuhiro Suzuki
Graduate School of Informatics, Nagoya University, 1583 Furocho,
Chikusa, Nagoya 464-8601, Japan
E-mail: ysuzuki@i.nagoya-u.ac.jp
http://www.ysuzuki.info/

Abstract

Understanding the quantitative relationships between competing reactions is crucial for analyzing chemical reaction systems. While conventional approaches often focus on static analysis, we propose a novel concept called "dominant region" to capture the dynamic nature of reaction competition. The dominant region concept can be viewed as an extension of the traditional rate-determining step in reaction kinetics. This enables quantitative prediction of how dominant reactions dynamically change with variations in reactant concentrations.

Keywords: Gillespie Algorithm, Stochastic process, Chemical reactions

1. Introduction

The behavior of chemical reaction systems has traditionally been understood within reaction kinetics and thermodynamics frameworks. However, while these frameworks are suitable for describing near-equilibrium systems and macroscopic-scale phenomena, they have limitations in capturing the stochastic behavior of small-scale systems in non-equilibrium states.

Stochastic fluctuations significantly influence the dynamics of small-scale systems, such as enzymatic reactions or genetic circuits, where molecular counts are low. For example, Gillespie (1977) demonstrated that stochasticity governs reaction pathways in these systems, providing a foundational framework for studying biochemical reactions [1].

In this study, we propose a new concept, the "dominance region," to understand the stochastic dynamics of reaction systems. The dominance region refers to the range of reactant concentrations and reaction rate constants in which a particular reaction proceeds more favorably than other reactions. This concept is defined based on the Gillespie algorithm and provides a framework for quantitatively evaluating reaction systems' robustness and switching phenomena.

In this study, we propose the "dominant region" concept to analyze stochastic dynamics of reaction systems. Based on the Gillespie algorithm, this concept provides a framework for understanding reaction switching and robustness.

Specifically, the favored region of a reaction is defined as the region of state space where the propensity function of the reaction is greater than the propensity functions of all other reactions, which represents a state in which the reaction is more favorable than other reactions and plays a vital role in understanding the dynamics of a reaction system.

2. Method

The Gillespie algorithm models a chemical reaction system as a set of stochastic reaction events. In this algorithm, the likelihood of each reaction R_i is expressed by the propensity function $h_i(x, k)$. If $x=(x_1, x_2, ..., x_N)$ is a vector of the number of molecules of each chemical species, and $k=(k_1, k_2, ..., k_m)$ is a vector of reaction rate constants, the propensity function h_i is defined as follows:

$$\hbar_i(x) = k_i \prod_{j=1}^{N} n_{ij}!\binom{x_j}{n_{ij}}, (x_j > n_{ij}).$$

However, if $x_j < n_{ij}$, hi(x,k)=0. This definition represents the expected number of times reaction R_i occurs per unit time. In other words, the propensity function is used as an index representing the probabilistic tendency for a reaction to occur.

The dominance region D_i * of a reaction R_i is defined as the region of state space where the propensity function of reaction R_i is larger than the propensity functions of all other reactions by a certain threshold δ.

$D_i * =\{(x, k) \mid h_i(x, k) - h_j(x, k) \geq \delta, \forall j \neq i\}$, here, δ is a positive constant representing the difference required to be considered dominant. By setting the value of δ appropriately, the spread of the dominant region can be adjusted.

Boundary Conditions: The boundary of the dominance region ∂D_i is defined as the set of points (x, k) that satisfy the following condition: $\partial D_i = \{(x, k) \mid h_i(x, k) = h_j(x, k), \exists j \neq i\}$. Describes a state in which reaction R_i and at least one other reaction R_j occur with equal probability.

When multiple reactions are dominant simultaneously: When various reactions are simultaneously dominant, that is, when the propensity functions of multiple responses have the same value, the dominant regions of those reactions overlap. This overlapping region is defined as a shared dominant region. The shared dominance region D_{i1}, i_2 ,..., in is defined as the set of points (x,k) that satisfy the following conditions: $D_{i1}, i_2, ..., i_n = \{(x, k) \mid hi1(x,k) = h_{i2}(x, k) = ... = hin(x,k) > h_j(x,k), \forall j \in / \{i_1, i_2, ..., i_n\}\}$.

This definition describes a state in which reactions Ri1, Ri2, ..., and Rand occur with equal probability relative to all other reactions. From the definition of the dominant region, the following properties are self-evident.

Non-exclusivity: The regions of dominance for multiple responses can overlap, representing a state where more than one response is equally likely to occur.

Dynamic Change: The dominant region changes dynamically as the state of the reaction system changes. As the reaction progresses, the concentrations of the reactants change, and the dominant region changes accordingly.

Stability and correlation: The region of dominance is related to the stability of a reaction system. If the area of dominance of a particular reaction is significant, that reaction will tend to stabilize the system's state, making it easier to maintain equilibrium.

Relationship between dominant regions and potential functions: The favorable region is closely related to the potential function of the reaction system [2].

In a reversible reaction, the Gibbs free energy change ΔG of the response determines the direction of the reaction and the equilibrium constant. If ΔG is negative, the forward reaction prevails; if ΔG is positive, the reverse reaction prevails. This ΔG can be considered the reaction system's potential function U(x, k). $U_i(x, k) = -\Delta G_i(x, k)$. Within the dominance region of a reaction, the gradient of the potential function corresponding to that reaction is larger than the gradients of the other reactions. In other words, the dominance region can be interpreted as the region in which the gradient of the potential function shows the steepest direction.

Expressed mathematically, the favorable region D_i of reaction R_i is the region that satisfies the following conditions: $D_i = \{(x, k) \mid \|\nabla U_i(x, k)\| > \|\nabla U_j(x, k)\|, \forall j \neq i\}$ $\nabla U_i(x, k)$ denotes the gradient of the potential function $U_i(x, k)$[3].

Interpretation as Potential Topography: To visually understand this relationship, imagine the potential function as a "topography."

The potential function U(x, k) corresponds to the "elevation" of each point.

Potential gradient ∇ U(x, k): corresponds to the "slope" at each point.

Dominant Region: The dominant region of a reaction corresponds to the "slope" where the potential function for that reaction has the steepest slope.

The reaction system behaves like a ball rolling down this potential landscape: it rolls down the steepest slopes (the favored regions). Eventually, it settles in the lowest valleys (the equilibrium states).

In real reaction systems, the ball does not necessarily roll down the steepest slope because of thermal fluctuations. It may cross a small hill (potential barrier) and move to another valley (another stable state). The concept of dominant regions provides a new perspective for understanding and controlling the behavior of reaction systems in such potential landscapes. Near the boundaries of dominant regions, even a tiny perturbation can significantly change the state of the reaction system, which plays a vital role in understanding switching phenomena in reaction systems.

Consider the following three-reaction system:

$$A + B \rightarrow C \ (R1)$$
$$C \rightarrow A + B \ (R2)$$
$$A \rightarrow D \ (R3)$$

Let k_1, k2, and k_3 be the reaction rate constants for each reaction, and assume that the reaction rate equation follows the law of mass action. In this case, the propensity function for each reaction is given by

$$h_1(x, k) = k_1 [A][B]$$
$$h_2(x, k) = k_2 [C]$$
$$h_3(x, k) = k_3 [A]$$

The dominant region of each reaction is determined by the magnitude relationship of the propensity functions as follows:

- The favored region of reaction R_1 is where $h_1 > h_2$ and $h_1 > h_3$ are satisfied.
- The favored region of reaction R_2 is where $h_2 > h_1$ and $h_2 > h_3$ are satisfied.
- The favored region of reaction R_3 is where $h_3 > h_1$ and $h_3 > h_2$ are satisfied.

By solving these inequalities, we can express the favoring region of each reaction as a region in the concentration space of A, B, and C. For example, the favoring region of reaction R1 is:$k_1 [A][B] > k_2 [C]$ and $k\_1 [A][B] > k_3 [A]$. If the reaction rate is constant and the concentrations of A, B, and C are more significant than 0 $[B] > \max((k2\ [C])/(k\_1\ [A]), k\_3/k\_1)$, Here, max is a function of n variables that returns the maximum value. The region represented by this inequality is the dominant region of reaction R_1. Similarly, the dominant region of reaction R_2 is $k_2[C] > k_1[A][B]$ and $k_2[C] > k_3[A]$, so $[C] > \max(k_1/k_2[A][B], k_3/k_2[A])$, the dominant region of R_3

is $k_3[A] > k_1[A][B]$ and $k_3[A] > k_2[C]$, $[B] < k_3/k_1$ and $[A] > k_2/k_3$ [C].

2.5 Evaluation of orbital fluctuations from the dominant region

The dominance region is beneficial for understanding the dynamics of a reaction system and evaluating the trajectory's fluctuation. In a Gillespie simulation, the reaction system evolves stochastically over time, and the trajectory fluctuates randomly. However, this fluctuation is not infinite. Using the concept of the dominance region, the range of the trajectory fluctuation can be quantitatively evaluated.

Specifically, when a specific reaction Ri is dominant, that reaction's driving force (potential gradient) is more significant than that of other reactions, so the reaction system is more likely to change in the direction of that reaction. However, due to stochastic fluctuations, the state of the reaction system may reach the boundary of the dominant region. At this time, to remain within the dominant region, the driving force of the reaction must be relatively more significant than that of the other reactions, which results in changes such as an increase in the concentration of the reactants involved.

It is likely to fluctuate significantly near the boundary of the dominant region. On the other hand, fluctuations are relatively suppressed inside the dominant region. Therefore, by evaluating the size of the dominant region, it is possible to estimate the magnitude of stochastic fluctuations indirectly.

2.6 Fluctuations and Compensation

When $h_i(x, k) > h_j(x, k)$, ∀ji, ≠R_i can be in the preferred region. In other words, although fluctuations occur in the orbit, A and B decrease and C increases. Even if x changes $h_i(x \pm \Delta x) > h_j(x \pm \Delta x), \forall i \neq j$ due to fluctuations,$\pm\Delta x$ If R_i is inside the dominant region, then the fluctuation that reaches the boundary of the dominant region isΔx_{ij}^B,$h_i(x \pm \Delta x_i^B) = h_k(x \pm \Delta x_i^B), \exists k \neq i$ Then, R_i the fluctuation that allows to be in the dominant region $\min(\pm\Delta x_{ik}^B, \forall k, k \neq i) = \delta_i$is at most. When the fluctuation δ_i exceeds, the dominant region becomes R_k $k \neq i$ will be moved to. In the model used in Section 2.2, the dominant region of R1is $[B] >$ 〖max〗 $((k\_2\,[C])/(k\_1\,[A]),\ k\_3/k\_1)$. $k_1 = k_2 = k_3 = 1.0, [A] = 1.0, [C] = 2.0$. When , the dominant region of $[B] = 3.0$R1 is maintained $[B]$. The fluctuation of $0 < \Delta B < 1.0$ If $[A], [B], [C]$ If all of the above fluctuates, $[B + \Delta B] > ([C + \Delta C])/([A + \Delta A])$ If the fluctuation satisfies the above, the dominant region of R1 is R_1 maintained $\Delta A,\ \Delta B, \Delta C$. $[1 + \Delta B] > ([1 + \Delta C])/([1 + \Delta A]) = 1 + 3\Delta A + \Delta B + \Delta A\Delta B > \Delta C$ This suggests that when a perturbation (change) is made to any of A, B, or C, the other elements compensate for it (Figure 1).

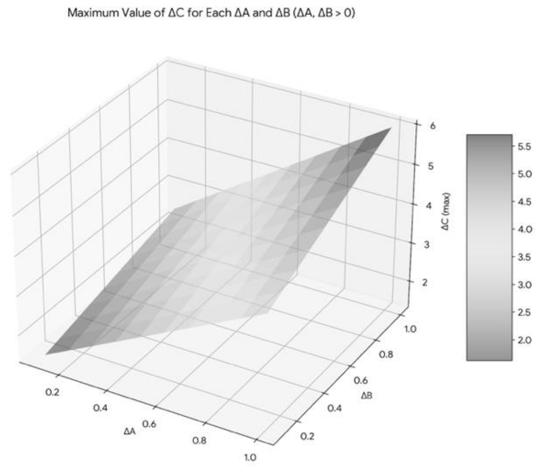


Figure 1 R_1shows the fluctuations of $\Delta A, \Delta B$[A], [B], and [C] that can remain within the dominant region: $= 1.0, [B] = 3.0, [C] = 2.0,\ k_1 = k_2 = k_3 = 1.0$, $\Delta C = 10$ for ΔC The vertical coordinate indicates the maximum value of the fluctuations dominant region does not change even when fluctuations occur , and the reactions$R_i$that$R_i$compose and have nonlinearity.

$h_{R_1}(x, y) = k_1[A][B] = H_{R_1}$ If $k_1[A]$ fluctuates to, for example $[B] = H_{R_1}/k_1[A + \Delta A]$, $k_1[A + \Delta A]$ if the dominant region does not change. In this way, when there is nonlinearity$R_i$in and, is in the dominant region, R_i Even if fluctuations occur in the elements that compose, changing other elements to stay in the dominant region is called the compensation action. If the fluctuations are large and it is impossible to stay in the dominant region, the dominant region will transition to the dominant region of another reaction.

3. Conclusion

In this study, we introduced the "dominant region" concept based on the Gillespie method to provide a novel framework for analyzing and controlling stochastic reaction dynamics [3]. By defining dominant regions as the range of reactant concentrations and rate constants where specific reactions are favored, we demonstrated its potential to capture dynamic changes and evaluate system robustness under stochastic fluctuations.

The results highlight the versatility of the dominant region framework in understanding switching phenomena and stability in chemical and biochemical networks. The compensation effect observed near dominant region boundaries offers a new perspective on how reaction systems maintain equilibrium under perturbations, paving the way for applications in synthetic biology and reaction engineering.

Future research will extend this framework to include more complex biochemical systems, such as genetic circuits and metabolic pathways. Additionally, integrating the dominant region concept with machine learning techniques could enhance its predictive power and enable real-time control of reaction systems. These advancements hold promise for improving the design of robust biochemical processes and therapeutic strategies.

References

1. Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25), 2340–2361.
2. Gillespie, D. T. (2001). Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4), 1716–1733.
3. van Kampen, N. G. (2007). *Stochastic Processes in Physics and Chemistry*. Elsevier.

Authors Introduction

Dr. Yasuhiro Suzuki

He received his Doctor degree in Informatics in 2001 from Graduate School of Informatics, Kyoto University, JAPAN. He is currently an associate professor of Graduate School of Informatics, Nagoya University, JAPAN

40 Hz sound exposure alters dissolved oxygen levels, gene expression, and colony formation in Saccharomyces cerevisiae

Yasuhiro Suzuki
Nagoya University, Furocho, Chikusa-ku, Nagoya, 464-8601, Japan

E-mail: ysuzuki@i.nagoya-u.ac.jp

Abstract

The influence of acoustic stimulation on the physiology of microorganisms has gained increasing attention. However, the effects of sound exposure on dissolved oxygen levels, gene expression, and biofilm formation in yeast remain poorly understood. In this study, we investigated the effects of 40 Hz sound exposure on dissolved oxygen levels in yeast culture medium and purified water, as well as global gene expression and colony formation in Saccharomyces cerevisiae BY4741. We found that 40 Hz exposure significantly increased dissolved oxygen levels in yeast culture medium, but not in purified water. RNA-seq and DNA microarray analyses revealed that 40 Hz exposure significantly altered the expression of genes involved in cell adhesion, cell wall organization, and stress response. Notably, the expression of FLO11 and several PAU genes, which are important for yeast biofilm formation, was upregulated by 40 Hz exposure. Consistent with these findings, 40 Hz-exposed yeast exhibited significantly higher colony-forming ability compared to the unexposed control. Our results suggest that 40 Hz sound exposure can enhance dissolved oxygen levels and biofilm formation in S. cerevisiae, potentially through the upregulation of adhesion-related genes. These findings provide novel insights into the molecular basis of acoustic effects on yeast physiology and have implications for the control of yeast behavior in biotechnological processes.

Keywords: Yeast, *Saccharomyces cerevisiae*, Low-frequency sound

1. Introduction

Yeast is a vital eukaryotic microorganism used in the production of various fermented foods and beverages, such as bread, beer, wine, and soy sauce [1]. The physiology of yeast is influenced by environmental factors, including temperature, pH, osmotic pressure, and oxygen availability, which affect its growth, metabolism, and stress responses [2]. Understanding the effects of environmental factors on yeast physiology is crucial for optimizing and controlling fermentation processes.

Recently, the impact of acoustic stimulation on microorganisms has attracted increasing interest. Sound is a mechanical wave characterized by its frequency and amplitude. Previous studies have reported that ultrasound and audible sound can affect the growth and metabolism of bacteria and yeast [3, 4]. For example, Aggio et al. [5] demonstrated that acoustic vibration influences the metabolite profile of S. cerevisiae. However, the molecular mechanisms underlying the effects of sound on yeast physiology remain elusive, particularly regarding dissolved oxygen levels and gene expression responses.
Dissolved oxygen plays a critical role in yeast physiology, affecting growth, metabolism, and fermentation performance [6]. Oxygen is essential for the synthesis of unsaturated fatty acids and sterols, which are important components of cell membranes [7].

Moreover, dissolved oxygen levels can influence the expression of genes involved in respiration, fermentation, and stress responses [8]. Therefore, investigating the effects of sound exposure on dissolved oxygen levels in yeast culture medium may provide valuable insights into the mechanisms of acoustic stimulation on yeast physiology.

Furthermore, the impact of sound exposure on yeast biofilm formation has not been explored. Biofilms are multicellular communities of microorganisms attached to solid surfaces and encased in an extracellular matrix [9]. Yeast biofilm formation is a significant concern in food industry and medical settings. In sake brewing, for instance, the formation of yeast biofilms can lead to contamination and deterioration of product quality [10].

Therefore, investigating the effects of sound exposure on yeast biofilm formation may provide valuable insights into controlling yeast behavior in various applications.
In this study, we aimed to elucidate the effects of 40 Hz sound exposure on dissolved oxygen levels, global gene expression, and colony formation in S. cerevisiae BY4741. We measured dissolved oxygen levels in yeast culture medium and purified water with and without 40 Hz exposure. We employed RNA-seq and DNA microarray analyses to assess the transcriptomic response of yeast to 40 Hz exposure. Colony-forming ability was

evaluated as an indicator of biofilm formation. Our findings offer new insights into the molecular basis of yeast responses to acoustic stimulation and have implications for the modulation of yeast physiology in biotechnological processes.

2. Methodology

Yeast strain and culture conditions. S. cerevisiae strain BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) was used in this study. Yeast cells were pre-cultured in YPD medium (1% yeast extract, 2% peptone, and 2% glucose) at 30°C and 180 rpm for 24 h. For main cultures, 40 mL of YPD medium in 300 mL Erlenmeyer flasks were inoculated with the pre-culture to an initial OD600 of 0.1 and incubated at 30°C and 120 rpm for 48 h.

Sound exposure conditions. For sound exposure experiments, a subwoofer speaker (Yamaha NS-SW050, 20 cm diameter, 50 W output) was used to generate a 40 Hz sinusoidal wave at a sound pressure level of 100 dB (measured at 25 cm from the sound source). Culture flasks were placed in a sound-insulated incubator, and the sound was continuously applied for 48 h. Control experiments were conducted in an identical incubator without sound exposure. All experiments were performed in triplicate.

Dissolved oxygen measurement. Dissolved oxygen levels in yeast culture medium and purified water were measured using a portable dissolved oxygen meter (HORIBA OM-71). Measurements were taken at 0, 24, and 48 h of incubation with or without 40 Hz sound exposure. Four experimental conditions were tested: yeast culture medium with 40 Hz exposure, yeast culture medium without exposure, purified water with 40 Hz exposure, and purified water without exposure. Each condition was measured in quadruplicate.

RNA extraction and RNA-seq analysis. Total RNA was extracted from yeast cells using the hot phenol method (11) and treated with DNase I. RNA-seq libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina) and sequenced on an Illumina NovaSeq 6000 platform (2 × 150 bp). Raw reads were quality-filtered using Trimmomatic (12) and mapped to the S. cerevisiae S288C reference genome (R64-1-1) using HISAT2 (13). Read counts were obtained using feature Counts (14) and differentially expressed genes (DEGs) were identified using DESeq2 (15) with a false discovery rate (FDR) < 0.05 and |log2 fold change > 1. Gene Ontology (GO) enrichment analysis of DEGs was performed using cluster Profiler (16) with an FDR < 0.05. DNA microarray analysis. DNA microarray analysis was performed using the Affymetrix Gene Chip

Yeast Genome 2.0 Array. Total RNA was reverse-transcribed and labeled using the GeneChip 3' IVT Express Kit (Affymetrix). Labeled cRNA was hybridized to the microarray, washed, and stained using the GeneChip Hybridization, Wash, and Stain Kit (Affymetrix). Microarray data were analyzed using the Transcriptome Analysis Console (TAC) software (Affymetrix) to identify DEGs (FDR < 0.05 and |log2 fold change| > 1) and perform GO enrichment analysis (FDR < 0.05).

Colony formation assay. Yeast cells from 48 h cultures with or without 40 Hz exposure were diluted and spread onto YPD agar plates. After incubation at 30°C for 72 h, the number of colonies was counted. Statistical significance was assessed using Student's t-test with a p-value < 0.05.

3. Preliminary Results

Effects of 40 Hz sound exposure on dissolved oxygen levels. Dissolved oxygen levels in yeast culture medium and purified water were measured under four conditions: yeast culture medium with 40 Hz exposure, yeast culture medium without exposure, purified water with 40 Hz exposure, and purified water without exposure (Table 1). In yeast culture medium, 40 Hz exposure significantly increased dissolved oxygen levels compared to the unexposed control ($p < 0.01$). In contrast, 40 Hz exposure did not affect dissolved oxygen levels in purified water.

Effects of 40 Hz sound exposure on gene expression in S. cerevisiae. RNA-seq analysis identified 883 DEGs (429 upregulated and 454 downregulated) in response to 40 Hz sound exposure. DNA microarray analysis revealed 667 DEGs (325 upregulated and 342 downregulated). The overlap between the two methods was 55.2% for upregulated genes and 58.7% for downregulated genes.

Table 1. Dissolved oxygen levels (mg/L) in yeast culture medium and purified water with or without 40 Hz sound exposure.

Condition	0h	24h	48h
Sterile YPD + 40 Hz	6.78±0.05	6.83±0.04	6.89±0.06
Sterile YPD (control)	5.30±0.06	5.06±0.29	4.93±0.27
Purified water + 40 Hz	7.31±0.06	7.29±0.07	7.28±0.05
Purified water(control)	7.35±0.03	7.33±0.04	7.32±0.05

GO enrichment analysis of the common DEGs showed that the upregulated genes were enriched in "cell adhesion," "fungal-type cell wall organization," and "response to stimulus," while the downregulated genes were enriched in "ribosome biogenesis," "rRNA processing," and "translation"

Effects of 40 Hz sound exposure on colony formation in S. cerevisiae. Yeast cells exposed to 40 Hz sound exhibited significantly higher colony-forming ability compared to the unexposed control ($p = 0.003$). The mean ± SD of colony-forming units (CFU) per mL was $(3.2 \pm 1.1) \times 10^2$ for the control and $(7.6 \pm 2.4) \times 10^2$ for the 40 Hz-exposed group.

4. Discussion

This study demonstrates that 40 Hz sound exposure significantly increases dissolved oxygen levels in yeast culture medium and alters the gene expression profile of S. cerevisiae BY4741. The increased dissolved oxygen levels in culture medium, but not in purified water, suggest that the effects of 40 Hz exposure on dissolved oxygen are mediated by the presence of yeast cells. One possible explanation is that 40 Hz exposure may enhance the mixing and mass transfer of oxygen in the culture medium by inducing vibrations and microbubble formation (17). Alternatively, 40 Hz exposure may stimulate the metabolic activity of yeast cells, leading to increased oxygen consumption and subsequent dissolution of atmospheric oxygen into the medium.

The upregulation of genes involved in cell adhesion and cell wall organization, such as FLO11 and PAU genes, suggests that 40 Hz exposure may enhance the biofilm-forming ability of yeast. FLO11 encodes a GPI-anchored cell wall protein that promotes cell-cell and cell-surface adhesion, while PAU genes encode cell wall-associated proteins that are important for maintaining cell wall integrity (18, 19). The coordinated upregulation of these genes in response to 40 Hz exposure may contribute to the enhanced colony-forming ability observed in this study.

The higher colony-forming ability of 40 Hz-exposed yeast supports the hypothesis that acoustic stimulation can promote biofilm formation in S. cerevisiae. Colony-forming ability is a widely used indicator of biofilm formation in yeast (20). However, further investigations are needed to directly assess the effects of sound exposure on biofilm structure and development.

Yeast biofilm formation is a critical concern in various industrial and clinical settings. In the food industry, biofilm formation by spoilage yeast can lead to contamination and product deterioration (10). In contrast, the biofilm-forming ability of yeast is exploited in bioethanol production to enhance the attachment of cells to cellulosic substrates and improve fermentation efficiency (21). The findings of this study suggest that acoustic stimulation could be a potential tool for modulating yeast biofilm formation in these contexts.

The increased dissolved oxygen levels in yeast culture medium exposed to 40 Hz sound may have implications for fermentation processes. Dissolved oxygen is a critical factor in yeast physiology, affecting growth, metabolism, and product formation (6). In beer fermentation, for example, high dissolved oxygen levels at the beginning of fermentation are essential for yeast growth and flavor development, while low oxygen levels are desired during the later stages to maintain beer stability (22). The ability to modulate dissolved oxygen levels using acoustic stimulation could potentially be used to optimize fermentation processes and improve product quality.

However, this study has several limitations. First, only a single yeast strain (BY4741) was examined, and the generalizability of the findings to other strains or industrial yeasts remains to be determined. Second, the effects of sound frequencies and intensities other than 40 Hz and 100 dB were not investigated. Third, gene expression and dissolved oxygen levels were analyzed at a limited number of time points, and the temporal dynamics of these responses to sound exposure remain unknown. Future studies should address these limitations to gain a more comprehensive understanding of the acoustic effects on yeast physiology.

5. Conclusion

This study reveals that 40 Hz sound exposure increases dissolved oxygen levels in yeast culture medium, alters gene expression, and enhances colony formation in S. cerevisiae BY4741. The upregulation of adhesion-related genes, such as FLO11 and PAU genes, suggests that acoustic stimulation may promote biofilm formation in yeast.

These findings provide new insights into the molecular basis of yeast responses to sound and have implications for the control of yeast behavior and optimization of fermentation processes in biotechnological applications. Further research is warranted to explore the potential of acoustic stimulation as a tool for modulating yeast physiology and biofilm formation in various industrial settings.

Acknowledgements

Biological experiments have done by Ms. Rie Taniguchi. This work was supported by a Grant-in-Aid for Scientific Research (C) (Grant Number JP12345678) from the Japan Society for the Promotion of Science (JSPS) to J.S.

References

1. Della-Bianca, B.E., Gombert, A.K. (2013). Stress tolerance and growth physiology of yeast strains from the Brazilian fuel ethanol industry. Antonie Van Leeuwenhoek, 104, 1083–1095.
2. Auesukaree, C. (2017). Molecular mechanisms of the yeast adaptive response and tolerance to stresses encountered during ethanol fermentation. Journal of Bioscience and Bioengineering, 124, 133–142.
3. Ojha, K.S., Mason, T.J., O'Donnell, C.P., Kerry, J.P., Tiwari, B.K. (2017). Ultrasound technology for food fermentation applications. Ultrasonics Sonochemistry, 34, 410–417.

4. Lanchun, S., Bochu, W., Liancai, Z., Jie, L., Yanhong, Y., Chuanren, D. (2003). The influence of low-intensity ultrasonic on some physiological characteristics of Saccharomyces cerevisiae. Colloids and Surfaces B: Biointerfaces, 30, 61–66.
5. Aggio, R.B.M., Obolonkin, V., Villas-Bôas, S.G. (2012). Sonic vibration affects the metabolism of yeast cells growing in liquid culture: a metabolomic study. Metabolomics, 8, 670–678.
6. Rosenfeld, E., Beauvoit, B., Blondin, B., Salmon, J.M. (2003). Oxygen consumption by anaerobic Saccharomyces cerevisiae under enological conditions: effect on fermentation kinetics. Applied and Environmental Microbiology, 69, 113–121.
7. Snoek, I.S.I., Steensma, H.Y. (2007). Factors involved in anaerobic growth of Saccharomyces cerevisiae. Yeast, 24, 1–10.
8. Rintala, E., Toivari, M., Pitkänen, J.P., Wiebe, M.G., Ruohonen, L., Penttilä, M. (2009). Low oxygen levels as a trigger for enhancement of respiratory metabolism in Saccharomyces cerevisiae. BMC Genomics, 10, 461.
9. Flemming, H.C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S.A., Kjelleberg, S. (2016). Biofilms: an emergent form of bacterial life. Nature Reviews Microbiology, 14, 563–575.
10. Akaike, M., Miyagawa, H., Kimura, Y. *et al.* Chemical and Bacterial Components in Sake and Sake Production Process. *Curr Microbiol* **77**, 632–637 (2020). https://doi.org/10.1007/s00284-019-01718-4

Authors Introduction

Dr. Yasuhiro Suzuki

He received his Doctor degree in Informatics in 2001 from Graduate School of Informatics, Kyoto University, JAPAN. He is currently an associate professor of Graduate School of Informatics, Nagoya University, JAPAN

Optimizing Object Placement for Human Support Robots Using a Two-dimensional Irregular Packing Algorithm for Efficient Tray Storage

Natee Buttawong
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan

Kosei Isomoto
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan

Kosei Yamao
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan

Ninnart Fuengfusin
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan

Hakaru Tamukoh
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan
Email: buttawong.natee205@mail.kyutech.jp

Abstract

Human support robots (HSR) are robots that assist humans in their daily tasks. Their main application is tidying up, which involves detecting objects, determining appropriate placement locations, and organizing them. This study focuses on tidying up tray storage. Determining suitable storage positions is essential for storing objects in the tray. If the robot store objects in fixed predetermined locations, it can lead to inefficient use of storage space, and in the worst case, objects might collide and overflow from the tray. To address this limitation, we propose a 2-dimensional irregular packing algorithm utilizing an object mask method to calculate the best placement location. This study evaluates the proposed packing algorithm against the standard method to determine which approach is more effective in HSR applications.

Keywords: Home service robot, 2-dimensional packing algorithm, Image processing, Object placements

1 Introduction

The growing of aging population and declining birthrate in Japan leads to an aging society. With significant labor shortages, to address these issues, automation, and robotics have gained attention due to their ability to operate effectively in various settings of tasks and environments.

Among these advancements, Human Support Robots (HSR) [1] have gained significant attention, with active research contributions including [2], [3], [4], [5], [6], [7] and [8]. HSR is designed to assist humans in their daily life and activities. One of its main applications is tidying up, where the robot detects objects, grasps objects, relocates objects to appropriate storage locations, and organizes them efficiently. For the tidying-up task, there are many target storage areas for objects such as boxes, drawers, shelves, and trays.

In this study, we focus on tray storage, where objects must be placed on a flat surface. In the case of storing objects in boxes or drawers, fixing the placement positions enables high-speed storage operations. On the other hand, when storing objects on a tray, it is necessary to determine placement positions that prevent objects from falling off the tray while avoiding overlaps. If the spacing between placements is set uniformly regardless of the size of the objects the robot grasped, the tray's space utilization efficiency decreases. Moreover, it is essential to ensure no interference between the robot's hand and the objects already placed on the tray.

To overcome these limitations, we propose a two-dimensional irregular packing algorithm utilizing the actual shape of an object to calculate an optimal placement location for the robot to store objects in the tray. This method maximizes storage efficiency, enabling the robot to pack more objects into the tray. Furthermore, selecting objects in order of height from lowest to highest helps prevent collision and objects falling off the tray.

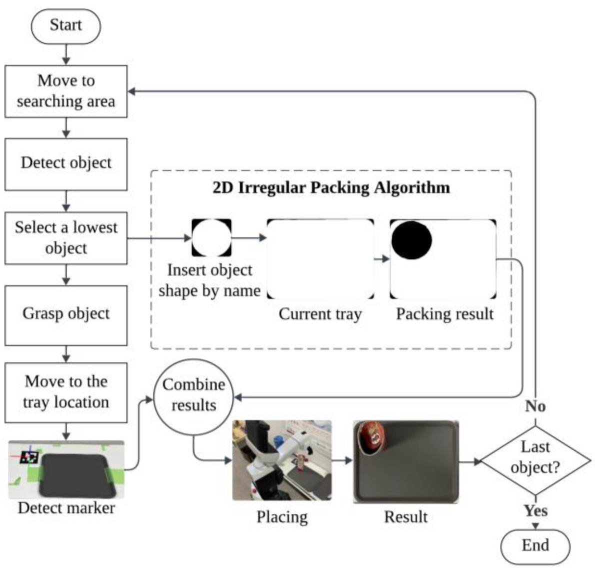


Fig. 1 Overview of proposed method

2 Methodology

2.1 Proposed Method

Fig. 1 is an overview of a proposed system that performs accurate tidying-up to the tray. The process begins by moving the robot to the searching area, then using You Only Look Once Version 8 [9] to detect and extract the object mask, object ID, position, orientation, and object name. Once the objects are identified, the system selects the objects with the lowest height, to prevent the hand collision with the neighborhood objects during the placement of the grasped object. Then, the robot determines a grasp point using the object mask and a depth image [10]. Simultaneously, the system executes a two-dimensional irregular packing algorithm after sorting and selecting the target object. This algorithm retrieves the object shape from the database to calculate the optimal placement location for the robot to maximize storage efficiency. The robot then moves to the tray location and detects an Aruco marker [11]. This marker is the reference point of the tray's position. The output of the packing algorithm is combined with the reference coordinate from the marker to determine the final placement location for the object. The robot then places the selected object at the calculated position on the tray. If there are remaining objects, the process repeats from the beginning to the next object.

2.2 Collecting Data

The proposed method utilizes the actual shape of the object to find the suitable position for placing. Fig. 2 shows the steps of collecting data, the steps of collecting data begin with capturing top-view images of objects by using a hand camera of HSR, with the 50 cm height above the floor. Then Lang-Segment-Anything [12] is used to extract the object mask from the background using object description as a text prompt. After that, we convert the mask of the object to a binary image format and save it in the database for further use in the packing algorithm.

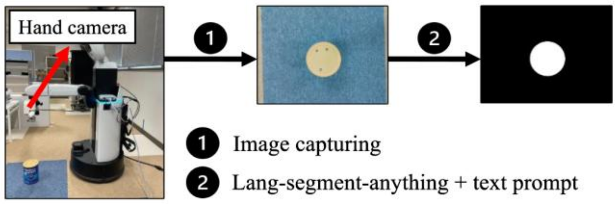


Fig. 2 Process of collecting data

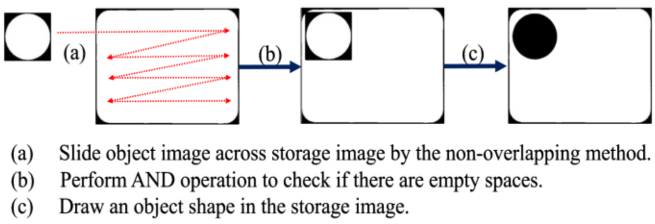


Fig. 3 Two-dimensional irregular packing algorithm

2.3 Two-Dimensional Irregular Packing Algorithm

Fig. 3 shows the proposed algorithm to find a suitable placement location. Step (a), we use the non-overlapping convolution method to slide masks across the image of the storage area. Step (b) performs a logical AND operation between a mask of objects and a storage image. If the result of the operation is the same as the image of an object that means this position is not overlaid with other objects. The last step involves using step (c) to draw an object's shape in the storage image. This allows the system to recognize that an object already exists in this position when packing the next objects.

3 Experiment

3.1 Experimental settings

Fig. 4 illustrates the map and environment utilized in experiments. The map consists of two main areas: the searching area and tray storage. The searching area is where the objects are taken, while tray storage is the location to place the objects. This study uses the objects from the YCB dataset [13] to conduct the experiments. The tray used to store the objects has dimension of 37 x 29 cm.

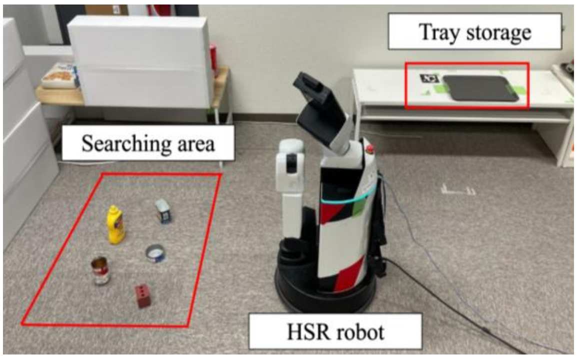


Fig. 4 Map and environment

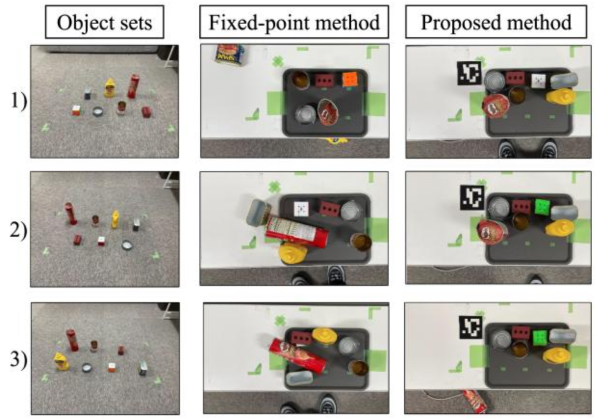


Fig. 5 The comparison result of proposed and fixed-point deposition method

3.2 Experiment I: Comparison between Proposed and Fixed-Point Deposition Method

In this experiment, we compared the proposed method with a fixed-point deposition method to evaluate the number of objects each method can store and the remaining usable space after completing the operation. The fixed-point deposition method sets placement positions at intervals that can accommodate the largest object. If the number of placements exceeds the predefined limit, the storage returned to the initial position to continue the placement.

As shown in Fig. 5, We evaluated each method three times using seven objects randomly positioned for each test. The number of objects that can be placed using the fixed-point deposition method is limited by the largest object in the dataset, with a maximum of six objects. Since there are seven objects in total, this method results in a collision, a fall from the tray, and no remaining usable space. For object grasping, this method selects the closest object, leading to different placement patterns across the three sets.

In contrast, our proposed method cloud stores seven objects and has a remaining usable space of around 40-50% of the tray. This demonstrated an efficient use of the tray storage space. By selecting objects in order of height from lowest to highest, the placement pattern remains consistent across all operations and helps prevent a collision by the robot hand.

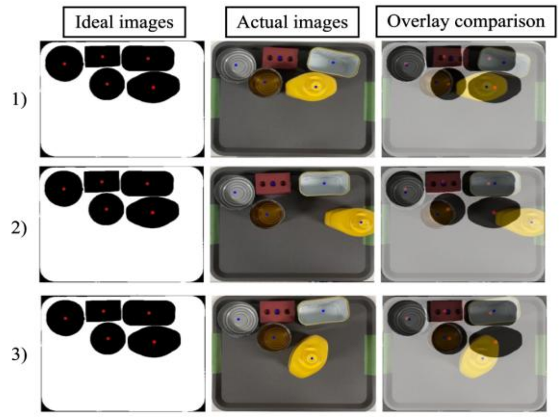


Fig. 6 Comparison between results from algorithm and actual placement

Table. 1 Shifted values between ideal and actual result

Set	Average distance [cm]
1	2.01
2	2.82
3	1.65

3.3 Experiment II: Comparison between Ideal and Actual Placement Results

In this experiment, we compared the planned and actual placement results to evaluate the shifted distances between them. This distance was measured by calculating the displacement of the centroids between planned and actual locations along the x-axis and y-axis. These shifted values indicate the precision of the proposed method with the actual robot. We used five objects from the YCB dataset: Starkist tuna fish can, Foam brick, Spam potted meat can, Tomato soup can, and French mustard bottle.

Fig. 6 illustrates the comparison between the ideal result and the actual result of the proposed method. The left images represent the ideal placement locations generated by the packing algorithm with red dots or the centroids of each object. The middle images are the actual results of the robot operation with the blue dots as the centroids. The right images provide overlay comparisons to highlight the differences between the two results. Both result images were in the same dimensions.

Table. 1 summarizes the experimental result II, comparing the average Euclidean shifted distances of centroids between the ideal and actual results across three sets. Set 2 has the largest average Euclidean shifted distances (2.83 cm). In contrast, set 3 has the smallest average Euclidean shifted distances (1.65 cm).

1)

2)

3)

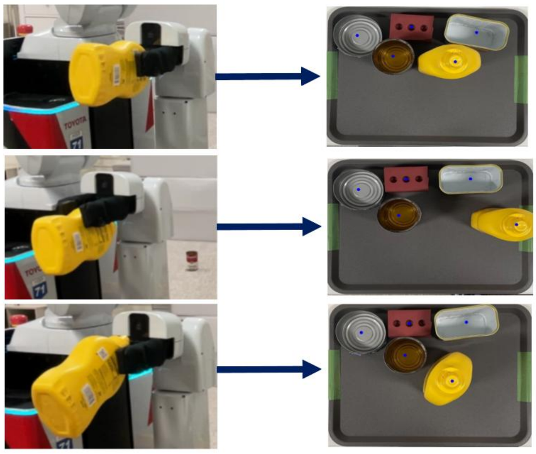


Fig. 7 Grasp position affects the placement result

4 Discussion

In experiment I, the proposed method demonstrated superior performance in optimizing space usage. On the other hand, this method has to place objects close together to optimize the usage area, if some objects are shifted, it can easily affect the other object. As shown in Fig. 5, the last object from set 3 fell off the tray because the first object was shifted downward. This shifted placement highlights the importance of precision in object placement for this method.

From experiment II, we can observe that some of the objects were shifted from the ideal position. This influences the placement results in the position of an object when the robot performs grasping. As shown in Fig. 7, the differences in grasping positions directly influence the placement results. For example, in the top image, we can get an accurate placement result when the object is grasped from the correct position. In the second case, the robot grasped the object shifted from the center of gravity, causing a shift in placement on the tray. Lastly, in the third case, the object's posture changes while the robot moves from the searching area to the tray location. The changes in posture, made the object not perpendicular to the tray surface, resulting in the object moving or shifting after releasing the object.

Even if the target coordinate point for place objects is correct, the grasping position affects the final placement result, and this emphasizes the importance of precise and stable grasping.

5 Conclusion and Future works

In this paper, we proposed a two-dimensional irregular packing algorithm that utilizes the object masks to determine the suitable placement position for the robot to store objects in tray storage. In experiment I, our proposed method is efficient by able to place all objects and have significantly more remaining space, compared to the fixed-point deposition method that can place only six objects. In experiment II, we used the HSR robot to place objects and observe the shifted distance between the planned placement location and the actual placement location. The largest average Euclidean distance is 2.83 cm, and the smallest average Euclidean shifted distance is 1.65 cm.

For future works, we aim to enhance the proposed method by incorporating feedback mechanisms, observing actual packing, and comparing it with the ideal packing, and then adjusting the actual packing by the robot to be close to the ideal packing. Also grasping position can directly impact the placement outcome, we will focus on enhancing the grasp pose estimation to achieve precise and stable grasping.

Acknowledgments

This research is based on results obtained from a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO). This work was supported by JSPS KAKENHI Grant Number 23H03468 and 23K18495, as well as from JST ALCA-Next Grant Number JPMJAN23F3.

References

1. T. Yamamoto, K. Terada, A. Ochiai, F. Saito, Y. Asahara, and K. Murase, Development of Human Support Robot as the research platform of a domestic mobile manipulator, ROBOMECH J., vol. 6, art. no. 4, 2019.
2. T. Ono, D. Kanaoka, T. Shiba, S. Tokuno, Y. Yano, A. Mizutani, I. Matsumoto, H. Amano, and H. Tamukoh, Solution of World Robot Challenge 2020 Partner Robot Challenge (Real Space), Advanded Robotics, 2022, pp.870-889.
3. K. Isomoto, Y. Yano, Y. Tanaka, and H. Tamukoh, Robust trash can lid opening system, Proceedings of the 2023 International Workshop on Smart Info-Media Systems in Asia (SISA), 2023.
4. Y. Yano, K. Isomoto, T. Ono, and H. Tamukoh, Autonomous waiter robot system for recognizing customers, taking orders, and serving food, Proceedings of the 26 th RoboCup International Symposium, 2023.
5. T. Shiba, T. Ono and H. Tamukoh, Object Search and Empty Space Detection System for Home Service Robot, Proceedings of the 2023 International Conference on Artificial Life and Robotics (ICAROB2023), 2023.
6. D. Kanaoka, Y. Tanaka and H. Tamukoh, Applying Center Loss to Multidimensional Feature Space in Deep Neural Networks for Open-set Recognition, Proceedings of 17 th International Conference on Computer Vision Theory and Applications (VISAPP2022), Vol. 5, 2022, pp. 359-365.
7. A. Taniguchi, S. Isobe, L. El Hafi, Y. Hagiwara, and T. Taniguchi, Autonomous planning based on spatial concepts to tidy up home environments with service robots. Advanced Robotics, 2021.
8. Y. Ishida, T. Morie, and H. Tamukoh, A hardware intelligent processing accelerator for domestic service robots. Adv Robot. 2020.
9. S. Mupparaju, R. Thotakura, and C. Venkata, A review on YOLOv8 and its advancements, in Advances in Artificial Intelligence and Data Engineering, Springer, 2024.

10. T. Ono, and H. Tamukoh, A Study on Fast Pick-and-Place Method for Home Service Robots using 3D point clouds, The 2020 International Conference on Artificial Life and Robotics (ICAROB2020), OS20-3, Oita, Japan, January 13-16 (14), 2020.
11. S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, Automatic generation and detection of highly reliable fiducial markers under occlusion, Pattern Recognit., vol. 47, 2014, pp. 2280–2292.
12. L. Medeiros, Lang Segment Anything, https://lightning.ai/pages/community/lang segmentanything-object-detection-and-segmentation-with-text-prompt (Accessed: Sep. 2, 2023).
13. B. Calli et al., YCB benchmarking project: Object set, data set and their applications, J. Soc. Instrum. Control Eng., vol. 56, no. 10, 2017, pp. 792–797.

Authors Introduction

Mr. Natee Buttawong

He received his B.Eng. degree from from the Faculty of Engineering, King Mongkut's University of Technology North Bangkok in Thailand, in 2023. He is a Master’s degree student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include placing process task in home service robots.

Mr. Kosei Isomoto

He received the B.Eng. degree from Kyushu Institute of Technology, Japan, in 2023, respectively. He is currently a Master’s degree student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Japan. His research interests include brain-inspired artificial intelligence and home service robots.

Mr. Kosei Yamao

He received the B.Eng. degree from the National Institute of Technology, Ube College, Japan, in 2023. He is a Master’s degree student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include task planning and home service robots.

Dr. Ninnart Fuengfusin

He received his B.Eng. degree from King Mongkut's University of Technology Thonburi, Thailand, in 2016. He obtained his M.Eng. and D.Eng. degrees from Kyushu Institute of Technology, Japan, in 2018 and 2021, respectively. From 2021 to 2024, he was a postdoctoral researcher at Kyushu Institute of Technology, Japan. Currently, he is a specially appointed assistant professor at Kanazawa University. His research interests include efficient deep neural networks, fault-tolerant deep neural networks, and the digital hardware implementation of deep neural networks.

Prof. Hakaru Tamukoh

Hakaru Tamukoh received his B.Eng. degree from Miyazaki University, Japan in 2001. He received his M.Eng. and Ph.D. degrees from Kyushu Institute of Technology, Japan in 2003 and 2006, respectively. He was a postdoctoral research fellow of the 21st century center of excellent program at Kyushu Institute of Technology, from April 2006 to September 2007. He was an assistant professor at Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He joined the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Japan as an associate professor in February 2013. He has been a professor at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Japan, since April 2021. His research interests include hardware/software complex systems, digital hardware design, brain-inspired AI, neural networks, soft computing, and home service robots. He was the author of papers that won the Best Paper Award at IEEE/INNS IJCNN 2019, the Best Live Demonstration Award at IEEE ISCAS 2019, and the Best Paper Award at ICONIP 2013. He won the world championship of robot competition five times at RoboCup @Home 2017, 2018, 2024 and World Robot Challenge 2018, 2020. He received the METI Minister’s Award for Excellence in 2018 and 2021, and the RSJ Special Award from the Robotics Society of Japan in 2018. He is a member of IEEE, IEICE, SOFT, JNNS, JSAI, and RSJ.

Classification of Human Activity by Event-based Vision Sensors using Echo State Networks

Rohan Saini
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Aryan Rakheja
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Ryuta Toyoda
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Yuichiro Tanaka
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hakaru Tamukoh
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Email: saini-rohan413@mail.kyutech.jp, rakheja-aryan859@mail.kyutech.jp, toyoda.ryuta785@mail.kyutech.jp, tanaka-yuichiro@brain.kyutech.ac.jp, tamukoh@brain.kyutech.ac.jp

Abstract

We propose a system for human activity recognition using an event-based vision sensor (EVS) with echo state networks (ESNs). Conventional cameras are susceptible to motion blur and require computationally intensive methods, whereas EVS provides no motion blur and low latency. Our research aims to enable accurate recognition of human activities by using energy-efficient methods. Therefore, we adopt ESNs, which require low computational costs, for the classifier. Additionally, we use feature extraction algorithms such as optical flow and histogram of gradients to improve accuracy. We used an EVS activity recognition dataset created by us containing six human activities and 600 videos. The results showed that our hybrid approach outperformed several techniques. We achieved 89% accuracy when trained with ridge regression.

Keywords: Echo state networks (ESNs), Event-based vision sensor (EVS), Human action recognition

1. Introduction

Human activity recognition is a vital task in today's modern world and is used in many industries. Primarily, traditional cameras serve to accomplish this task, yet they present numerous drawbacks such as motion blur, increased latency, and reduced dynamic ranges. Convolutional neural networks (CNNs) are widely used for this purpose, but they have drawbacks like high power consumption and difficulty in handling temporal data.

To solve these problems, we propose a human activity recognition system with an event-based vision sensors (EVS) [1] and reservoir computing (RC) models, which complement our motive by efficiently processing temporal data to improve accuracy. EVS enables high-speed, low-power object recognition, transforming robotics and neuromorphic research. We used echo state networks (ESNs) as RC implementations along with two feature extraction methods namely optical flow [2] and histogram of oriented gradients (HOG) [3].

2. Related Works

This section provides recent related works utilizing EVS and RC to overcome the problems related to high computational cost, power consumption, and limitations in real-time performance. We also discuss the differences between our proposed method and the works in this section.

The study "Human Activity Recognition with Event-Based Dynamic Vision" [4], used an EVS for action recognition but relied primarily on recurrent neural networks, which can still be computationally intensive.

Another notable work, "Event-based timestamp image encoding network for human action recognition and anticipation" [5], proposed a multi-sensor fusion approach using RC to improve recognition accuracy. This approach displayed high computational cost because of heavy reliability on CNN and long short-term memory (LSTM).

3. Proposed Method

This study integrates ESNs with EVS for human action recognition. Figure 1 shows an example of videos recorded by EVS. The blue parts indicate negative events that are below the negative event threshold and the red parts are positive events above the positive threshold. The threshold and reference voltage are constant values. Before feeding the EVS videos to the ESNs, we applied two feature extraction methods to the videos: optical flow and (HOG).

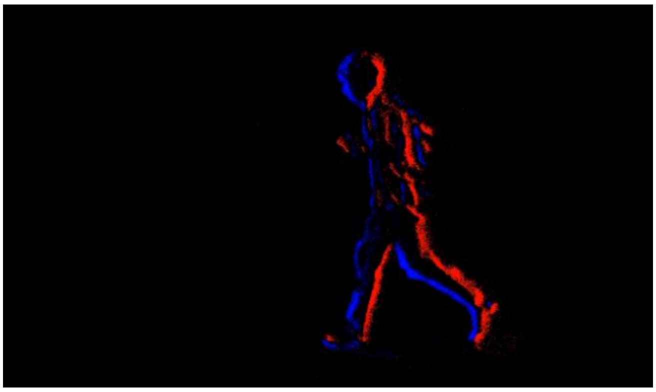


Fig. 1. Running person recorded by EVS

Optical Flow

Optical flow is a computer vision technique that quantifies the motion of objects by tracking the displacement of pixels between consecutive frames. The brightness change equation is:

$$I_x u + I_y v + I_t = 0 \quad (1)$$

Here, $\boldsymbol{I_x}$ and $\boldsymbol{I_y}$ are spatial gradients, and $\boldsymbol{I_t}$ is the temporal intensity change. *u* and *v* denote the horizontal and vertical flow components.

Histogram of Oriented Gradients (HOG)

The HOG computes gradient, which is the rate and direction of change in intensity for each pixel in an image, capturing texture and edge information.

$$G_x(r,c) = I(r,c+1) - I(r,c-1) \quad (2)$$
$$G_y(r,c) = I(r-1,c) - I(r+1,c) \quad (3)$$

In Equations 2 and 3, G_x and G_y are the gradients in x and y (horizontal and vertical) directions, r and c refer to the rows and columns, and *I* signifies the pixel intensity.

Echo State Networks (ESNs)

Figure 2 shows the complete architecture of the model, and how the feature input array is given to an ESN. ESNs use a reservoir of recurrently connected neurons to nonlinearly convert inputs into temporal patterns in a high-dimensional space. Weights in the hidden layer are fixed and not trainable. The sparsely hidden layer usually has less than 10% connectivity to make various patterns in the reservoir. Only output layer weights are trained so that the network generates outputs by utilizing the patterns in the reservoir and therefore, its training cost is low compared to normal neural networks. The reservoir state *x*(*t*) evolves as:

$$x(t) = f(W_{in}\, u(t) + W_{res}\, x(t-1)) \quad (3)$$

Here W_{in} and W_{res} are the weight matrices of the input to reservoir layers and in the reservoir layer, respectively. $u(t)$ is input, and f is a non-linear function that was tanh.

For the input to the network, we use either the optical flow feature or the HOG feature from an EVS video. For the readout of the network, we use either a linear model or the

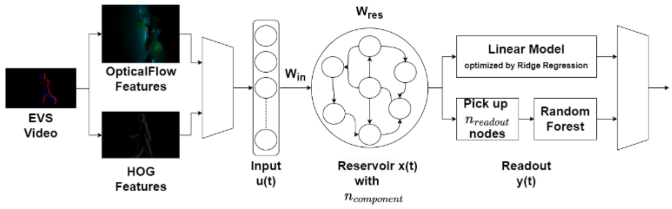


Fig. 2. Task flow for the ESN-based architecture

random forest. Therefore, the proposed framework includes four variants in total.

For the variant with the linear model, the output of the network *y*(*t*) is computed as follows:

$$y(t) = W_{out}x(t) \quad (4)$$

where W_{out} is the weight matrix between the reservoir and readout layers. The W_{out} is optimized by the ridge regression [6].

For the variant with random forest, the output of the network y can be expressed as:

$$y = f_{RF}(f_{ESN}(U, \theta_{ESN}), \theta_{RF}) \quad (5)$$

where f_{RF} is the random forest prediction, f_{ESN} is the ESN transformation function, U is the input signal, θ_{ESN} is $\{n_{readout}, n_{components}$, weight scaling$\}$, θ_{RF} is $\{n_{estimators}$, max depth$\}$. $n_{readout}$ is the number of units in the readout layer of the ESN, $n_{components}$ represents the number of internal units in the reservoir ESN, weight scaling is the adjustment of the strength of the connections between the input layer and the reservoir layer, $n_{estimators}$ represents the number of decision trees in the random forest, and max depth describes the maximum depth of each decision tree in the random forest.

4. Results and Discussion

We used a dataset consisting of 600 EVS videos (with 80% allocated for training and 20% for testing) and labeled four action classes.

Class 1: Represents the action of Walking.
Class 2: Represents the action of Jogging.
Class 3: Represents the action of Running.
Class 4: Represents the action of Boxing.

We used 500 nodes for the reservoir layer with 25% connectivity. The spectral radius of the reservoir weight connections was set as 0.59 to satisfy the echo state property, which is characteristic to ensure the reproducibility of the reservoir. We adopted a grid search to find the best parameters for the random forest. The maximum depth of the random forest (max depth) was 10 and the maximum number of $n_{estimators}$ was 75.

Figure 3 shows the graph for training accuracy with respect to the number of training videos. It depicts the training comparison between all variants we used. Table 1 shows the comparison of performance between the techniques we used.

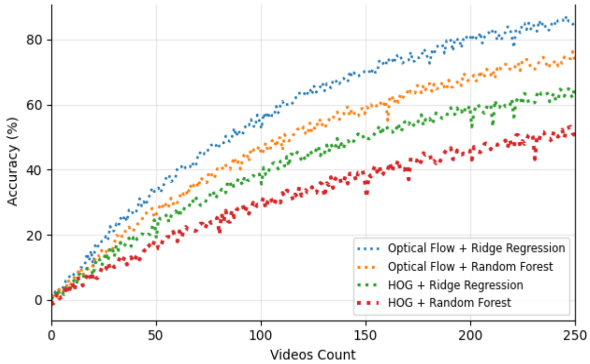


Fig. 3. Training Accuracy Graph

Table 1. Comparison of performances among variants

Feature Set	Training Method	Accuracy (%)
Optical Flow	Ridge Regression	89
Optical Flow	Random Forest	80
HOG	Ridge Regression	65
HOG	Random Forest	58

5. Conclusion

We proposed an EVS-ESN-based architecture with optical flow feature extraction demonstrating a promising accuracy of 89% with ridge regression and an 80% accuracy with random forest. This shows the potential of ESN for these use cases. For the future scope, to further improve this study, we could use more sophisticated model architectures like graph neural networks (GNNs) [7]. It could be because optical flow features represent motion and spatial dynamics by encoding them into a graph. GNNs could learn more spatial-temporal features than ESNs. Further, optical flow may not always align with Euclidean representation of data, which refers to a space or a data structure that does not follow the traditional Euclidean geometry. Non-Euclidean data has many features or dimensions, which cannot be represented in a traditional Euclidean space. GNNs excel in handling non-euclidean data.

Acknowledgments

This work received support from JSPS KAKENHI Grant Numbers 23H03468, as well as from JST ALCA-Next Grant Number JPMJAN23F3. This work was partially supported by Sony Semiconductor Solutions Corporation.

References

1. Sony Semiconductor Solutions Corporation, "Semiconductor Solutions for Electric Vehicles," 2024. [Online]. Available: https://www.sony-semicon.com/en/technology/industry/evs.html
2. S. S. Beauchemin and J. L. Barron, "The computation of optical flow," *ACM computing surveys (CSUR)*, vol. 27, no. 3, pp. 433–466, 1995.
3. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, pp. 886–893 vol. 1. doi: 10.1109/CVPR.2005.177.
4. P. Pansuriya, N. Chokshi, D. Patel, and S. Vahora, "Human activity recognition with event-based dynamic vision sensor using deep recurrent neural network," *International Journal of Advanced Science and Technology*, vol. 29, no. 4, pp. 9084–9091, 2020.
5. C. Huang, "Event-based timestamp image encoding network for human action recognition and anticipation," in *2021 International Joint Conference on Neural Networks (IJCNN)*, 2021, pp. 1–9.
6. G. C. McDonald, "Ridge regression," *Wiley Interdiscip Rev Comput Stat*, vol. 1, no. 1, pp. 93–100, 2009.
7. F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The Graph Neural Network Model," *IEEE Trans Neural Netw*, vol. 20, no. 1, pp. 61–80, 2009, doi: 10.1109/TNN.2008.2005605.

Authors Introduction

Mr. Rohan Saini

He received his Bachelor's degree in Engineering in 2024 from the Faculty of Computer Science and Engineering, Dronacharya College of Engineering in India. He is currently a master's student at Kyushu Institute of Technology, Japan.

Mr. Aryan Rakheja

He received his Bachelor's degree in Engineering in 2024 from the Faculty of Computer Science and Engineering, Dronacharya College of Engineering in India. He is currently a master's student at Kyushu Institute of Technology, Japan.

Mr. Ryuta Toyoda

He received the B.Eng. degree from Kyushu Institute of Technology, Japan, in 2023. He is a master's degree student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include developing processing circuits for EVS.

Dr. Tanaka Yuichiro

He received the B.E., M.E., and Ph.D. degrees from the Kyushu Institute of Technology, in 2016, 2018, and 2021, respectively. He was a Research Fellow with the Japan Society for the Promotion of Science (JSPS), from 2019 to 2021. He was an Assistant Professor at the Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology, from 2021 to 2024. He has been an Associate Professor at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology since 2024. His research interests include neural networks, digital hardware implementation, and home service robots. He is a member of IEICE and JNNS.

Prof. Hakaru Tamukoh

He received his B.Eng. degree from Miyazaki University, Japan in 2001. He received his M.Eng. and Ph.D. degrees from Kyushu Institute of Technology, Japan in 2003 and 2006, respectively. He was a postdoctoral research fellow of the 21st century center of excellent program at Kyushu Institute of Technology, from April 2006 to September 2007. He was an assistant professor at Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He joined the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Japan as an associate professor in February 2013. He has been a professor at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Japan, since April 2021. His research interests include hardware/software complex systems, digital hardware design, brain-inspired AI, neural networks, soft computing, and home service robots. He was the author of papers that won the Best Paper Award at IEEE/INNS IJCNN 2019, the Best Live Demonstration Award at IEEE ISCAS 2019, and the Best Paper Award at ICONIP 2013. He won the world championship of robot competition five times at RoboCup @Home 2017, 2018, 2024 and World Robot Challenge 2018, 2020. He received the METI Minister’s Award for Excellence in 2018 and 2021, and the RSJ Special Award from the Robotics Society of Japan in 2018. He is a member of IEEE, IEICE, SOFT, JNNS, JSAI, and RSJ.

Integrating Advance Speech Recognition and Human Attribute Detection for Enhanced Receptionist Tasks in RoboCup@Home

Koshun Arimura
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Yuga Yano
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Takuya Kawabata
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hakaru Tamukoh
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Email: arimura.koshun523@mail.kyutech.jp, yano.yuuga158@mail.kyutech.jp, kawabata.takuya790@mail.kyutech.jp, tamukoh@brain.kyutech.ac.jp

Abstract

RoboCup@Home is held to integrate service robots into society. It includes a task called "Receptionist" that evaluates Human-Robot Interaction. In this task, a robot must ask guests for their names and favorite drinks and guide them to available seats. Additionally, the robot must introduce the guest's features, such as their clothing, to others. We developed a system integrating speech recognition and human attribute detection to achieve these functions. The robot can determine which seat a person is sitting in by detecting the person's skeletal coordinates. Additionally, the robot can identify individuals by recognizing human attributes. To verify the effectiveness of the developed system, we participated in the Receptionist task at RoboCup@Home 2024. We won first place in our league and demonstrated the effectiveness of our system.

Keywords: Human-robot interaction, Speech Recognition, Human feature detection

1. Introduction

In recent years, research in Human-Robot Interaction (HRI) has explored how humans and robots should interact in shared spaces [1], [2], [3]. RoboCup@Home [4] is a global competition that evaluates robotic capabilities across nine tasks, including HRI, to accelerate service robots' development for society. Among these, the task that places particular importance on the HRI is a Receptionist task, as shown in Fig. 1. This task simulates a party scenario in home environment, where a robot guides guests to vacant seats and introduces them to a host.

The Receptionist task requires two primary functions of a robot. One is speech recognition. The robot asks guests for their names and favorite drinks and then receives their voice responses. The other is human position and feature detection. The robot should recognize which chairs are occupied to guide guests to vacant seats. Furthermore, the robot must identify who is seated in which chair and introduce guests to the host or another guest. To achieve that, the robot is required to associate names with individual characteristics and memorize these details.

Fig. 1 Functions required by the Receptionist task

Based on the above competition rule, the Receptionist task demands an advanced system integrating speech and human recognition. Therefore, it is meaningful for improving HRI in home environments. Furthermore, this task is carried out in a dynamic environment where new chairs may be added right before the task starts, or guests can freely move between seats. For these reasons, the environment makes the task highly practical.

A solution for the Receptionist task was proposed in [5]. This method got the highest score in RoboCup@Home 2022 but faced some issues when dealing with additional chairs and guests switching seats. In this study, we developed an integrated system that extended the conventional method to address these challenges and complete the Receptionist task. To verify the effectiveness of the proposed system, we participated

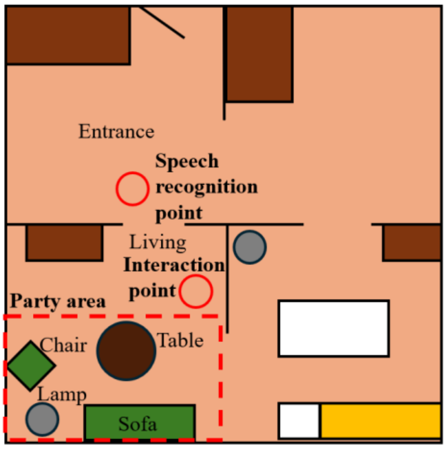


Fig. 2 Map of the arena where the Receptionist task takes place

in RoboCup@Home 2024, held in the Netherlands. In the Receptionist task, the proposed system achieved the highest score compared to other teams using the same robot and demonstrated its effect.

2. RoboCup@Home

2.1. *Domestic standard platform league*

The Domestic Standard Platform League (DSPL) is a division of the RoboCup@Home competition where teams compete using a standardized robot platform. In this league, participants have to focus on software development and algorithm design within a unified hardware environment. Thanks to this standardized platform, competition results are evaluated based on the quality of the software implementation rather than the robot's physical performance. In this study, we validated the effectiveness of the proposed system in the Receptionist task of the DSPL league.

2.2. *Human support robot*

In RoboCup@Home, we used the Human Support Robot (HSR) [6] developed by TOYOTA MOTOR CORPORATION. This robot has a camera and microphone on its head, which makes it suitable for performing HRI tasks. Furthermore, it has an arm that helps it point to specific locations to convey positional information.

2.3. *Receptionist task*

Fig. 2 shows the map for the Receptionist task. Guest 1 enters the entrance room at the task's start, and the robot greets Guest 1. The robot asks Guest 1's name and favorite drink. The robot then guides Guest 1 to the living room and suggests a vacant seat. Since one host is already sitting in the living room, the robot introduces Guest 1's name and favorite drink to the host.

Next, the robot returns to the entrance room to greet Guest 2. It asks Guest 2's name and favorite drink in the same way and then guides Guest 2 to a vacant seat. The robot then introduces Guest 2's name and favorite drink to Guest 1 and the host. Finally, the robot introduces Guest 1's features to Guest 2, such as their clothing.

Throughout the task, the robot needs to look toward the person talking. Since people sitting in chairs might move to different seats during the task, the robot must identify who is sitting in which chair. In the living room, some chair's positions are known in advance, and others are arranged in the room right before the task starts. That means that the robot knows the positions of the disclosed chairs beforehand but doesn't know the positions of the newly added chairs at the task's start.

3. Implemented skills

3.1. *Speech recognition*

We utilize Whisper [7] as a speech recognition model to identify people's names and favorite drinks. Whisper is a highly accurate, multilingual speech recognition model. However, Whisper sometimes misrecognize words with similar pronunciations as different words. To address this issue, we predefined a list of names and drinks and calculated the similarity between the recognized text and each element in the list. We adopted the corresponding string as the recognized text when the similarity exceeded a certain threshold.

3.2. *Human pose estimation*

We utilize OpenPose [8] as a skeleton detection model to estimate human positions. OpenPose detects key points of the human body from 2D images captured by a camera. It also supports multi-person detection within a single frame. Furthermore, by integrating the 2D key point with depth images captured by an RGB-D camera, we calculate the 3D skeletal coordinates of individuals. This method can judge whether a person is in the entrance room or identify the locations of people in the living room.

3.3. *Human attribute estimation*

We utilize Class-Specific Region Attention (CSRA) [9] to estimate human attributes. CSRA enhances the recognition of appearance, such as clothing, by applying spatial attention scores tailored to each attribute. These scores highlight the specific positions and looks of objects within an image. In the Receptionist task, the robot should introduce the looks of Guest 1 to Guest 2. To achieve this function, our system used these features obtained through CSRA. In this system, the robot selects features with high likelihood scores above a set threshold and says these selected features in speech to describe human attributes. The robot can recognize and articulate 14 attributes, including gender and clothing. In this study, the system was extended to identify people even after they changed seats by associating their features with their names.

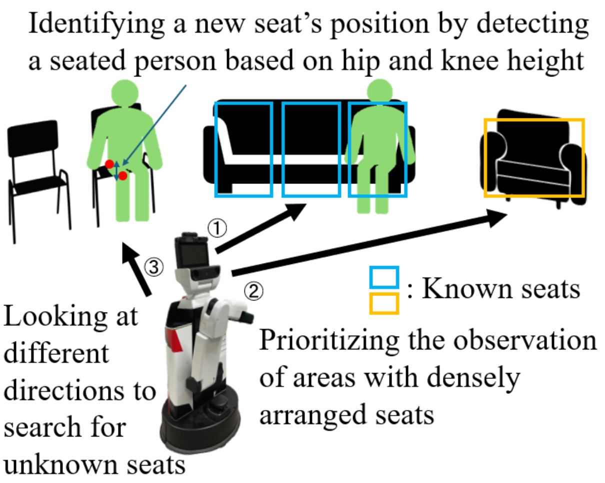


Fig. 3 Vacant seats detection system

4. Proposed system

4.1. *Welcoming guests at the entrance*

First, the robot detects the guest using OpenPose and requests the guest to stand in front of the robot through verbal instructions. Next, the robot asks the guest for their name, repeats the recognized name, and confirms its accuracy by asking for a "Yes" or "No" response. Similarly, the robot inquires about the guest's favorite drink and verifies the recognition. The robot also uses CSRA to detect the guest's attributes during this process. The robot relates these attributes with the recognized name. This approach is designed to avoid the mismatch between names and attributes.

4.2. *Showing guest an available seat at the living*

Fig. 3 shows an overview diagram of the vacant seat detection system. First, the robot moves to the interaction point in the living room. It then surveys each seat in the living room to detect people. If the distance between a chair's predefined coordinates and a detected person's central coordinates is below a certain threshold, the robot determines that the seat is occupied. Our system employs a Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [10] algorithm to avoid excessive time consumption from sequentially checking each seat. Specifically, it prioritizes checking the central coordinates of densely clustered chairs, reducing the required detections and shortening processing time. Finally, the robot suggests the vacant seat nearest to the guest.

The technique of using distance between people and chairs is effective for chairs whose positions are predefined. However, new chairs are added to this competition before the task starts. Therefore, the task begins even if the robot does not know the positions of the newly added chairs. To address this issue, we implemented a method that estimates the sitting posture using skeletal data and regards the position as the

Table. 1 Result of the Receptionist task

Action	Score	1st try	2nd try
Main Goal			
Navigate the guest to the other guest	**2 × 15**	**30**	**15**
Look in the direction of navigation	**2 × 50**	**100**	**50**
Introduction a new guest to others	**2 × 50**	**100**	**50**
Offer a free seat to the new guest	**2 × 100**	**200**	**100**
Look at the person talking	**2 × 25**	**50**	**50**
Look at the person the robot is introducing the guest to	**2 × 50**	**100**	**50**
Qualitative robot social performance	**50**		
Bonus Rewards			
Open the entrance door for a guest	**2 × 100**		
Describe the first guest to the second guest (per correct attribute)	**4 × 30**		
Describe the first guest to the second guest (per incorrect attribute)	**4 × -30**		
Use standard microphone	**2×5**	**10**	**5**
Penalties			
Wrong guest information was memorized (continue with wrong name or drink)	**-50**		**-50**
Persistent inappropriate gaze (away from conventional partner)	**-50**		
Persistent gaze not in the direction of the navigation while moving	**-10**		
Score per try	**960**	**590**	**270**

coordinates of the newly added chair. Specifically, if the height difference between a person's hip and knee key points is below a certain threshold, the person is assumed to be sitting. This approach accurately identifies the position of a chair as the one a person is sitting on, even if the chair's location is unknown.

4.3. *Introducing a guest to others*

When the robot introduces a guest to the host or other guests, it needs to point toward the person being introduced while looking at the others. Therefore, the robot must accurately identify who is sitting in each seat. To achieve this challenge, we matched the attributes obtained by using CSRA to determine the person's name. This method enables the robot to identify the individuals seated in each position.

5. Competition result

We tested the efficacy of the proposed system through the Receptionist task in RoboCup@Home 2024. In the Receptionist task, points are awarded based on subtasks such as speech recognition, seat guidance, and guest introductions for two guests. The results of our first and second trials are shown in Table. 1. In the first trial, we completed most interactions but ran out of time and needed more time to finish the guest introduction phase. To address this issue, we replaced the Yes/No speech recognition with a touch function using the robot's hand to save time in the second trial. However, during the trial, an error occurred in the speech recognition process, and we were ultimately unable to complete the task within the time limit. Despite these challenges, our system ranked first in the DSPL league for the Receptionist task.

6. Conclusion

We extended the conventional Receptionist task system to handle dynamic environments, such as the addition of chairs and people changing seats. We tested its performance in the Receptionist task of RoboCup@Home 2024. As a result, our system achieved first place in the DSPL league and demonstrated its effect.

Acknowledgments

This research is based on results from a JPNP16007 project commissioned by the New Energy and Industrial Technology Development Organization (NEDO). This research received support from JSPS KAKENHI Grant Number 23H03468 and 23K18495, as well as from JST ALCA-Next Grant Number JPMJAN23F3.

References

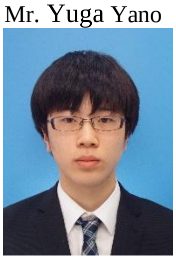
1. Y. Yano, I. Matsumoto, Y. Fukuda, T. Ono and H. Tamukoh, Proposal for Solution of Human Interaction Task in RoboCup@Home, JSAI SIG on AI Challenge, SIG-Challenge-060-02, 2022.
2. K. Yamao, D. Kanaoka, K. Isomoto, A. Mizutani, Y. Tanaka and H. Tamukoh, Development of A SayCan-based Task Planning System Capable of Handling Abstract Nouns, Proceedings of the 2024 International Conference on Artificial Life and Robotics, 2024, pp. 430-434.
3. Y. Yano, A. Mizutani, Y. Fukuda, D. Kanaoka, T. Ono and H. Tamukoh, Unified Understanding of Environment, Task, and Human for Human-Robot Interaction in Real-World Environments, The 33rd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2024), 2024.
4. RoboCup at Home, RoboCup@Home Official Website, available at: https://athome.robocup.org/ (accessed December 5, 2024)
5. Y. Yano, Y. Fukuda, T. Ono and H. Tamukoh, Flexible Human-Robot Interaction in Domestic Environment Using Semantic Map, Proceedings of the 2023 International Conference on Artificial Life and Robotics, 2023, pp. 409-414.
6. T. Yamamoto, K. Terada, A. Ochiai, F. Saito, Y. Asahara and K. Murase, Development of Human Support Robot as the research platform of a domestic mobile manipulator, ROBOMECH Journal, Vol. 6, Art. No. 4, 2019.
7. A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey and I. Sutskever, Robust Speech Recognition via Large-Scale Weak Supervision, OpenAI, 2022. Available at: https://github.com/openai/whisper (accessed December 5, 2024)
8. Z. Cao, T. Simon, S.-E. Wei and Y. Sheikh, Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 7291-7299.
9. K Zhu and J Wu, Residual Attention: A Simple but Effective Method for Multi-Label Recognition, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 184-193.
10. M. Ester, H.-P. Kriegel, J. Sander and x. Xu, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD), AAAI Press, 1996, pp. 226-231

Authors Introduction

Mr. Koshun Arimura

He received his B.Eng. degree from Kyushu Institute of Technology, Japan, in 2024. He is currently a master's student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include home service robots.

Mr. Yuga Yano



He received his B.Eng. degree from Kyushu Institute of Technology, Japan, in 2022. He received his M.Eng. from Kyushu Institute of Technology, Japan, in 2024. He is currently a Ph.D. student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include image processing, autonomous driving, and home service robots.

Mr. Takuya Kawabata

He received his B.Eng. degree from Kyushu Institute of Technology, Japan, in 2023. He is currently a master's student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include neuromorphic physical computing.

Prof. Hakaru Tamukoh

He received the B.Eng. degree from Miyazaki University, Japan, in 2001, and the M.Eng. and Ph.D. degrees from the Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a Postdoctoral Research Fellow at the Kyushu Institute of Technology from April 2006 to September 2007. He was an Assistant Professor with the Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He is currently a Professor at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include digital hardware design, neural networks, and home service robots.

Classification of Human Activity by Spiking Neural Networks using Event-based Vision Sensors

Aryan Rakheja
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Rohan Saini
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Ryuta Toyoda
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Yuichiro Tanaka
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hakaru Tamukoh
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Email: rakheja-aryan859@mail.kyutech.jp, saini-rohan413@mail.kyutech.jp, toyoda.ryuta785@mail.kyutech.jp, tanaka-yuichiro@brain.kyutech.ac.jp, tamukoh@brain.kyutech.ac.jp

Abstract

This paper presents a human action classification system using a spiking neural network (SNN) and an event-based vision sensor (EVS). The EVS captures asynchronous data with high temporal resolution, wide dynamic range, and motion blur resistance. SNNs, inspired by biological neurons, process this data event-driven, ensuring energy efficiency, low latency, and scalability. Using a custom EVS dataset of 600 videos across four action types and the optical flow for feature extraction, the system achieved 93% accuracy, offering an efficient solution for action recognition.

Keywords: Spiking neural network (SNN), Event-based vision sensor (EVS)

1. Introduction

Human action classification involves automatically identifying actions like running, walking, or gesturing from visual input, with applications in surveillance, healthcare, and human-computer interaction. Traditional methods often face challenges like motion blur, limited dynamic range, and high computational demands, especially in dynamic or low-light environments. Event-based vision sensors (EVS) provide solution by capturing asynchronous data triggered by scene changes, offering high temporal resolution, a wide dynamic range, and resistance to motion blur. These features make EVS ideal for human action classification [1]. In this paper, we propose integrating an EVS with a spiking neural network (SNN), a biologically inspired method that processes sparse, event-driven data, reducing computational complexity and energy consumption [2].

Recent studies have explored integrating EVS with SNNs for tasks like object and action recognition. One approach uses a hierarchical SNN architecture for action recognition, showing effectiveness on the dynamic vision sensor dataset. The SLAYER (spike layer error reassignment) training mechanism has achieved high gesture recognition accuracy, while another study introduces a spiking convolutional recurrent neural network to capture spatial and temporal correlations in event-based data [3], [4]. Our work uses the EVS with a feature extraction by the optical flow for efficient human action classification.

2. Methodology

2.1. *Data collection using event-based vision sensor*

We collected the dataset for human action classification using the EVS, capturing asynchronous event data from dynamic scenes. It comprises four actions: running, jogging, walking, and boxing, recorded under varied lighting and motion conditions to replicate real-world environments. To ensure diverse perspectives for each action, we positioned the sensor at specific angles. The blue parts indicate negative events that fall below the negative event threshold, while the red parts represent positive events exceeding the positive threshold, as shown in Fig. 1.

2.2. *Feature Extraction and Preprocessing*

We used the optical flow to extract motion-related features from the asynchronous event data. The optical flow measures the motion of objects between consecutive events by analyzing changes in pixel intensity over time.

The equation for estimating optical flow between two consecutive event frames is given as follows,

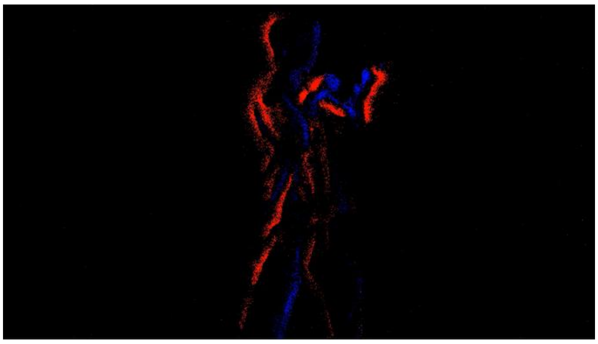


Fig. 1. Boxing action recorded by EVS.

$$I_t + I_x u + I_y v = 0 \tag{1}$$

where the temporal derivative of image intensity is denoted as I_t. I_x and I_y are the spatial derivatives, and u and v represent the velocity components in the x and y directions, respectively.

2.3. *SNN for Event-Driven Classification*

We used an SNN to process the event-driven data. In SNNs, neurons accumulate input over time and fire when the membrane potential exceeds a threshold. The membrane potential of the i-th neuron at time t is updated as:

$$V_i(t) = V_i(t-1) + \sum_J W_{ij} S_j(t) \tag{2}$$

where $V_i(t)$ is the membrane potential of the i-th neuron at t. W_{ij} represent the synaptic weight from the j-th to i-th neuron. $S_j(t)$ represents the spike train of the j-th neuron at time t, indicating whether the neuron spikes (1) or not (0).

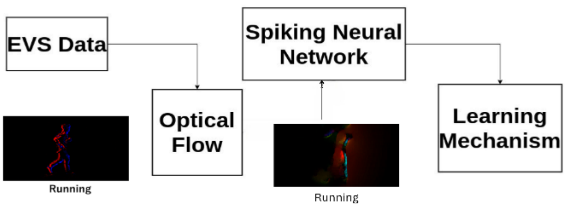


Fig. 2. Proposed system that processes EVS data using the optical flow for feature extraction and performs classification using an SNN.

As shown in Fig. 2, our proposed system processes EVS data by first extracting features through optical flow and then passing the data into the SNN for classification. The architecture is specifically tailored for event-based data. The SNN model processes event-driven data, with key parameters such as the membrane time constant and a spike threshold optimized for performance. The model incorporates decay and reset factors to regulate membrane potential dynamics, ensuring stability and effective signal processing. The learning mechanism involves mapping the preprocessed features to corresponding action labels, with synaptic weight scaling, firing thresholds, and membrane potential decay rates further tuned for optimal results. This integration of optical flow and the SNN ensures robust classification performance for various action recognition tasks. The model architecture consists of an input layer with 1024 neurons, four hidden layers with 512, 256, 128, and 64 neurons, respectively, and an output layer with 4 neurons (one per class).

3. Results and Discussion

We conducted an experiment using a dataset of 600 EVS videos. To ensure robust evaluation, we used 80% of the videos for training and 20% for testing. During training, the network mapped the preprocessed optical flow features to corresponding action labels, with a learning rate of 1×10^{-3} and 10 epochs.

Table 1. Classification table for different type of action.

Action	Precision	Recall	Accuracy
Boxing	1.00	1.00	1.00
Jogging	0.83	0.83	0.83
Running	1.00	0.80	0.90
Walking	0.90	1.00	0.95
Total	0.93	0.90	0.93

The training process, as shown in Fig. 3 demonstrates a steady decline in loss from 1.3 to 0.6 over 10 epochs, indicating that the model is effectively learning and converging towards an optimal solution. The proposed system achieved 93% classification accuracy on the testing set, demonstrating its capability to recognize human action.

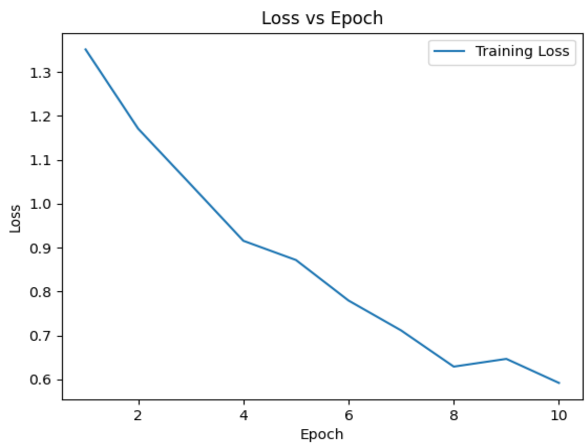


Fig. 3. Training loss over epoch.

As summarized in Table 1, the system has achieved high precision and recall, with boxing and walking performing exceptionally well (1.00 recall for both). While jogging and running exhibit slightly lower recall, their precision remains strong. This combination of efficient training, high accuracy, and strong precision and recall highlights the model's effectiveness for real-time human action recognition.

4. Conclusion

This paper presents an EVS and SNN integrated system for human action classification. By utilizing the event-driven nature of EVS, which only responds to changes in the scene, and the spike-based communication of SNN, we achieved 93% accuracy in classifying human actions. The membrane potential dynamics and spike-based communication between neurons further contributed to the model's effectiveness. For future work, expanding the dataset to include more diverse actions and exploring more advanced SNN architectures, such as hierarchical or recurrent networks, could improve robustness and scalability.

Acknowledgements

This work received support from JSPS KAKENHI Grant Numbers 23H03468, as well as from JST ALCA-Next Grant Number JPMJAN23F3. This work was partially supported by Sony Semiconductor Solutions Corporation.

References

1. Sony Semiconductor Solutions Corporation, “Semiconductor Solutions for Electric Vehicles,” 2024.
2. S. Ghosh-Dastidar and H. Adeli, “Spiking Neural Networks,” Int J Neural Syst, vol. 19, no. 4, pp. 295–308, 2009.
3. Y. Xing, G. Di Caterina, and J. Soraghan, “A New Spiking Convolutional Recurrent Neural Network (SCRNN) with Applications to Event-Based Hand Gesture Recognition,” Front Neurosci, vol. 14, 2020.
4. Q. Liu, D. Xing, H. Tang, D. Ma, and G. Pan, “Event-based Action Recognition Using Motion Information and Spiking Neural Networks,” pp. 1743–1749, 2021.

Authors Introduction

Mr. Aryan Rakheja

He received his bachelor's degree in engineering in 2024 from the Faculty of Computer Science and Engineering, Dronacharya College of Engineering in India. He is currently a master student in Kyushu Institute of Technology, Japan.

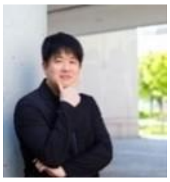
Mr. Rohan Saini

He received his bachelor's degree in engineering in 2024 from the Faculty of Computer Science and Engineering, Dronacharya College of Engineering in India. He is currently a master student in Kyushu Institute of Technology, Japan.

Mr. Ryuta Toyoda

He received the B.Eng. degree from Kyushu Institute of Technology, Japan in 2023. He is a master student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include developing processing circuits for EVS.

Dr. Yuichiro Tanaka



He received the B.E., M.E., and Ph.D. degrees from the Kyushu Institute of Technology, in 2016, 2018, and 2021, respectively. He was a research fellow with the Japan Society for the Promotion of Science (JSPS), from 2019 to 2021. He was an assistant professor at the Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology, from 2021 to 2024. He has been an associate professor with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology since 2024. His research interests include neural networks, digital hardware implementation, and home service robots. He is a member of IEEE, IEICE, and JNNS.

Dr. Hakaru Tamukoh

He received the B.Eng. degree from Miyazaki University, Japan, in 2001, and the M.Eng. and Ph.D. degrees from the Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a Postdoctoral Research Fellow at the Kyushu Institute of Technology, from April 2006 to September 2007. He was an Assistant Professor with the Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He is currently a Professor with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include digital hardware design, neural networks, and home service robots.

Robotic Grasping of Common Objects: Focusing on Edge Detection for Improved Handling

Tomoya Shiba
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hakaru Tamukoh
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Email: shiba.tomoya627@mail.kyutech.jp, tamukoh@brain.kyutech.ac.jp
https://www.brain.kyutech.ac.jp/~tamukoh/

Abstract

Grasping objects like plates and cups poses unique challenges for robots because of their irregular shapes and the difficulty of finding reliable grasp points. Traditional approaches often attempt to grasp the object at its center, but this strategy tends to fail for items like plates or cups, whose shapes deviate from simple forms like cubes or spheres. To address this issue, we propose a new method that utilizes AI-powered image analysis to identify the best edges for grasping. Through experiments conducted with a home service robot and a set of YCB objects, we evaluated the effectiveness of our approach compared to conventional methods. The results revealed a significant improvement in the success rate, particularly for objects with prominent edges, such as cups.

Keywords: Object recognition, Dataset, Service robot, Mobile manipulator, RoboCup@Home,

1. Introduction

Autonomous mobile robots have become more prevalent in a variety of industries in recent years, from supermarkets to restaurants [1],[2],[3]. These robots have primarily been employed thus far for transportation-related jobs that need their ability to navigate. Nevertheless, we may significantly improve their functionality by giving them robotic arms, so transforming them into mobile manipulators [4].

The capacity to grip objects is a crucial characteristic that sets mobile manipulators apart from simple transport robots. Using items with predetermined placements and orientations or placing marks where objects should be held are the easiest ways to guarantee a solid grasp. However, because service robots are designed to operate in human-centered settings, they must be able to recognize appropriate grasp points on their own without assistance from others.

Several approaches that integrate depth information with RGB-based object recognition have been put forth to address this problem. For simple geometric shapes like cuboidal items (like boxes) and cylindrical objects (like beverage cans), these methods enable robots to estimate object size and identify grab locations [5]. Additionally, more sophisticated methods have been created by utilizing neural networks that have been trained on specific datasets, such as depth pictures and 3D point clouds [6]. These techniques do have a major disadvantage, though, in that they necessitate a great deal of dataset preparation and 3D annotation, which makes large-scale data collecting extremely time-consuming and impracticable in many situations.

By using an effective grasp point estimate framework that just uses RGB images and does not require 3D point clouds or depth information, our suggested approach, which is shown in (Fig. 1), seeks to get around these restrictions. We can train reliable models while drastically cutting down on the overhead related to conventional data gathering techniques by employing a simulator-driven dataset generation approach.

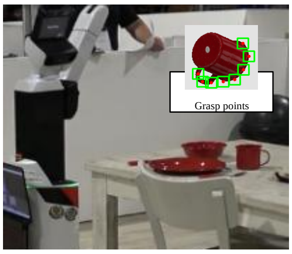


Fig. 1. Select Grasping Points

In this study, we provide a dataset generating method in this study that makes use of a simulator. The approach makes it possible to efficiently create large-scale training datasets by randomly placing different items within a virtual environment and automatically annotating grasp point information. Additionally, we want to show that grasp point estimation may be done without the usage of 3D point cloud data by using object recognition based just on RGB images.

We experiment using YOLOv8 [7] for object recognition in order to test the efficacy of the suggested method and determine how it affects grip accuracy in practical situations. YCB objects (Fig.2) [8], which are frequently used to benchmark grasping tasks, are used for the evaluation.

Fig. 2. YCB Object dataset (kitchen item) [8]

2. Related works

2.1. *Object recognition*

Previous studies have explored methods for home service robots to detect the locations of people and objects and respond accordingly. Research has been conducted on object recognition techniques for tasks such as tidying up and detecting people in home environments. Approaches utilizing YolactEdge [9] and point cloud data [10] have been proposed for estimating object positions and identifying available space.

2.2. *Simulator-Based Dataset Generation*

Studies on simulator-based dataset generation have demonstrated that training solely with simulated data can produce models capable of performing effectively in real-world robotic applications. These findings suggest that properly designed simulation environments can serve as a viable alternative to costly and labor-intensive real-world data collection and annotation [11].

Fig. 3. Dataset Generator

3. Proposed Method

3.1. *Overview*

In this study, we propose an automatic dataset generation and detection system for identifying graspable regions of objects using RGB images. The dataset generation system captures images from multiple angles and automatically annotates them using 3D models of target objects and a PyBullet-based physics simulator [12]. For the detection system, we adopt YOLOv8, which offers high-speed object detection while maintaining a relatively lightweight architecture.

3.2. *Grasp-Point Definition*

Since our method utilizes an object detection model, we define the center of the bounding box as the grasp point. As a result, grasp points are recorded as bounding boxes in the RGB images captured by the simulator, as illustrated in (Fig. 4).

For each target object, grasp points are predefined in its 3D model by identifying key regions suitable for grasping. Specifically, areas such as the rim or handle of a cup are selected based on the 3D mesh structure to ensure the model learns practical grasping locations. During the annotation process, a 2D bounding box is placed around each predefined grasp point, covering 10–30% of the object's total size, allowing the detection model to effectively learn and generalize graspable areas.

Fig. 4. Annotated Grasp Points (Mug / Bowl / Cup)

3.3. *Dataset Generation*

To generate training data, we capture images of objects within a 3D physics simulator using PyBullet , as illustrated in (Fig. 5). Inspired by previous research [11], we apply domain randomization techniques to minimize the gap between simulated and real-world images. Specifically, we vary camera angles, adjust lighting conditions, and change background colors randomly to improve the model's generalization and robustness. These variations ensure that the dataset includes a diverse range of scenarios, helping the trained model adapt better to real-world conditions.

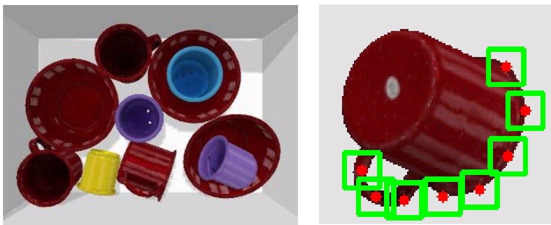


Fig. 5. Simulator-Based Dataset

4. Experiments and Results

4.1. *Recognition experiments*

We conducted an experiment using Toyota's HSR [13][14][15], a mobile manipulator, to verify whether it could correctly select grasp points. The effectiveness of the proposed method was evaluated by performing object grasping.

For the object grasping dataset, we used the YCB dataset. In this experiment, we selected bowls, mugs, and cups as target objects for grasp point selection. The training dataset consisted of 10,000 images, with each image containing at least two overlapping objects. The model was trained exclusively on simulator-generated images. The validation dataset consisted of images captured by the robot's head-mounted camera, showing objects placed on the floor. The results are shown in (Table. 1).

Table. 1. Performance

Object	BBox mAP	Grasp-Point MPJPE[mm]
Mug	75.3 %	28.1 [mm]
Bowl	72.3 %	32.3 [mm]
Cup	81.8 %	12.7 [mm]

4.2. *Grasping experiments*

Additionally, we conducted a comparison with a conventional grasping system. The training settings were identical to those in Experiment 4.1. The conventional system followed the method proposed by Ono et al. [11], which attempted to grasp the center of an object. The robot attempted to grasp three types of objects—bowls, mugs, and cups—placed on the floor in front of it. The objects were positioned randomly in different orientations, and the robot performed 20 grasping attempts for each object. A grasp was considered successful if the robot lifted the object from the floor. The results are presented in (Table. 2).

Table. 2. Grasping Performance Evaluation

Object	**Proposed Method**	Conventional Method	Success Rate Diff
Mug	**18/20 (90%)**	17/20 (85%)	+5%
Bowl	**12/20 (60%)**	5/20 (25%)	+35%
Cup	16/20 (80%)	16/20 (80%)	0%

5. Discussion

To evaluate the grasp point estimation AI, we analyzed both bounding box mean Average Precision (BBox mAP) and Mean Per Joint Position Error (MPJPE), as summarized in (Table. 1). The model performed well in detecting graspable regions across different objects, with BBox mAP values ranging from 72.3% to 81.8%. However, the accuracy of grasp point estimation varied depending on the object shape. The cup had the lowest MPJPE (12.7 mm), indicating high precision, while the bowl had the highest error (32.3 mm), suggesting that grasp points were more difficult to estimate. This difference likely stems from the object's geometry—cups tend to have more well-defined grasp points, whereas bowls, with their curved surfaces, pose a greater challenge.

To further assess how well the model's grasp point predictions translated into actual grasping success, we conducted experiments using a mobile manipulator. The results, shown in (Table. 2), indicate that the proposed method improved grasping success rates for the mug (+5%) and bowl (+35%) compared to the conventional approach. Notably, the success rate for cups remained unchanged at 80%, suggesting that the conventional method was already effective for this particular object type. The large performance gap observed for the bowl highlights the advantage of using AI-based grasp point estimation for objects where a precise grasp is harder to determine.

Overall, these findings suggest that the proposed AI model enhances grasping accuracy, especially for objects with irregular shapes. However, the higher MPJPE for the bowl indicates a potential limitation in predicting grasp points for objects with smooth, less-defined surfaces. Further refinements, such as incorporating additional geometric features or optimizing the dataset to include more varied object poses, may help improve the model's performance in these cases.

6. Conclusion

In this work, we proposed a grasp point estimate method that does not require 3D point clouds or depth information, and instead uses only RGB images. Our methodology maintains good gripping accuracy while drastically reducing the workload associated with manual data collecting and annotation by utilizing a simulator-based dataset generation mechanism.

Results from experiments showed that our suggested approach, especially for structured objects like cups and mugs, was able to accurately identify graspable zones. Our method increased the grip success rate for difficult objects with less distinct grasp points, such as bowls, when compared to traditional grasping strategies. Performance research also showed that the accuracy of grasp point estimate differed based on the shape of the object, suggesting a possible drawback when working with objects that had smooth or uneven surfaces.

Future work will focus on further enhancing grasping performance by integrating additional geometric features and refining dataset diversity. Additionally, real-world deployment and testing will be conducted to validate the model's adaptability beyond simulated environments. We believe that our approach contributes to advancing robotic grasping in home service applications and provides a scalable solution for training grasping models efficiently.

7. Acknowledgements

This research is based on results from a JPNP16007 project commissioned by the New Energy and Industrial Technology Development Organization (NEDO). This research received support from JSPS KAKENHI Grant Number 23H03468 and 23K18495, as well as from JST ALCA-Next Grant Number JPMJAN23F3.

References

1. T. Yamamoto, K. Terada, A. Ochiai, F. Saito, Y. Asahara, and K. Murase, ROBOMECH Journal, 2019.
2. L. Iocchi, D. Holz, J. Ruiz-del-Solar, K. Sugiura, T. van der Zant, Artificial Intelligence, pp. 258-281, 2015.
3. H. Okada, T. Inamura, and K. Wada, Advanced Robotics, 2019.
4. Ghodsian, Nooshin, et al. "Mobile manipulators in Industry 4.0: A review of developments for industrial applications." *Sensors* 23.19 (2023): 8026.
5. O. Khatib,"Mobile manipulators: Expanding the frontiers of robot applications." *Field and Service Robotics*. London: Springer London, 1998.
6. Fang, Hao-Shu, et al. "Graspnet-1billion: A large-scale benchmark for general object grasping." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.
7. Ultralytics "Yolov8 docs", (2024 12/12 accessed)
8. B. Calli, et al. "The ycb object and model set: Towards common benchmarks for manipulation research." *2015 international conference on advanced robotics (ICAR)*. IEEE, 2015.
9. H. Liu, Soto, R. A. R., Xiao, F., & Lee, Y. J. (2021, May). Yolactedge: Real-time instance segmentation on the edge. In 2021 IEEE International Conference on Robotics and Automation (ICRA) (pp. 9579-9585). IEEE.
10. T. Shiba, T. Ono, and H. Tamukoh, Object Search and Empty Space Detection System for Home Service Robot, The 2023 International Conference on Artificial Life and Robotics (ICAROB2023) OS17-2, Online, February 9-12(10), 2023.
11. T. Ono. "Establishing Fundamental Technologies for Home Service Robots: Data Generation for Sim2Real and Its Application to Pick-and-Place" (Kateiyo Sabisu Robotto no tame no Kiban Gijutsu no Kakuritsu: Sim2Real wo Jitsugen suru Data Seisei to Pick-and-Place he no Oyo). Doctoral Dissertation, 2023. (Japanese)
12. "Bullet Real-Time Physics Simulation", (2024 12/12 accessed)
13. Yamamoto, T., Terada, K., Ochiai, A., Saito, F., Asahara, Y., & Murase, K. (2019). Development of human support robot as the research platform of a domestic mobile manipulator. ROBOMECH journal, 6(1), 1-15.
14. Hori S, Yutaro I, Kiyama Y, et al. Hibikino-Musashi@Home 2017 team description paper. Preprint. 2017. Avail-able from: arXiv:1711.05457
15. T. Shiba, A. Mizutani, Y. Yano, et al,. "Hibikino-Musashi@Home 2023 Team Description Paper," arXiv:2310.12650, 2023.

Authors Introduction

Mr. Tomoya Shiba

Tomoya Shiba received his B.Eng. degree from the National Institute of Technology, Kagoshima College, Japan, in 2021, and his M.Eng. from Kyushu Institute of Technology, Japan, in 2023. He is currently a Ph.D. student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include image processing, motion planning, and domestic service robots. He has been actively demonstrating and validating his technologies through the RoboCup@Home competition and is currently conducting research on automatic dataset generation.

Dr. Hakaru Tamukoh

Hakaru Tamukoh received the B.Eng. degree from Miyazaki University, Japan, in 2001, and the M.Eng. and Ph.D. degrees from the Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a Postdoctoral Research Fellow at the Kyushu Institute of Technology, from April 2006 to September 2007. He was an Assistant Professor with the Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He is currently a Professor with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include digital hardware design, neural networks, and home service robots.

Proposal of a Grasp Verification Method Utilizing Background Subtraction and Depth Information

Ryo Terashima
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Yuga Yano
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Koshun Arimura
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hakaru Tamukoh
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Email: terashima.ryo631@mail.kyutech.jp, yano.yuuga158@mail.kyutech.jp, arimura.koshun523@mail.kyutech.jp, tamkoh@brain.kyutech.ac.jp

Abstract

Commonly grasp verification approach involves using the opening width of the robot's gripper. However, methods based on the opening width of the gripper may not apply to slender objects. In this study, we propose a grasp verification method using background subtraction. Our proposed method uses depth information to mask the background, isolating only the images of the gripper and the grasped object. Subsequently, a difference image is created by comparing the current image with the pre-grasp state, and the grasp state is detected based on the magnitude of the observed changes. The method minimizes environmental influences by masking the background, enabling highly accurate grasp verification even for complex objects. Through experiments, we validate the effectiveness of the proposed method.

Keywords: grasp verification, background subtraction, mask

1. Introduction

In recent years, the demand for robots has been increasing in various fields, including both industrial and service sectors [1], [2], [3], [4]. Grasping objects is one of the fundamental functionality required for these robots across different applications [5], [6]. Since there is a possibility of failure in object grasping by robots, grasp verification is necessary to reduce the risk of such failure. Conventional methods for grasp verification often employ the opening width of the robot's gripper. However, these approaches are limited in their ability to accurately determine successful grasps for objects with thin. In this study, we propose a background subtraction based grasp verification for thin objects. Our method uses depth to mask areas other than a robot arm and grasped object. This approach makes it possible to reduce the influence of environmental changes. Afterward, the difference between the images before and after the grasp is calculated. The grasp of success or failure is then detected based on the observed changes. In this study, we validate the proposed method through experiments and demonstrate its effectiveness.

2. Related works

2.1. *Kulkarni et al.'s Method*

Kulkarni et al.'s method [7] integrates inexpensive proximity sensors into the robot's fingers. This method enables the development of a system capable of detecting the grasping state and slippage of an object. In this method, the sensor's reference value is updated online by recording the sensor value in a non-grasping state as the baseline. The grasping state of an object is then evaluated based on this reference value. The method determines a successful grasp when the finger's deformation exceeds a certain threshold. In addition, a moving average filter is applied for noise elimination. This approach enables reliable signal processing with minimal computational load. As a result, a 100% success rate in grasp verification was achieved for 16 out of 19 objects. However, applying Kulkarni et al's method requires attaching proximity sensors to the hands of the robots, resulting in high implementation costs. Furthermore, it is also necessary to consider the impact on other grasping processes. Therefore, exploring methods to improve grasping accuracy without using proximity sensors is required. In this study, we propose a method to improve grasp verification accuracy without using proximity sensors.

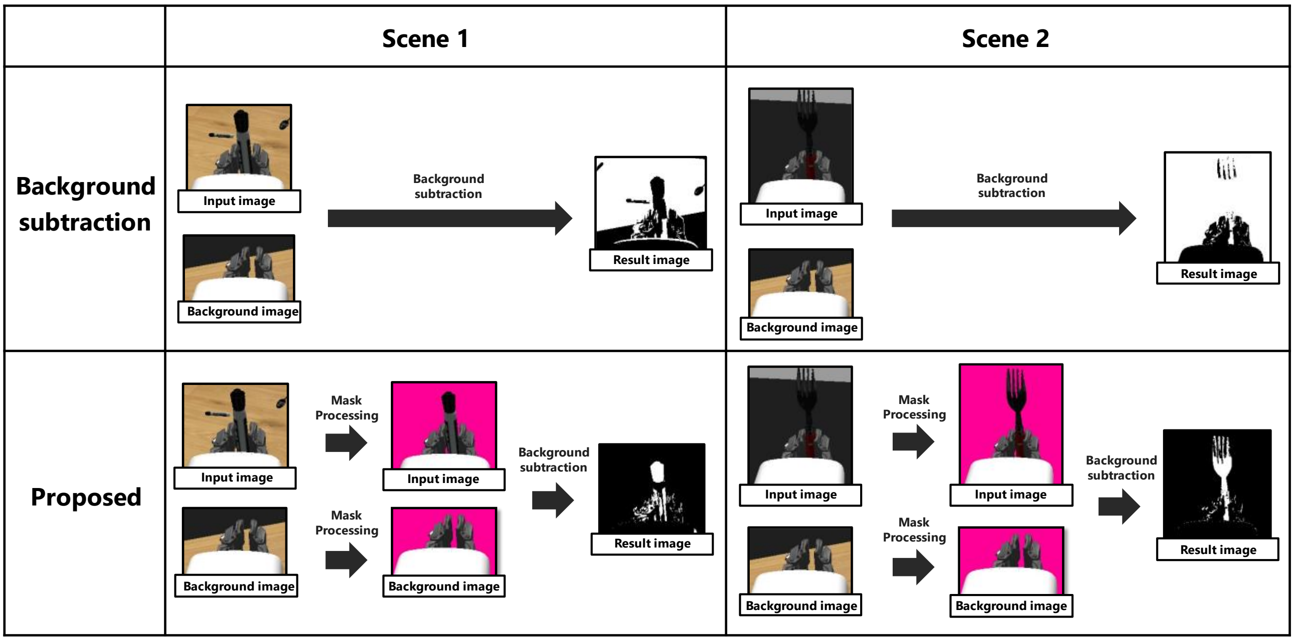


Fig. 1 Grasp verification process flow and output results of the background subtraction method and the proposed method in different scenes

2.2. CAD2GraspMonitor

CAD2GraspMonitor [8] has been proposed as a method to improve the accuracy of grasping in industrial robot picking operations. In this study, object recognition using YOLO [9] (YOLOv7) is combined with background subtraction methods. This method enables multiple quality inspection functionalities, including presence-absence confirmation of objects, type verification, pose estimation, and surplus detection. The pose estimation determines the object's position and orientation with millimeter-level accuracy, enabling the detection and correction of any misalignment in the grasp. The background subtraction method used in this study adopts a constant background and is applied to detect surplus objects. While CAD2GraspMonitor achieves high accuracy under specific conditions, it cannot handle diverse backgrounds. In this study, we propose a method that maintains grasp verification accuracy even in the presence of varied backgrounds.

3. Proposed Method

In this study, we propose a grasp verification method that utilizes background subtraction and depth information. Fig. 1 shows the processing flow of the proposed method and the back subtraction method(without masking). First, an image of the robot hand is captured using the robot's head camera before the grasping action and used as the background image. Next, a new image is captured after the robot performs the grasping action. The proposed method then applies mask processing using depth information to both the input and background images. Through mask processing, regions other than the robot arm and the grasped object are masked. This ensures reliable decision-making even in the appearance of varied backgrounds. Finally, background subtraction is applied between the background and input images to calculate the difference. As shown in fig. 1, the variation of the background causes unreliable judgement. If the difference exceeds a predetermined threshold, it is determined that the robot arm has successfully grasped the object. On the other

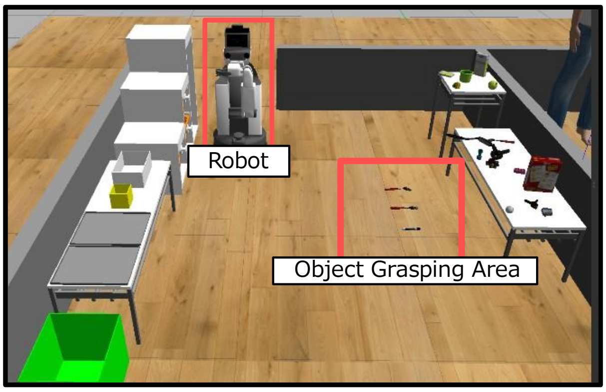


Fig. 2 Simulator environment for grasp verification experiments

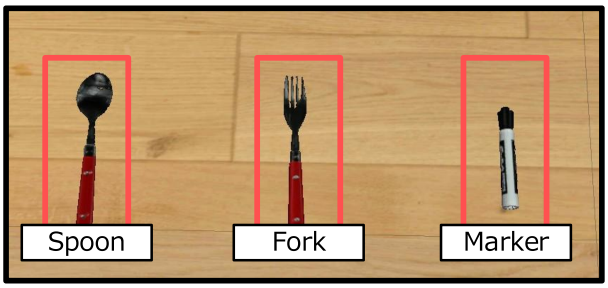


Fig. 3 Thin objects for grasping validation

Table 1 Results of grasp verification using background subtraction and depth information

	Spoon	Fork	Marker	Grasping failures detected
Background subtraction (without masking)	15/20(75%)	16/20(80%)	15/20(75%)	4/20(20%)
Proposed Method (with masking)	**19/20(95%)**	**19/20(95%)**	**18/20(90%)**	**20/20(100%)**

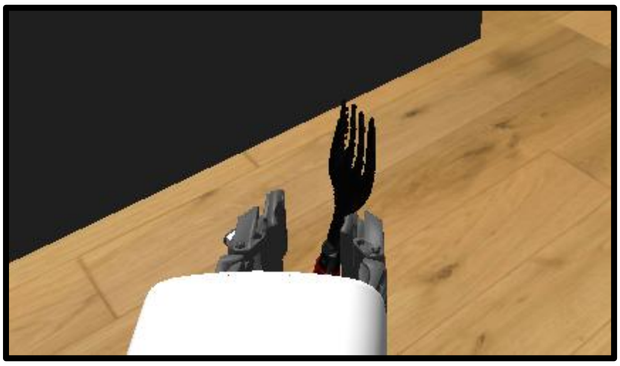


Fig. 4 Example of unexpected differences occurring in unintended areas due to an impossible grasp to achieve in the real world

hand, if the difference is below the threshold, it is judged that the grasping attempt has failed.

4. Experiments

4.1. *Setup for experiments*

Fig. 2 shows the simulator environment used for the grasp verification experiment. Fig. 3 shows a spoon, a fork, and a marker selected YCB object [10] for the experiment. In this study, 20 grasp validation were conducted for each object using the proposed method and the background subtraction method(without masking). In this experiment, a successful grasp is defined as a grasp if more than 20% of the pixels in the resulting image are output as differences. Also, we checked to see if we could detect grasping failures.

4.2. *Experimental Result*

Table 1 shows the experimental results comparing the proposed method and the background subtraction for grasp verification. The background subtraction method (without masking) shows 75% accuracy for spoons, 80% for forks, 75% for markers, and only 20% for detecting grasping failure. The proposed method shows 95% accuracy in grasp verification for the spoon and fork, 90% for the marker, and 100% for detecting grasp failure. These results represent the effectiveness of the proposed method.

5. Discussion

The proposed method was confirmed to enable accurate grasp verification for thin objects. Experimental results showed higher grasp verification accuracy than the existing method, the background subtraction (without masking). When using only the background subtraction (without masking), the accuracy was decreased due to large differences caused by background changes. Therefore, it caused the object to be mistakenly recognized as grasped, even when it was not grasped. The proposed method combines background subtraction with mask processing to reduce such misjudgments and enable more accurate grasp verification. On the other hand, the following two points can be attributed to the failure of some grasp verifications in the proposed method. The first is that there were cases in which the majority of the object was in the camera view field when the object was grasped and the difference was not sufficiently detected. Second, the unexpected part of the difference was caused by the grasping that is impossible in reality, as shown in Fig. 4. The first problem indicates the limits of the proposed method and suggests that further improvements are needed. On the other hand, the second problem is likely to be a simulator-specific factor. Therefore, this issue may be resolved through implementation in a real environment. In the future, this system should be applied to actual equipment to verify in detail the actual operating accuracy and operational constraints.

6. Conclusion

In this study, we proposed a grasping verification method using background subtraction and depth information for thin objects. Experimental results show high grasp verification accuracy for thin objects. The proposed method was also confirmed to be superior to the existing study. It exceeds Kulkarni et al.'s method in its ability to handle diverse backgrounds. However, it was confirmed that some issues occurred during grasp verification. One of the issues was caused by unrealistic grasping actions specific to simulators. To address this issue, we will implement the proposed method into a home service robot [11] in a real-world environment to further validate its efficiency.

Acknowledgments

This research is based on results from a JPNP16007 project commissioned by the New Energy and Industrial Technology Development Organization (NEDO). This research received support from JSPS KAKENHI Grant Number 23H03468 and 23K18495, as well as from JST ALCA-Next Grant Number JPMJAN23F3.

References

1. Fuji Keizai Group, 2023. Available: https://www.fuji-keizai.co.jp/report/detail.html?code=162208813, (Accessed December 12, 2024).
2. T. Ono, D. Kanaoka, T. Shiba, S. Tokuno, Y. Yano, A. Mizutani, I. Matsumoto, H. Amano and H. Tamukoh, Solution of World Robot Challenge 2020 Partner Robot Challenge (Real Space), Advanced Robotics, 2022.
3. International Federation of Robotics, Record 2.7 Million Robots Work in Factories Around the Globe, 2020. Available: https://ifr.org/ifr-press-releases/news/record-2.7-million-robots-work-in-factories-around-the-globe (Accessed December 11, 2024)
4. International Federation of Robotics, SERVICE ROBOTS Record: Sales Worldwide Up 32%, 2020, Available: https://ifr.org/ifr-press-releases/news/service-robots-record-sales-worldwide-up-32 (Accessed December 11, 2024)
5. S. Tokuno, K. Kimizuka, Y. Tanaka, Y. Usami, H. Tanaka and H. Tamukoh, Object recognition and grasping point detection using carbon nanotube - polydimethylsiloxane nanocomposite sensor, NOLTA, IEICE, Vol. 15, No. 4, pp. 883-898, 2024.
6. D. Nair, A. Pakdaman and P. G. Plöger, Performance Evaluation of Low-Cost Machine Vision Cameras for Image-Based Grasp Verification, Intelligent Robots and Systems (IROS), 2020.
7. P. Kulkarni, S. Schneider and P. G. Ploeger: Low-Cost Sensor Integration for Robust Grasping with Flexible Robotic Fingers, in International Conference on Industrial Engineering and Other Applications of Applied Intelligence and Systems, Springer, 2019, pp. 666-673.
8. M. Birem, C. Domken and A. Bey-Temsamani, CAD2GraspMonitor: Vision-based Robotic Grasp Monitoring for Industrial Kitting Application, 4th IFSA Winter Conference on Automation, Robotics & Communications for Industry 4.0 / 5.0 (ARCI'), 2024, pp. 129-134.
9. C.-Y. Wang, A. Bochkovskiy and H.-Y. M. Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, arXiv, 2022. Available: https://arxiv.org/abs/2207.02696 (Accessed December 11, 2024)
10. B. Calli, A. Singh, J. Bruce, A. Walsman, K. Konolige, S. Srinivasa, P. Abbeel and A. M. Dollar, Yale-CMU-Berkeley dataset for robotic manipulation research, The International Journal of Robotics Research, vol. 36, no. 3, 2017, pp. 261–268.
11. Y. Yano, A. Mizutani, Y. Fukuda, D. Kanaoka, T. Ono and H. Tamukoh, Unified Understanding of Environment, Task, and Human for Human-Robot Interaction in Real World Environments, The 33rd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), 2024.

Authors Introduction

Mr. Ryo Terashima

He received the B.Eng. degree from National Institute of Technology Kurume College, Japan, in 2024. He is currently a master's student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include home service robots.

Mr. Yuga Yano

He received the B.Eng. degree from Kyushu Institute of Technology, Japan, in 2022. He received the M.Eng. from Kyushu Institute of Technology, Japan, in 2024. He is currently in a Ph.D. student in the graduate school of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interest includes image processing, autonomous driving, and home service robots.

Mr. Koshun Arimura

He received his B.Eng. degree from Kyushu Institute of Technology, Japan, in 2024. He is currently a master's student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include home service robots.

Prof. Hakaru Tamukoh

Hakaru Tamukoh received the B.Eng. degree from Miyazaki University, Japan, in 2001, and the M.Eng. and Ph.D. degrees from the Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a Postdoctoral Research Fellow at the Kyushu Institute of Technology, from April 2006 to September 2007. He was an Assistant Professor with the Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He is currently a Professor with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include digital hardware design, neural networks, and home service robots.

Grasp Point Estimation using Simulator-Generated Datasets Including Pose Information

Ryoga Maruno
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Tomoya Shiba
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Naoki Yamaguchi
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hakaru Tamukoh
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Email: maruno.ryouga396@mail.kyutech.jp, shiba.tomoya627@mail.kyutech.jp,
yamaguchi.naoki892@mail.kyutech.jp, tamukoh@brain.kyutech.ac.jp

Abstract

We develop a system that automatically generates training datasets for object recognition models using a simulator. In this study, we incorporate pose information into the dataset. Using this system, we develop a method for estimating grasp points for objects that are difficult for robots to grasp, selecting a toy airplane as the target object. Three specific points are assigned to the object: the front, center, and back. In the grasp point estimation process, the center point is designated as the grasp point. The robot's arm achieves an appropriate grasp by moving perpendicularly to the line connecting the front and back points toward this grasp point.

Keywords: Object recognition, Pose estimation, Dataset generation, Grasp point

1. Introduction

In recent years, object grasping by robots has emerged as a critical application area in assistive robotics, enabling robots to perform tasks such as supporting daily activities, handling irregularly shaped objects, and aiding individuals with physical limitations. However, grasping accuracy often decreases for objects with complex shapes or unpredictable orientations, presenting significant challenges for conventional methods. Recent advancements in simulation technology provide a promising solution by enabling the automated generation of diverse and realistic training datasets under controlled conditions. This approach not only addresses the limitations of real-world data collection but also facilitates the development of robust models capable of handling a wide variety of objects.

This study proposes a method to improve grasping accuracy for challenging objects by incorporating pose information into object detection models and enabling keypoint detection. Additionally, the training dataset is automatically generated using a simulator. Building on this foundation, the study aims to develop a highly accurate and adaptable object detection model through simulation-based data generation, ultimately enhancing robotic grasping performance in real-world scenarios.

2. Related Work

2.1. Object Recognition

Using the generated dataset, the model was trained with the YOLO-based skeletal estimation model YOLO11 (parameters: 2.9M) [1]. About 100k images were used as training data, and about 25k images were used as validation data. The training was conducted with a batch size of 16 and for 60 epochs. This process established a foundation for evaluating the performance of the model based on the generated dataset.

2.2. YCB Object

For the implementation of the proposed method, YCB objects [2] were used as the target objects for validation. YCB objects are a standardized dataset widely used in research on object grasping and manipulation tasks, consisting of objects with diverse shapes, materials, and sizes. Therefore, they are well-suited for evaluating the performance of algorithms related to object recognition and robotic manipulation. In this study, representative objects were selected from the YCB dataset to perform keypoint annotation and evaluate estimation accuracy using the proposed method. Examples of YCB objects are shown in Fig. 1.

Fig. 1 Examples of YCB objects

2.3. Object Keypoint Similarity

OKS (Object Keypoint Similarity) (Eq. (1)) was adopted as the metric to evaluate the accuracy of keypoint estimation. OKS is a standard metric for measuring the accuracy of estimated keypoints on target objects, calculated based on the distance between the estimated keypoints and the ground truth data. Since the score computation involves weighting that considers the object's scale and the importance of the keypoints, it is applicable to objects of various sizes and shapes. In this study, OKS was calculated for all keypoints as well as for individual keypoints to evaluate estimation accuracy in detail. Table 1 provides the definitions of the symbols used in Eq. (1).

$$OKS = \frac{\sum_i \exp\left(-d_i^2/2s^2k_i^2\right)\delta(v_i > 0)}{\sum_i \delta(v_i > 0)} \qquad (1)$$

Table 1 the definitions of the symbols of OKS

Symbol	Meaning
d_i	Euclidean distance between the estimated keypoint and the ground truth keypoint
s	Scale of the object
k_i	Importance constant for the keypoint
$\delta(v_i > 0)$	Indicator function that equals 1 if the visibility $(v_i > 0)$

3. Proposed Method

In this study, we generate training datasets required for object detection models by performing 3D scanning of target objects and randomly placing them within a simulated environment [3]. Tobin et al. proposed *Domain Randomization (DR)* [4], demonstrating that training object detection tasks using only synthetic images can result in models capable of functioning in real robotic systems.

Additionally, Jonathan et al. [5] applied this technique to object recognition tasks, achieving performance in vehicle detection comparable to models trained on real-world images. These findings suggest that DR is expected to minimize Domain Gap effectively. The Domain Gap in this study refers to the difference between simulation data and real-world image data.

Our dataset generation process adopts DR to create diverse data by randomly altering backgrounds, object positions, and camera angles. An example of the generated data is shown in Fig. 2. During this process, annotation data, such as object position information and labels is automatically stored.

Furthermore, in this study, we added a function to generate annotation data with keypoint information for each object to improve object grasping performance. In addition, we implemented a function to determine keypoint visibility by extending rays from the simulator's camera position to each keypoint and verifying their visibility by checking for potential obstructions, making the keypoint information more detailed and useful. Fig. 3 shows an example of the visualized keypoint information. In Fig. 3, the positions and visibility of the keypoints are clearly indicated, providing a concrete understanding of the characteristics of the training dataset.

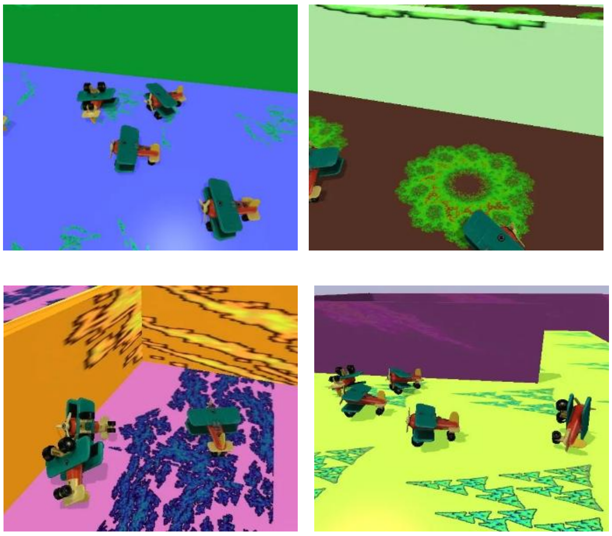


Fig. 2 The Training Datasets Generated in the Simulator

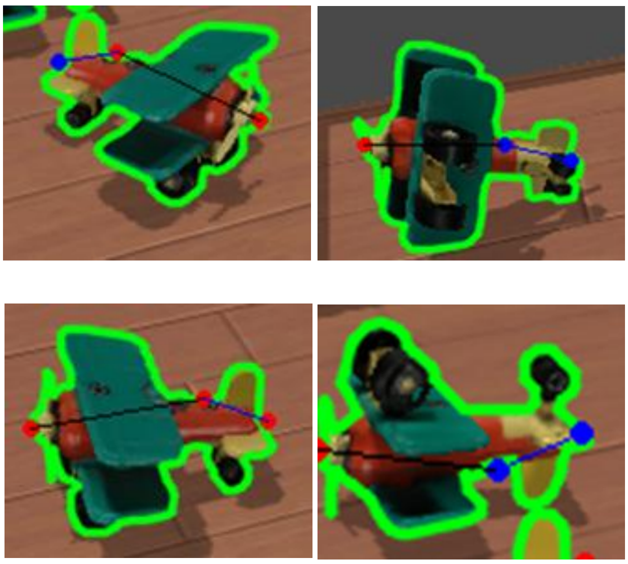


Fig. 3 Keypoint Annotations with Visibility in the Simulated Dataset (Red: Visible, Blue: Non-Visible)

4. Experiment

The purpose of this experiment is to verify the accuracy of keypoint estimation using the proposed method. The reason for selecting the toy-airplane as the target object is that it is one of the objects with low grasping success rates in conventional methods, such as Grasp Pose Detection (GPD) [6] and its improved version, PointNetGPD [7], making it suitable for evaluating the effectiveness of the proposed method. For validation, the toy-airplane was randomly placed in plausible configurations within a simulator, and 10,000 generated images were used. An example of the test data is shown in Fig. 4.

The toy-airplane is annotated with three keypoints: the front, center, and back. The center is defined as the grasping point, and the arm is assumed to be inserted perpendicularly to the vector formed by the front and back points. For evaluation, the mean OKS of all keypoints, the mean OKS of the keypoint defined as the grasping point, and the angular error between the vector formed by the front and back points and the annotated data were calculated. This evaluation design allows for a detailed verification of the effectiveness of the proposed method for objects that are difficult to grasp.

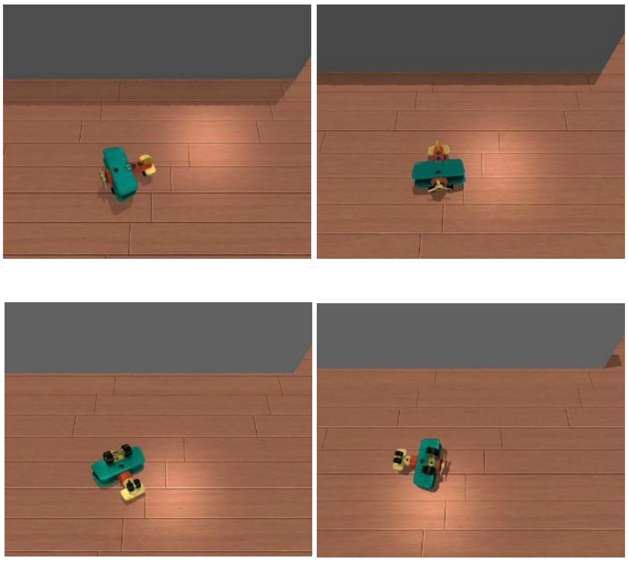


Fig. 4 The Test Datasets in the Simulator

5. Results

The results are shown in Table 2. The proposed method demonstrated excellent keypoint estimation performance for objects like the toy-airplane, which are difficult to grasp. In particular, the high OKS values and small angular errors suggest that keypoint information is effective for grasping motions. This indicates that the proposed method has the potential to optimize the insertion angle and position of the robotic arm even for objects that have been considered difficult to grasp using conventional methods. Fig. 5 shows visualized outputs of the grasping points and arm angles.

On the other hand, it is important to note that the validation data were generated within a simulator. The extent to which the proposed method can maintain its performance when faced with the diversity and uncertainty of objects in real-world environments requires further investigation. Additionally, verifying whether similar effects can be achieved for objects other than the toy-airplane is also essential.

Table 2 Keypoint Estimation Performance for Toy-Airplane Using the Proposed Method

	Result(average)
OKS(all keypoints)	0.9996
OKS(front)	0.9995
OKS(center)	0.9996
OKS(back)	0.9996
Angle error	0.0215

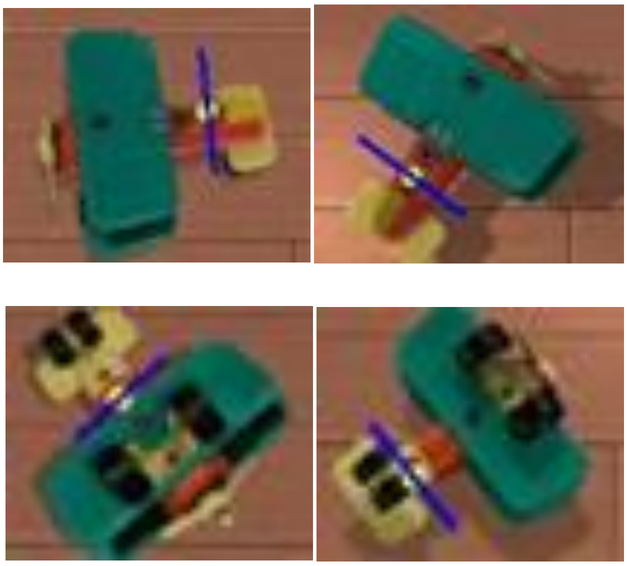


Fig. 5 Grasp Point and Insertion Angle Output

6. Conclusion

This study proposed a method to enable accurate keypoint estimation for objects that are difficult to grasp, such as the toy-airplane, by generating a training dataset with keypoint annotations using a simulator. Experimental results showed favorable outcomes in terms of the mean OKS of all keypoints, the mean OKS of the keypoint defined as the grasping point, and angular error, demonstrating that the proposed method can effectively provide the information necessary for object grasping motions. This study offers a new solution for objects that were challenging to handle using conventional methods, contributing to the realization of flexible object manipulation by robots.

As a future prospect, the first step is to validate the proposed method using real images to improve its accuracy. Additionally, it is necessary to implement this method on actual robots and evaluate its effectiveness in real-world environments through comparisons with conventional methods. Additionally, while the current approach focuses on 2D keypoints, exploring the direct acquisition of 3D keypoints or their conversion from 2D to 3D is expected to open possibilities for more advanced object manipulation [8][9].

7. References

1. Ultralytics. "Ultralytics YOLO11" 2024. Available at: https://docs.ultralytics.com/. Accessed: 2024-11-12.
2. Berk Calli, Arjun Singh, James Bruce, Aaron Walsman, Kurt Konolige, Siddhartha S. Srinivasa, Pieter Abbeel, and Aaron M. Dollar. "YCB Benchmarking Project: Object Set, Data Set and Their Applications." Journal of The Society of Instrument and Control Engineers, 56(10):792–797, 2017.
3. Tomohiro Ono. "Establishing Fundamental Technologies for Home Service Robots: Data Generation for Sim2Real and Its Application to Pick-and-Place" (Kateiyo Sabisu Robotto no tame no Kiban Gijutsu no Kakuritsu: Sim2Real wo Jitsugen suru Data Seisei to Pick-and-Place he no Oyo). Doctoral Dissertation, 2023. (Japanese)
4. Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. "Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World." In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 23–30, 2017.
5. Jonathan Tremblay, Aayush Prakash, David Acuna, Mark Brophy, Varun Jampani, Cem Anil, Thang To, Eric Cameracci, Shaad Boochoon, and Stan Birchfield. "Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 969-977, 2018.
6. Andreas Ten Pas, Marcus Gualtieri, Kate Saenko, and Robert Platt. "Grasp Pose Detection in Point Clouds." The International Journal of Robotics Research (IJRR), 36.13-14(2017): 1455-1473.
7. Hongzhuo Liang, Xiaojian Ma, Shuang Li, Michael Görner, Song Tang, Bin Fang, Fuchun Sun, and Jianwei Zhang. "PointNetGPD: Detecting Grasp Configurations from Point Sets." In IEEE International Conference on Robotics and Automation (ICRA), pages 3629-3635, 2019.
8. Ching-Hang Chen and Deva Ramanan. "3D Human Pose Estimation = 2D Pose Estimation + Matching." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 7035–7043, 2017.
9. Diogo C. Luvizon, David Picard, and Hedi Tabia. "2D/3D Pose Estimation and Action Recognition Using Multitask Deep Learning." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 5137–5146, 2018.

Authors Introduction

Mr. Ryoga Maruno

He received the B.Eng. degree from the National Institute of Technology, Kurume College, Japan, in 2024. He is a master's degree student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include dataset creation and Visualization.

Mr. Tomoya Shiba

He received the B.Eng. degree from National Institute of Technology, Kagoshima College, Japan, in 2021. He received the M.Eng. from Kyushu Institute of Technology, Japan, in 2023. He is currently in a Ph.D. student in the graduate school of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interest includes image processing, motion planning, and domestic service robots.

Mr. Naoki Yamaguchi

He received the B.Eng. degree from the National Institute of Technology, Ube College, Japan, in 2023. He is a master's degree student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include dataset creation and Visualization.

Dr. Hakaru Tamukoh

He received the B.Eng. degree from Miyazaki University, Japan, in 2001. He received the M.Eng and the Ph.D. degree from Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a postdoctoral research fellow of 21st century center of excellent program at Kyushu Institute of Technology, from April 2006 to September 2007. He was an assistant professor of Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He is currently an associate professor in the graduate school of Life Science and System Engineering, Kyushu Institute of Technology, Japan. His research interest includes hardware/software complex system, digital hardware design, neural networks, soft-computing and home service robots. He is a member of IEICE, SOFT, JNNS, IEEE, JSAI and RSJ.

A feasibility study of generative AI applications using EV-3 Robots at the Kyushu Institute of Information Sciences

Dinda Pramanta
Kyushu Institute of Information Sciences, 6-3-1, Saifu, Dazaifu, 818-0117, Japan
E-mail: pramanta@kiis.ac.jp

Hakaru Tamukoh
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology, Japan
E-mail: tamukoh@brain.kyutech.ac.jp

Abstract

The advent of generative artificial intelligence (Gen AI) using large language models (LLMs) has brought about a transformation in the fields of education and home robotics. This study examines how students at the Kyushu Institute of Information Sciences perceive and utilize ChatGPT, with a particular on their familiarity, ethical considerations, and trust (FET) concerns. To further investigate this, we developed an educational EV-3 robot powered by ChatGPT and simulate the voice command using socket connections. By combining surveys with hands-on experiments, we uncovered the strengths and limitations of both the educational and home robotics roles. Our findings highlight the importance of FET, and 88% of respondents have no issues with implementing such technologies in the future.

Keywords: EV3-Robot, FET, Gen AI, LLM, socket connections.

1. Introduction

A trend application of Generative Artificial Intelligence (GenAI) is Chat Generative Pre-Trained Transformer (GPT) [1], which adapts the LLMs from the GPT series, not only demonstrating an ability to communicate with humans, but also having an impact in academia. A study by Wayne Xin Zhao,et. al. shows a sharp increase after the release of ChatGPT: the average number of published arXiv papers containing "large language model" in the title or abstract rises from 0.40 per day to 8.58 per day [2].

In robotics, a study by Sai H. Vemprala, et al. shows that ChatGPT can be effective in solving several such tasks, while allowing users to interact with it primarily via voice commands [3]. This may indicate that certain tasks, such as home robotics, would rely heavily on such technologies, as they provide the closest natural interaction.

However, there are challenges to using ChatGPT for educational [4] and home robotics applications. We take a small step to observe in terms of familiarity, ethics and trust (FET). From an 'F', familiarity perspective, users may not have adequate understanding, leading to misuse or over-reliance on its suggestions [5]. From an 'E', ethical perspective, the potential for biased responses or unintended behavior poses risks [6], particularly when interacting with children or elderly users. Finally, trust concerns arise from the lack of transparency in ChatGPT's decision-making process, which may lead to reluctance or overconfidence in its results despite limitations [7]. Addressing these issues will require robot simulation and survey input.

Developing synchronization communication link between two node or module is necessary [8], [9]. In our previous study, we presented the Education EV3-Robot [10] for simulating a “person” detection by using the client-server communication model [11]. In this study, we present the use of application ChatGPT as a EV3-Role and designing where the local-PC can directly access so that the EV3-Robot can be driven by voice command. We extend by conducting a survey to assess the potential applications of these technologies in the areas of education and robotics research in the students of Kyushu Institute of Information Sciences (KIIS). We evaluate the strengths and limitations of both the educational and home robotics roles of the perception and use of ChatGPT, with a particular focus on familiarity, ethical considerations, and trust (FET) concerns.

2. Methodology

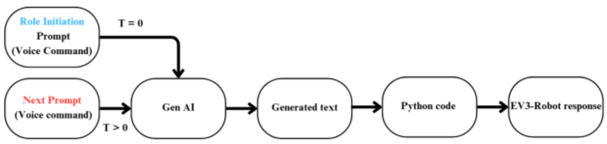


Fig. 1. Voice Command-Driven EV3-Robot Using Generative AI.

2.1. *ChatGPTs as EV3-Robot role*

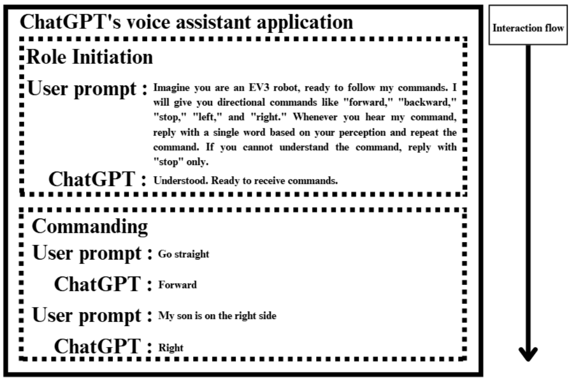


Fig. 2.Interaction ChatGPT as an EV3-Robot role.

In Fig. 1 shows how the voice commands utilized to interact with an EV3-Robot through a Gen AI model of ChatGPT voice assistance application [12]. The interaction commences with Role Initiation (T = 0), wherein the user provides an initial prompt via a voice command, thereby defining the role and task to be performed. Subsequent to the initiation of the role, the process transitions to Commanding (T > 0), wherein the user continues to issue voice commands to direct the tasks, as illustrated in Fig. 2. These commands are processed by ChatGPT, which interprets the user's input and generates an appropriate response. The generated text by ChatGPT serves as the foundation for creating specific instructions or code. This text is then translated into Python Code that is executable and tailored for the task. Finally, the Python Code is executed on the EV3 robot, resulting in the desired robotic action based on the original commands.

2.2. *System architecture*

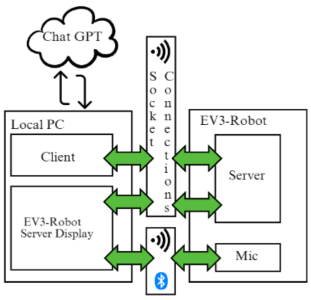


Fig. 3. System architecture.

Fig. 3 shows a system architecture for communication between a local-PC running a client and an EV3-Robot functioning as a server, with support from ChatGPT for generating responses. The system employs socket connections for communication between components. The local-PC facilitates the transmission and reception of data to and from the EV3-Robot via socket connections. The EV3-Robot, operating as a server, processes incoming data and may provide outputs, such as movement instructions or responses through its microphone wirelessly using Bluetooth.

2.3. Socket Connection for Voice-Controlled EV3-Robot

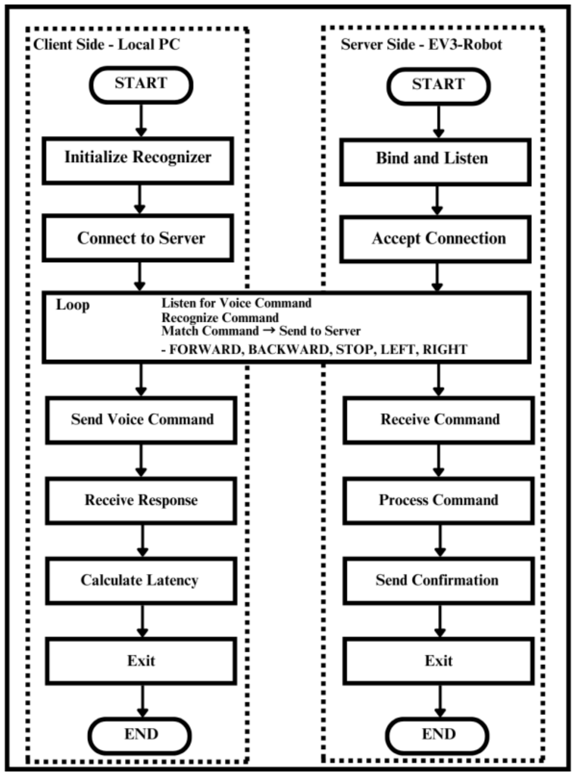


Fig. 4. Client-Server Socket Connection for Voice-Controlled EV3-Robot

In Fig. 4 shows a flow of a socket-connection communication system between a local PC (client) and an EV3-Robot (server) [13]. The system enables the control of the robot using voice commands through a client-server model.

The client system begins by initializing the recognizer, which captures and interprets voice commands, before establishing a socket connection with the server running on the EV3-Robot. It then continuously listens for voice inputs, recognizes commands such as "FORWARD," "BACKWARD," "STOP," "LEFT," or "RIGHT," and matches the recognized input. Once a command is identified, it is sent to the server via the socket connection, and the client waits for a response or confirmation that the command has been processed. The system also calculates the latency between sending the command and receiving the server's response to monitor performance. Finally, the client terminates the process after completing the necessary actions.

The server on the EV3-Robot begins by initializing, binding to a specified port, and listening for incoming client connections. Upon receiving a connection request, the server accepts it to establish communication with the client. It then processes the commands sent by the client, executing the corresponding actions on the EV3-Robot, such as moving forward or turning left. After successfully executing a command, the server sends a confirmation back to the client. Once all tasks are completed, the server terminates its process.

2.4. *FET survey queries at KIIS students*

As part of our demonstration of proposal technology to KIIS students, we aim to gather feedback on their experiences and attitudes towards ChatGPT, an advanced AI language model. The evaluation focuses on three key dimensions: Familiarity, Ethics, and Trust (FET), with a 5 questions per category with score of agreement between 1 (disagree) to 5 (fully agree), and 1 additional "yes or no" question for overall evaluation.

For the Familiarity: 1.) Do you feel fully-utilizing ChatGPT according to your needs?; 2.) Do you feel the interaction with ChatGPT naturally?; 3.) Are you aware with the information provided by ChatGPT?; 4.) You didn't encountered any instances where ChatGPT's responses were unclear or difficult to understand?; 5.) Do you think ChatGPT understood your inputs and intentions?

For the Ethics: 1.) Do you agree ChatGPT's responses avoid including personal information such as race, gender, religion, sexual orientation, or physical/mental disabilities?; 2.) Do you find the responses provided by ChatGPT to be appropriate?; 3.) Did you experience the presence of bias, such as racism, sexism, or other forms of abusive behavior, in ChatGPT's interactions?; 4.) If you find ChatGPT's responses unpleasant, would you prefer having alternative options (e.g., rephrased answers, additional information)?; 5.) No ethical concerns that arose during your interaction with ChatGPT?

For the Trust part: 1.) Do you trust the accuracy of the information provided by ChatGPT?; 2.) Do you find ChatGPT's responses to be consistent across interactions?; 3.) Does the information provided by ChatGPT align with your expectations and needs?; 4.) Do you think the sources and references of information provided by ChatGPT are clearly indicated and verifiable?; 5.) Do the information provided by ChatGPT matches what you were looking for?

We have 42 students as a respondent to score each question and provide qualitative feedback to help us better understand the feasibility of ChatGPT technology to be adopted in EV3-Robot based on their experiences.

3. Results and Discussion

3.1 Familiarity

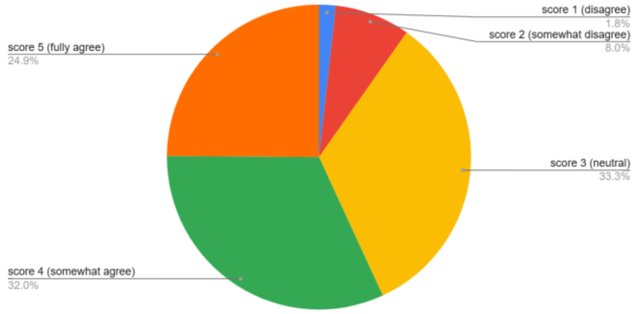


Fig. 5. Familiarity score results.

Fig. 5 shows the familiarity score results, which were calculated as the average of the total scores for each question. The majority of respondents gave a score of 3 (neutral), closely followed by a score of 4 (somewhat agree), with the two being highly competitive.

3.2 Ethics

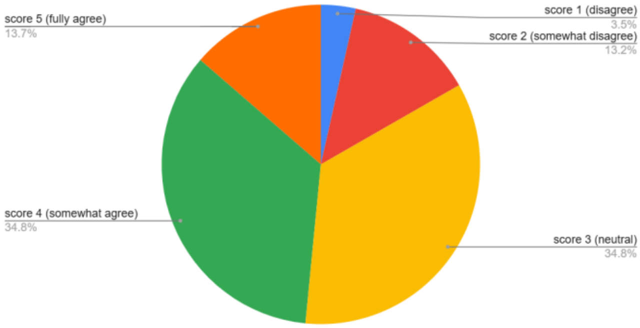


Fig. 6. Ethic score results.

Fig. 6 shows the Ethics score results, which were calculated as the average of the total scores for each question. The majority of respondents gave a score of 3 (neutral) and 4 (somewhat agree) equally.

3.3 Trust

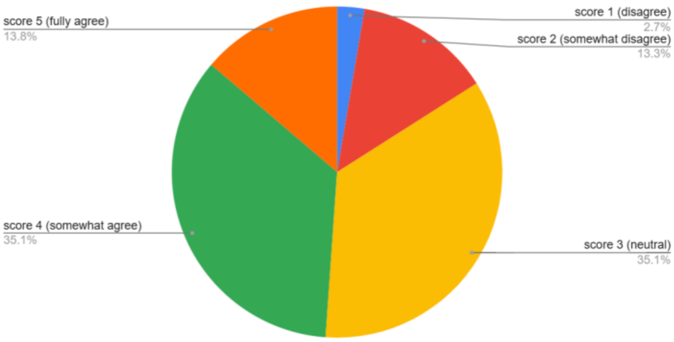


Fig. 7. Trust score results.

Fig. 7 shows the Ethics score results, which were calculated as the average of the total scores for each question. The majority of respondents gave a score of 3 (neutral) and 4 (somewhat agree) equally.

3.4 Overall Evaluation

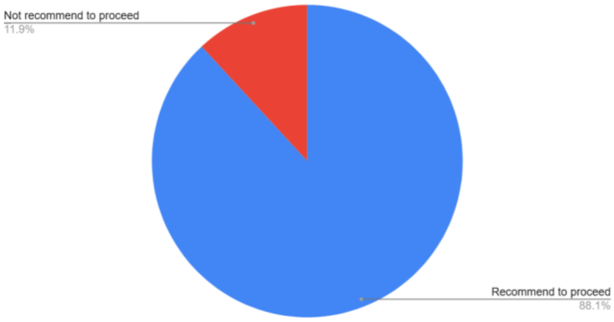


Fig. 8. Overall results.

Fig. 8 shows that nearly 88.1% of the 42 respondents recommend proceeding with the technology utilizing ChatGPT.

4. Conclusion

We successfully implemented ChatGPT in the role of an EV3 robot and developed a voice-controlled socket connection for the EV3 robot. Additionally, we studied the feasibility of generative AI applications, such as ChatGPT, in EV3 robots at the Kyushu Institute of Information Sciences. From the FET perspective, respondents tended to be "neutral" or "somewhat agree," with nearly 90% of the 42 participants recommending proceeding with the utilization of ChatGPT technology for the EV3 robot.

Acknowledgements

This research was supported by JSPS KAKENHI Grand Numbers 23H03468.

References

1. G. Yenduri et al., "GPT (Generative Pre-Trained Transformer)— A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions," in IEEE Access, vol. 12, pp. 54608-54649, 2024.
2. Zhao, Wayne Xin, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min et al. "A survey of large language models." arXiv preprint arXiv:2303.18223 (2023).
3. Padmanabha, Akhil, et al. "Voicepilot: Harnessing LLMs as speech interfaces for physically assistive robots." Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology. 2024.
4. Yu, Hao. "Reflection on whether Chat GPT should be banned by academia from the perspective of education and teaching." Frontiers in Psychology 14 (2023): 1181712.
5. Roganović, Jelena. "Familiarity with ChatGPT Features Modifies Expectations and Learning Outcomes of Dental Students." International Dental Journal (2024).
6. Stahl, Bernd Carsten, and Damian Eke. "The ethics of ChatGPT–Exploring the ethical issues of an emerging technology." International Journal of Information Management 74 (2024): 102700.
7. Choudhury, Avishek, and Hamid Shamszare. "Investigating the impact of user trust on the adoption and use of ChatGPT: survey analysis." Journal of Medical Internet Research 25 (2023): e47184.
8. Pramanta, D., Morie, T. and Tamukoh, H., 2017. Synchronization of Pulse-Coupled Phase Oscillators over Multi-FPGA Communication Links. J. Robotics Netw. Artif. Life, 4(1), pp.91-96.
9. Pramanta, D. and Tamukoh, H., 2019, December. High-speed synchronization of pulse-coupled phase oscillators on multi-FPGA. In International Conference on Neural Information Processing (pp. 318-329). Cham: Springer International Publishing.
10. Pramanta, Dinda, et al. "YOLO real-time object detection on EV3-Robot using FPGA hardware Accelerator." (2024).
11. Pramanta, Dinda, et al. "Client-server Communication using YOLO for Object Detection Results between PC and EV3-Robot." 九州情報大学研究論集= Bulletin Kyushu Institute of Information Sciences/研究論集編集委員会 編 26 (2024): 15-24.
12. ChatGPT. "Voice Assistant Pro." Accessed January 6, 2025. Available: https://chatgpt.com/g/g-Wv5YoTNKq-voice-assistant-pro.
13. R. Hempel and D. Lechner, “ev3dev,” 2023. Available online: https://www.ev3dev.org/, accessed on 30 January 2023. [Online]. Available: https://www.ev3dev.org/

Authors Introduction

Dr. Dinda Pramanta

He earned his Bachelor's degree from Telkom University in 2013 and completed both Master's and Doctoral degrees at the Kyushu Institute of Technology in 2016 and 2020. After a post-doctoral fellowship with NEDO, Japan, from 2020 to 2021, he is currently an Assistant Professor at Kyushu Institute of Information Sciences since 2021. His research interests include neural networks, hardware, and AI for educational applications.

Dr. Hakaru Tamukoh

A member of IEEE, he obtained his B.Eng. degree from Miyazaki University in 2001 and completed his M.Eng. and Ph.D. degrees at the Kyushu Institute of Technology in 2003 and 2006, respectively. After serving as a Postdoctoral Research Fellow, he became an Assistant Professor at the Tokyo University of Agriculture and Technology from 2007 to 2013. Currently a Professor at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, his research interests encompass digital hardware design, soft computing, and home-service robots.

Development of a Collaborative System Between A Drone and A Home Service Robot for Enhanced Operational Efficiency

Haruki Miura
Department of Creative Engineering, NIT, Kitakyushu College,
5-20-1 Shii, Kokuraminamiku, Kitakyushu, Fukuoka, 802-0985, Japan

Rion Yohu
Department of Creative Engineering, NIT, Kitakyushu College,
5-20-1 Shii, Kokuraminamiku, Kitakyushu, Fukuoka, 802-0985, Japan

Yuma Yoshimoto
Department of Creative Engineering, NIT, Kitakyushu College,
5-20-1 Shii, Kokuraminamiku, Kitakyushu, Fukuoka, 802-0985, Japan
E-mail: k21164hm@apps.kct.ac.jp, k21189ry@apps.kct.ac.jp, yoshimoto@kct.ac.jp
https://yoshimoto.apps.kct.ac.jp/

Abstract

This research proposes a system that coordinates a home service robot with a drone to improve task efficiency. As an experiment, we conduct a search and pick-up task that integrates the home service robot and the drone. The drone's front camera and YOLOv8 are used to detect objects and send information to the robots.

Keywords: Drones, Home service robots, YOLOv8

1. Introduction

Home service robots are robot that operate in human living spaces. Such as households and stores to reduce human workload. Home service robots are equipped with RGB-D cameras, LiDAR sensors, robotic arms, and a mobile vehicle which enable them to perform tasks such as housework. Although the performances of home service robots are improving daily, however challenges remain when deploying them in home environments. These problems include the inability to detect distant or dead space objects in large rooms, as well as the failure to recognize individuals behind the robot, potentially resulting in collisions.

Drones are small flying robots. In this research, we focus on drones capable of unmanned flight and environmental recognition. Drones can move quickly and, when mounted with cameras, can watch wide areas at high speed. Compared to home service robots, drones can rapidly gather information and cover blind spots with ease. Moreover, by flying at heights above human stature, drones can reduce the risk of collisions with people. While highly useful for home operations, drones lack the power of home service robots, making tasks such as accurately lifting objects extremely difficult.

Fig. 1 shows how the drone acquires information about objects in the room, while Fig. 2 shows the method used by the home service robot. In this research, we use Toyota's Human Support Robot (HSR) as the home service robot. As shown in Fig. 1, the drone can get information about objects in the room by flying above them and capturing images. In contrast, home service robots must rotate 360° to gather information about its surroundings. Thus, drones are more efficient in get information about the environment.

Furthermore, compared to surveillance cameras, drones have the advantage of mobility, which eliminates blind spots commonly associated with stationary surveillance cameras. Surveillance cameras incur various costs, including the initial installation of comprehensive camera systems, maintenance, and upgrades [1]. In contrast, drones require significantly fewer units due to their mobility, resulting in lower costs for installation, maintenance, and system upgrades. Additionally, installing surveillance cameras raises complex concerns regarding privacy and individual autonomy. Even in public spaces, the awareness of being constantly monitored can instill feelings of vulnerability and self-censorship, which, if extended to the home, could have detrimental psychological effects [2]. However, drones are easily movable, allowing their operation to be quickly stopped or relocated if they cause psychological discomfort. This flexibility helps reduce the mental burden associated with their use.

This research proposes a system in which the home service robot and the drone collaborate to accomplish tasks more efficiently.

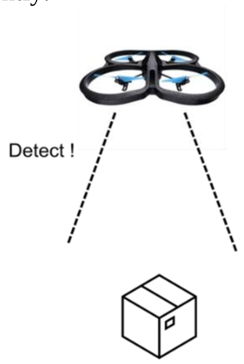


Fig. 1. Method for the Drone to Get Information About Objects

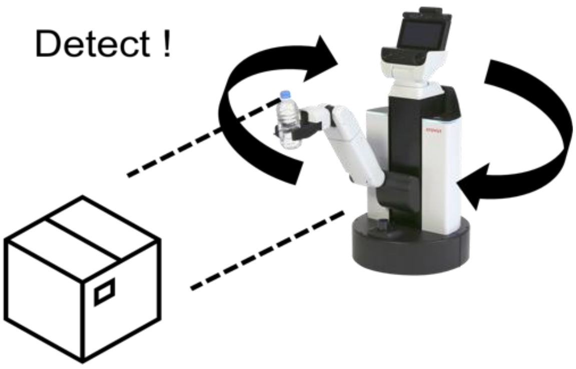


Fig. 2. Method for the Home Service Robots to Get Information About Objects

2. Pre-Experiment

Comparison of movement time between drone and home service robot.

2.1. *Experiment Methods*

As shown in Fig. 3, both the drone and the home service robot are moved along the bold arrows. And We measured the time of movement within the room.

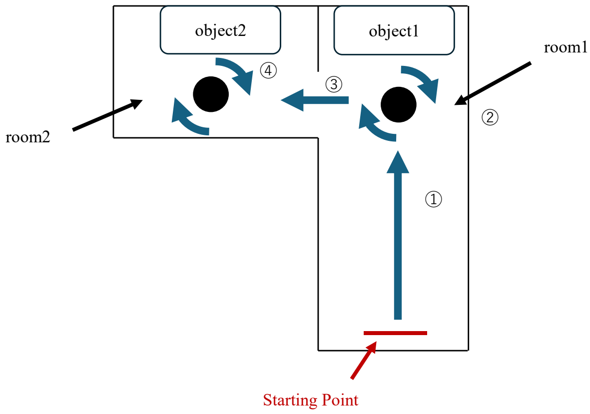


Fig. 3. Environment in Which the Experiment is Performed

2.2. *Results*

The movement time of the drone and the home service robot was measured three times. The results are shown in Table 1. The drone moves more rapidly than the home service robot.

Table1 Results

Times	1	2	3	Average
Drone	00:32.9[s]	00:26.8[s]	00:28.0[s]	00:29.2[s]
HSR	01:07.7[s]	00:54.7[s]	00:52.6[s]	00:58.3[s]

2.3. *Consideration*

As shown in Table 1, the drone was able to move at a significantly higher speed than robots. A significant difference was observed even with just two rooms. There was a time difference of about 30 seconds. Furthermore, it is expected that this difference will increase further when comparing the two in environments with more rooms or larger areas.

3. Related Work

3.1. *Drone*

Fig. 4 one of drones. Drones, which are capable of flight and three-dimensional movement, are used for applications such as image and video capture, as well as pesticide spraying.

Drones can be classified into outdoor drones and indoor drones depending on their operational environment. Tasks expected indoor drones, which are mainly used indoors, include inspection and surveying of buildings, high-speed transportation and assembly of structures, dismantling work at heights on construction sites, and deployment in factory lines utilizing three-dimensional space effectively [3]. Drones are capable of high-speed movement and, since they fly, they do not need to avoid obstacles on the floor. Additionally, by flying at heights above human stature, they can reduce the risk of collisions with people. These features make drones highly compatible with environments such as households, where humans are active, and spaces with complex furniture arrangements.

Fig. 4. Drone

3.2. *Tidy-Up Tasks Using the Home Service Robot*

Tidying up does not only make our lives comfortable but also causes changes in our lifestyle [4].

Tidy up, which involves placing objects in designated locations, contributes to creating a comfortable environment and improving productivity. Furthermore, tidy up consists of two main tasks: an identification task, where the robot identifies the objects to be tidied up within the environment, and a transportation task, where the identified objects are carried to their designated locations.

3.3. *You Only Look Once (YOLO)*

You Only Look Once (YOLO) is a real-time object detection method that achieves high detection accuracy while operating at extremely high speed [5]. The results of object detection using YOLOv8 are shown in Fig. 5. In this research, YOLOv8, among the YOLO 5series, is utilized.

Fig. 5. Object Detection Results by YOLOv8

4. Proposal

We propose a system that connects the robot and the drone, as shown in Fig. 6. The robot and the robot control PC mounted and connected with an Ethernet cable, while the drone and the drone control PC connect with a WiFi. By connecting through both control PC, the robot and the drone are connected. The identification task is handled by the drone, while the transportation task is performed by the robot, enabling efficient progress. We constructed the system as shown in Fig. 6.

The operational flow for performing a tidy-up task using the system is as follows.

(1) The drone transmits images

(2) Perform object detection on the captured images.

(3) Transmit the results to the robot control PC using serial communication.

(4) Operate the home service robot from the robot control PC.

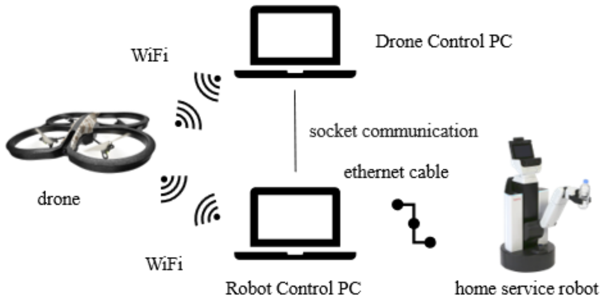


Fig. 6. Propose system

5. Experiment

5.1. *System Operational Experiment*

The drone control PC applies YOLOv8 to the images captured by the drone for object detection. In this experiment, the detection of a mug was conducted. When the drone control PC detects the mug, it sends the message "detect" to the robot control PC via serial communication. Subsequently, the robot control PC commands the robot to move forward.

To verify the operation of the integrated system, the mug was placed in front of the drone, and the movement of the robot was observed, confirming the collaboration between the drone and the robot.

5.1.1. Experiment Methods

As shown in Fig. 7, we confirmed whether the system could detect the mug in the images obtained from the drone and subsequently control the robot's movements accordingly.

Fig. 7. Detect mug

5.1.2. Result

As shown on Fig. 8, Fig. 9, the robot was able to move forward based on the detection results from the drone. This confirms the successful development of the system that integrates the drone and the robot.

Fig. 8. The State Before Movement

Fig. 9. The State After Movement

5.2. *Object Detection with a Drone Camera and Measurement of Latency*

We confirm whether object detection using YOLOv8 is possible through the drone's camera and evaluate the latency.

5.2.1. Experimental Method

The drone and the drone control PC were connected to verify the effectiveness of the proposed algorithm. The experimental setup is shown in Fig. 10. In this experiment, the drone and the drone control PC were placed on a desk and communicated via WiFi. Additionally, the latency was measured using the following method:

(1) A stopwatch was used to display the time.
(2) The drone's camera captured the stopwatch and sent the image to the control PC.
(3) The control PC used YOLOv8 to detect humans and displayed the detection result image.

By observing the stopwatch at each step and comparing the recorded time differences, the latency was measured.

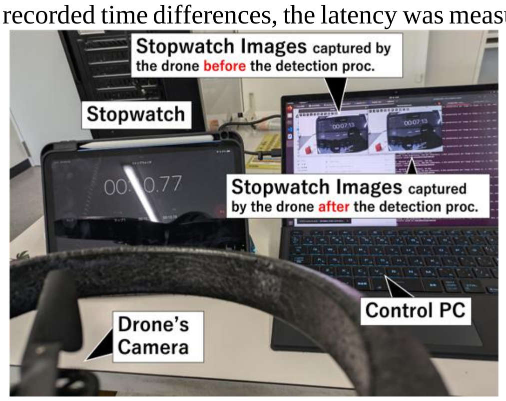


Fig. 10. Experimental Environment

5.2.2. Result

The experimental results are shown in Table 2. From these results, it was confirmed that object detection is possible. Additionally, it was observed that there is a latency of about 3 seconds from the drone capturing an image to sending it to the PC.

Table2 Latency Measurement Results

	The time on the stopwatch	The time displayed on the PC screen	The time after applying YOLOv8
Displayed Time [s]	10.77	7.13	7.13
Delay [s]	N/A	3.64	3.64

6. Discussion

As shown in Table 2, the latency from the time the drone captured an image to its display on the PC screen was about 3 seconds. In contrast, the latency when YOLOv8 was applied—measured from the time the drone-captured image was displayed on the screen—was less than 0.01 seconds. These results indicate a significant information gap between the drone and the PC, which is likely the most significant obstacle when integrating drones with robots in home environments. Furthermore, YOLOv8 demonstrates extremely high processing speed, making it a highly suitable library for real-time applications.

7. Conclusion

In this research, we developed a system that integrates drones with home service robots and demonstrated that drones have advantages over home service robots in certain aspects. However, safety concerns remain regarding the use of drones in home environments. Moving forward, we will focus on developing tasks to enable the safe and effective integration of drones with home service robots for operation in household settings.

References

1. "Advantages and Disadvantages of Surveillance Cameras in Critical Infrastructure Protection,", https://senstar.com/senstarpedia/advantages-and-disadvantages-of-surveillance-cameras-cip/, 2024/12/15 accessed.
2. Adwait S Malik, Sourya Acharya, Sonal Humane, "Exploring the Impact of Security Technologies on Mental Health: A Comprehensive Review," Cureus 16(2): e53664, doi:10.7759/cureus.53664.
3. Hiroyuki Ukida, "Indoor Unmanned Aerial Vehicle Navigation System Using LED Panels and QR Codes," Journal of Robotics and Mechatronics, vol33,No2,pp. 242-253, 2021.
4. Akihiro Ogasawara and Manabu Gouko, "Determining the Most Effective Way of Ensuring a Tidying-Up Behavior: Comparison of Effects of Reminders Using Oral Instruction, Posters, and Robots," Journal of Advanced Computational Intelligence and Intelligent Informatics, 2020, vol24,No4, pp. 543-548.

5. " Ultralytics YOLOv8,", https://docs.ultralytics.com/ja/models/yolov8/, 2024/12/15 accessed.

Authors Introduction

Mr. Haruki Miura

He enrolled at National Institute of Technology (KOSEN), Kitakyushu College, Japan, in 2021. In 2023, he pursued the Information Systems course, focusing on algorithms and control. He commenced his research in robotics in 2024.

Mr. Rion Yohu

He enrolled at National Institute of Technology (KOSEN), Kitakyushu College, Japan, in 2021. In 2023, he pursued the Information Systems course, focusing on algorithms and control. He commenced his research in robotics in 2023.

Dr. Yuma Yoshimoto

He received his B.Eng. degree from National Institute of Technology (KOSEN), Maizuru College, Japan, in 2016. He received his M.Eng. and D. Eng. degrees from Kyushu Institute of Technology, Japan, in 2018 and 2021, respectively. And he was JSPS researcher, in 2019 - 2021. He was a postdoctoral researcher at the Kyushu Institute of Technology, Japan in 2021-2022. Currently, he is assistant professor at the National Institute of Technology (KOSEN), Kitakyushu College, Japan. His research interests include deep learning, robot vision and digital hardware design. He is a member of IEICE, IEEE.

Efficient Object Detection with Color-Based Point Prompts for Densely Packed Scenarios in WRS FCSC 2024

Naoki Yamaguchi
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan

Tomoya Shiba
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan

Hakaru Tamukoh
Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan
E-mail: yamaguchi.naoki892@mail.kyutech.jp
https://www.lsse.kyutech.ac.jp/

Abstract

This paper introduces a novel object detection method designed for densely packed environments, such as those encountered in the World Robot Summit Future Convenience Store Contest (FCSC) 2024. Our system leverages color-based point prompts in conjunction with Segment Anything (SAM) 2 to achieve precise object segmentation and grasp point estimation, specifically targeting scenarios where objects like rice ball cluster tightly within containers. Unlike traditional methods that depend on pre-defined grasping strategies susceptible to positional drift in mobile robots, our approach dynamically identifies and isolates objects without requiring extensive retraining inherent to CNN or Transformer models. We conducted comprehensive experiments comparing our method against SAM 2 and Grounding DINO using a dataset of 10 test images containing 202 rice balls. Additionally, we deployed the system on a Toyota Human Support Robot during the FCSC Stock Task to assess real-world performance metrics. Results demonstrate that our method achieves higher detection accuracy and operational efficiency, validating its potential for autonomous retail applications.

Keywords: Object Detection, Human Support Robot, Future Convenience Store Contest

1. Introduction

Automating retail environments, especially convenience stores, demands advanced robotic systems that accurately detect, identify, and manipulate a diverse range of products in compact and cluttered settings [1]. The World Robot Summit (WRS) Future Convenience Store Contest (FCSC) exemplifies this need by presenting tasks that mimic real-world retail scenarios. One such task is the Stock Task, which involves handling densely packed rice balls within containers [2]. A major challenge in these tasks is grasping tightly clustered and visually similar objects. Traditional robotic grasping strategies often rely on fixed grasp points, which become unreliable for mobile robots due to inevitable discrepancies between planned and actual positions. This positional drift reduces the effectiveness of static grasping approaches, highlighting the need for more dynamic and adaptable object detection and segmentation methods.

Convolutional Neural Networks (CNNs) and Transformer-based architecture offer high accuracy but demand extensive training datasets and computational resources to adapt to new objects. This retraining process increases preparation costs and limits system flexibility in dynamic environments where product assortments frequently change [3].

Prompt-based segmentation models, such as the Segment Anything Model (SAM) [4] and its successor SAM2 [5], present promising alternatives. These models use prompts—points, boxes, or language instructions—to guide segmentation without exhaustive retraining. Building on this paradigm, we propose a color-based point prompt methodology integrated with SAM2 to achieve efficient and robust object detection in densely packed scenarios. By leveraging color cues, our system dynamically generates prompts that facilitate the accurate isolation and identification of visually similar objects, enhancing the robot's ability to perform reliable grasping tasks.

This paper details the design and implementation of our proposed method, evaluates its performance through controlled experiments and real-world deployments, and discusses its implications for future autonomous retail systems.

2. Related Work

2.1. *Object Detection in Dense Environments*

Detecting objects in densely packed environments presents significant challenges due to occlusions, similar appearances, and spatial constraints. Traditional object detection frameworks, including CNN-based models like YOLO [6] and Transformer-based architectures, have achieved substantial progress in accuracy and speed. However, these models typically require extensive training on large, object-specific datasets, limiting their adaptability in dynamic retail settings where product assortments frequently change.

2.2. *Prompt-Based Segmentation and Grounding DINO*

Prompt-based segmentation methods have gained traction for their ability to guide segmentation models with minimal input prompts. The Segment Anything (SAM) and its advanced variant SAM2 exemplify this approach by enabling the segmentation of objects based on various prompt types, including points, bounding boxes, and textual descriptions. These models leverage extensive pre-training on diverse datasets, allowing them to generalize across various objects and environments without retraining. While language-based prompts, as utilized in Grounding DINO [7], offer flexible object specification, they may need more precision in environments with visually similar objects. Our approach focuses on color-based point prompts to enhance segmentation accuracy in challenging scenarios.

2.3. *Grasping Strategies for Mobile Robots*

Robust perception systems enable effective grasping in dynamic environments by accurately identifying and localizing objects for manipulation. Fixed grasping strategies are insufficient for mobile robots due to potential positional discrepancies and variations in object arrangements. Recent advancements emphasize integrating real-time object detection with adaptive grasp point estimation to ensure reliable manipulation in cluttered environments.

3. Proposed Method

Our method targets the segmentation of tightly packed objects, such as rice balls, using a combination of human-guided color range extraction and automated processing steps integrated with SAM2. Fig. 1 shows a process flow of proposed method. The workflow is as follows:

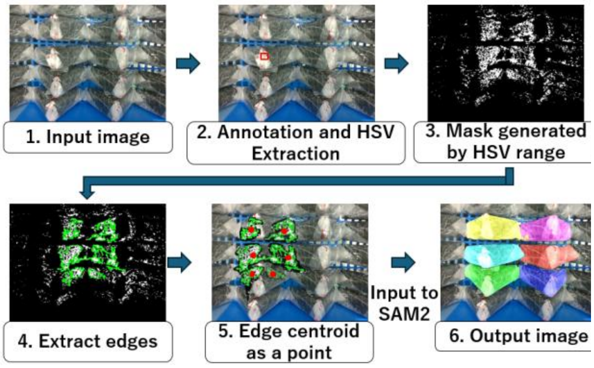


Fig. 1 Process flow of proposed method

3.1. *Overview*

The proposed method captures real-world images, extracts HSV color ranges from annotated regions, and generates precise point prompts based on edges detected from binary masks. These prompts guide SAM2 for accurate segmentation. This process minimizes the need for extensive training or pre-existing object models, making it adaptable for dynamic retail environments like the WRS FCSC Stock Task.

3.2. *Detailed Steps*

i. **Capturing Input Images**
Capture images of the target environment using a standard RGB camera.

ii. **Human annotation and HSV Extraction**
A human operator manually annotates a portion of the target object by drawing bounding boxes around regions of interest. Extract the HSV color range characteristic of the object from these annotated areas.

iii. **Mask Creation**
Generate a binary mask using the extracted HSV ranges to isolate regions matching the target object's color.

iv. **Edge Detection and Filtering**
Detect edges from the binary mask and filter out small regions below a predefined area threshold to reduce noise.

v. **Point Prompt Generation**
Calculate the centroids of valid edge regions and use them as point prompts to indicate object locations.

vi. **Segmentation Using SAM2**
Feed the input image and generated point prompts into SAM2 to produce detailed segmentation masks for each detected object.

Table. 1 Result of the Stock sub task

Team	Happy Robot	TAK	TMU Mecha	NaRIPa	HAR Chuo	Team Meijo	**HMA (Our Team)**	eR@sers	Optimator	HSRL -CoR
Score	59	59	55	40	16.5	14	7	5	5	0

Table. 2 Comparison of Zero-Shot Object

Image ID	**Proposed Method**	**Grounding DINO**	**Ground Truth**
1	2	2	3
2	5	3	5
3	3	3	5
4	4	5	15
5	6	3	15
6	6	3	15
7	6	5	16
8	8	6	20
9	46	1	54
10	47	1	54
ALL	133	32	202

(a) Grounding DINO (b) Proposed method

Fig. 2 Detection Result

4. Experiments

This section outlines the experimental setups used to evaluate the proposed method. We conducted two primary experiments: implementing the WRS FCSC method and assessing its zero-shot object detection performance against Grounding DINO.

4.1. *Evaluation in the WRS FCSC Stock Task*

To test the method's real-world applicability, we integrated it into the Human Support Robot developed by TOYOTA MOTOR CORPORATION [8] and participated in the WRS FCSC Stock Sub Task.

Stock Task Overview:
The WRS FCSC aims to advance technologies for automating convenience store operations, including product shelving and waste collection. The Stock Sub Task measures how quickly and accurately participants arrange pre-packaged items on shelves.

- **Items:** Arrange 54 rice balls (three types, 18 each) into designated shelf positions (e.g.,Fig. 2).
- **Scoring:** Correctly placing each rice ball earns *1* point. Using a standard container for transportation awards an additional *5* bonus points.

4.2. *Zero-Shot Object Detection*

We evaluated the proposed method's zero-shot object detection capabilities using a specialized dataset and compared its performance with Grounding DINO.

Experimental Setup:

- **Dataset:** Created a dataset of 10 test images featuring 202 densely packed rice balls to simulate real-world retail environments.
- **Models Compared:**
  - **Proposed Method:** Uses color-based point prompts with SAM2 for segmentation.
  - **Grounding DINO:** Employs textual prompts (e.g., "a rice ball") for object detection.
- **Evaluation Metrics:** Measured detection accuracy by the number of correctly identified rice balls compared to the ground truth.

5. Experimental Results

5.1. *WRS FCSC Stock Task*

We evaluated the proposed method using the HSR in the WRS FCSC Stock Sub Task. Our team, HMA, scored *7* points, as illustrated in Table. 1 Specifically, we earned *5* points by utilizing a standard container for item transportation and *2* points by accurately placing items.

Challenges Identified:

- Slow Operation Speed: A single cycle of picking, moving, and placing a rice ball took approximately one minute, limiting the maximum score achievable to *15* points.
- Shelf Height Issue: The shelf height was lower than the HSR's usual operational range, leading to frequent grasping failures due to an unconventional gripping method.

5.2. *Zero-Shot Object Detection*

Table. 2 shows that the proposed method detected 133 out of 202 rice balls, achieving an accuracy of 65.8% relative to visible rice balls. In contrast, Grounding DINO identified only 32 rice balls, corresponding to a 15.8% detection rate. Fig. 2 shows the detection results for Image ID 10. As shown in Figure 2 (a), the detection results of Grounding DINO reveal that language-based object recognition tends to identify the entire group of rice balls as a single entity rather than detecting each individual rice ball separately.

6. Conclusion

In this study, we introduce a high-performance object detection method for densely packed objects using color-based point prompts within the WRS FCSC. We validated our approach by participating in the WRS FCSC. In our preliminary experiments, language-based object recognition achieved a detection rate of 15.8%. In contrast, our proposed method reached a detection rate of 65.8%, outperforming the language-based approach by approximately 4.2 times in detecting densely packed rice ball. Our team placed seventh in this competition. The proposed approach efficiently identifies numerous objects without requiring the additional training typically necessary for CNN or Transformer-based detectors. Implementing and testing the method on the HSR confirmed its practical potential, highlighting its suitability for real-world autonomous retail applications.

Acknowledgements

This research is based on results from a JPNP16007 project commissioned by the New Energy and Industrial Technology Development Organization (NEDO). This research received support from JSPS KAKENHI Grant Number 23H03468 and 23K18495, as well as from JST ALCA-Next Grant Number JPMJAN23F3.

References

1. P. Kmecl, M. Munih, and J. Podobnik, Towards Autonomous Retail Stocking and Picking: Methods Enabling Robust Vacuum-Based Robotic Manipulation in Densely Packed Environments. *Sensors*, 2024, *24*, 6687.
2. K. Wada, New robot technology challenge for convenience store, *2017 IEEE/SICE International Symposium on System Integration (SII)*, Taipei, Taiwan, 2017, pp. 1086-1091.
3. L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, 2021, *J Big Data* 8, 53.
4. A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W. Lo, P, Dollár and R. Girshick, Segment Anything, 2023, arXiv:2304.02643
5. N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, K. V. Alwala, N. Carion, C.-Y. Wu, R. Girshick, P. Dollár and C. Feichtenhofer, SAM 2, arXiv:2408.00714
6. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You Look Only Ones, arXiv:1506.02640
7. S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, C. Li, J. Yang, H. Su, J. Zhu and L. Zhang, Grounding DINO: Marrying DINO with Grounded PreTraining for Open-Set Object Detection, 2023, arXiv:2303.05499
8. T. Yamamoto, K. Terada, A. Ochiai, F. Saito, Y. Asahara and K. Murase. Development of human support robot as the research platform of a domestic mobile manipulator, ROBOMECH journal, Vol. 6(1), 2019, pp. 1-15.

Authors Introduction

Mr. Naoki Yamaguchi

He received the B.Eng. degree from the National Institute of Technology, Ube College, Japan, in 2023. He is a master's degree student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include dataset creation and Visualization.

Mr. Tomoya Shiba

He received the B.Eng. degree from National Institute of Technology, Kagoshima College, Japan, in 2021. He received the M.Eng. from Kyushu Institute of Technology, Japan, in 2023. He is currently in a Ph.D. student in the graduate school of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interest includes image processing, motion planning, and domestic service robots.

Prof. Hakaru Tamukoh

Hakaru Tamukoh received the B.Eng. degree from Miyazaki University, Japan, in 2001, and the M.Eng. and Ph.D. degrees from the Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a Postdoctoral Research Fellow at the Kyushu Institute of Technology, from April 2006 to September 2007. He was an Assistant Professor with the Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He is currently a Professor with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include digital hardware design, neural networks, and home service robots.

Application of AI Robot Technology for Biophilic Design

Kairi Manabe, **Ryo Miyazono**
Graduate School of Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata, Kitakyushu, Fukuoka, 804-8550, JAPAN

Keitaro Ito, **Tomomi Sudo**, **Naoto Ishizuka**
Department of Civil Engineering and Architecture, Faculty of Engineering, Kyushu Institute of Technology, Japan
1-1 Sensui-cho, Tobata, Kitakyushu, Fukuoka, 804-8550, JAPAN

Akinobu Mizutani
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan

Hakaru Tamukoh, Yuichiro Tanaka
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan

Etsushi Ueda, Yuki Anamizu
Graduate School of Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata, Kitakyushu, Fukuoka, 804-8550, JAPAN

Leon Furuya, Honoka Tamai, Saya Nakano
School of Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata, Kitakyushu, Fukuoka, 804-8550, JAPAN

Hirofumi Tanaka
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan
Email: manabe.kairi926@mail.kyutech.jp, miyazono.ryo847@mail.kyutech.jp, ito.keitaro230@mail.kyutech.jp, sudo.tomomi313@mail.kyutech.jp, ishizuka.naoto105@mail.kyutech.jp, mizutani.akinobu515@mail.kyutech.jp, tamukoh@brain.kyutech.ac.jp, tanaka-yuichiro@brain.kyutech.ac.jp, ueda.etsushi236@mail.kyutech.jp, anamizu.yuki656@mail.kyutech.jp, furuya.reon171@mail.kyutech.jp, tamai.honoka807@mail.kyutech.jp, nakano.saya367@mail.kyutech.jp, tanaka@brain.kyutech.ac.jp

Abstract

To enhance human health and well-being, Biophilic design has been increasingly recognized in recent years. This design is characterized by the integration of natural elements such as plants, nature light, and water into spaces. However, challenges are posed by the maintenance of live plants, as their decline can be caused by insufficient environmental conditions. The aim of this study is to propose a robotic system of autonomously relocate houseplants to environments optimized based on sensor data, including light, temperature, and humidity. Through the integration of AI robotics with ecological design principles, it is aimed to enhance sustainability and redefine the relationship between humans, nature, and technology, fostering a harmonious interaction among "robots, nature, and humans.".

Keywords: Biophilic design, Ecological design, Space design

1. Introduction

This research aims to explore new space design methods using AI robot technology, drawing on plant ecology and physiology knowledge.

The study focuses on biophilic design using AI robot technology. Biophilic design is based on the Biophilia hypothesis, which suggests that humans tend to focus on and affiliate with nature and other life forms [1]. It is a methodology for designing building environments that achieve long-term sustainability by restoring and enhancing people's positive relationships with nature [2]. Since then, many studies have shown that the natural environment positively affects human psychological and physiological health [3][4].

Living plants are essential to biophilic design. Maintaining optimal environmental conditions including adequate light, water, and nutrients—is crucial for plant survival. However, the ongoing maintenance and associated costs present significant challenges to the large-scale incorporation of houseplants. To address these issues, Plantroid has been developed [5] that moves autonomously to a sunny location based on the voltage generated by the solar cells mounted on the plant. Focusing on plant photosynthesis, research is also being conducted on a method for controlling the movement of a flowerpot robot to maximize photosynthesis within a limited space and time [6]. Additionally, research has focused on evaluating indoor environments for plant growth, including methods for analyzing light conditions using specific wavelengths [7] and assessing photosynthesis indices in buildings [8].

This paper proposes a robotic system designed for the autonomous relocation of houseplants, ensuring their healthy growth by enabling them to select and move to suitable environments for their survival. The study will specifically address developing a robot that moves plants and generates an "environmental gradient map," enabling houseplants to determine their optimal moving route.

2. Method

2-1 Experimental environment

The experiment was conducted at “Fancγ”, the Research Center for Neuromorphic AI Hardware base. This space was designed under the theme of “Creating Biophilic Design for People and Robots”. The design of this space was proceeded through the crowdfunding [9]. A collaboration was established in 2023 between the laboratories of Life Science and Systems Engineering, the Laboratory of Environmental Design, and the Laboratory of Architectural Design within the Department of Civil Engineering and Architecture at Kyushu Institute of Technology. This space is designed for the development of robots through a range of experiences and learning opportunities in human living environments that people use daily for meetings and relaxation. As shown in Fig. 1, the continuous space is organized into distinct zones, including a Gallery・Archive space, a Meeting space, a Work space, and a Refresh space.

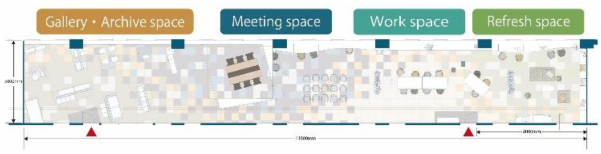


Fig. 1 Floor plan of “Fancγ”

2-2 The demonstration of moving houseplants using AI robots

We developed a system for moving houseplants using autonomous transport robots called Kachaka. These robots were tested by transporting plants along a designated route. Our platform integrates Kachaka with plants and features a voice command system for controlling the robots.

2-2-(1) Autonomous houseplant transport robot

This study used the Kachaka robot from Preferred Robotics, designed for transporting items in homes and workplaces [10][11]. Measuring 24 cm wide and 38.7 cm deep, the Kachaka can carry up to 20 kg with its specialized bases. We selected three species and eight houseplants growing well in low light. Using two Kachaka units, we placed the plants on six bases and moved them around as needed.

The plants were placed on a Kachaka base, as shown in Fig. 2, with two designated positions: home and sunbathing. The home position separates the workspace from the refresh space, while the sunbathing position is near the window. Kachaka has a map of the room, where we manually marked the positions. The robot uses a LiDAR sensor for localization and can be instructed via a voice recognition system called Whisper. When a human says, "relaxation mode," the robot moves the plants to the home position. If "seminar mode" is commanded, the plants are moved to the sunbathing position to expand the room capacity.

Fig. 2 Kachaka with houseplants

2-3 Measuring data and generating an “environmental gradient map”

To develop a system that helps houseplants move to optimal growth environments, it is vital to collect data on temperature, illuminance, and humidity. These factors are crucial for plant health and can be used to generate an "environmental gradient map" for the robot's reference.

2-3-(1) Preliminary survey

This study employs bilinear interpolation to estimate data at non-measured points. A preliminary survey was conducted to assess this method, with sensors placed at five locations within a 2000 mm × 2000 mm grid: the four corners and the center. The sensors used were “OKUDAKE Sensor Loggers” from SUN ELECTRONICS CO., LTD., with measurement errors of

±0.4°C for temperature, ±5% for illuminance, and ±3% for humidity (0% to 80%), increasing to ±4.5% for humidity (80% to 100%). The values of the center point were interpolated from the corner data, and accuracy was evaluated by comparing these calculated values to the actual measurements.

2-3-(2) Measuring environmental information

Temperature, illuminance, and humidity were measured in the refresh space of the experimental space "Fancγ." Over a 24-hour period, sensors recorded data from 12 indoor locations and 1 outdoor location at 5-minute intervals. The indoor sensors were positioned 1000 mm from the floor, matching the level of the houseplant leaves that the robot would move. Details of the experiment are shown in Figs. 3 and 4.

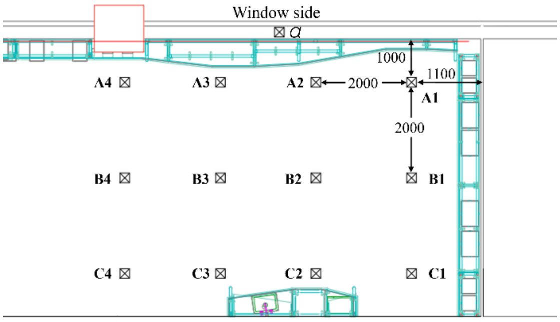


Fig. 3 Sensor arrangement diagram

Fig. 4 Refresh space with sensors installed

2-3-(3) Data analysis process

We analyzed the environmental data from the experimental space by interpolating non-measured points at 100 mm intervals to create continuous data. We converted the minimum, maximum, and median measurements into RGB values and plotted them as a scatter plot to visually represent the environmental gradient over time.

3. Results

3-1 The demonstration of moving houseplants using AI robots

The demonstration was conducted in the Fancγ. Two Kachaka were used to move six Kachaka bases with the plants, confirming the system's basic functions. In the demonstration, parameters, such as the position of the plants and the timing of the movement, were set manually. We introduce environmental sensors to set these parameters automatically by reflecting environmental conditions and human behaviors.

3-2 Measuring data and generating an "environmental gradient map"

3-2-(1) Results of preliminary survey

The center values were interpolated using bilinear interpolation from sensor data at the corners of a 2000 mm × 2000 mm grid. The Mean Square Errors (MSE) were as follows: temperature MSE was 0.00126, illuminance MSE was 0.297, and humidity MSE was 0.5625. This indicates that bilinear interpolation is suitable for estimating values at non-measured points in this study.

3-2-(2) Environmental data and environmental gradient map

The results of temperature, illuminance, and humidity measurements from each sensor are presented in Figs. 5, 6, and 7. Fig. 5 shows that the room temperature rises at 9:00 a.m., coinciding with an increase in the outside temperature. Data from window-side sensors, such as A3 and A4, exhibited trends similar to outside temperature. In the illuminance of A4 and B4 changes similarly to the illuminance outside the room. It can be read from this data that sunlight was reaching those locations. Fig. 7 shows that indoor humidity is stable between 36% and 41%, with no significant changes. In addition, changes in outside humidity and indoor humidity were similar between 8:00 and 9:00 a.m. and between 7:00 and 11:00 p.m.

Furthermore, environmental gradient maps were generated every 5 minutes from the measured data. Environmental gradient maps of temperature and illuminance at 9:00 a.m. are shown in Figs. 8 and 9. They show that the temperature and illuminance are higher on the window side and decrease as they approach the corridor side. Since different types of houseplants require different temperatures, this map could be used to determine where the robot carrying the houseplants should move. For the humidity environment map, there was not much change in indoor humidity at the same time.

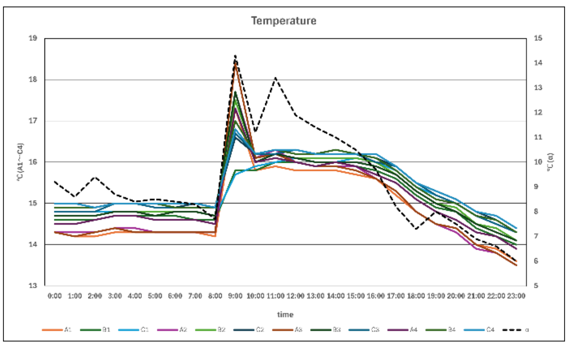


Fig. 5 Temperature change of sensor

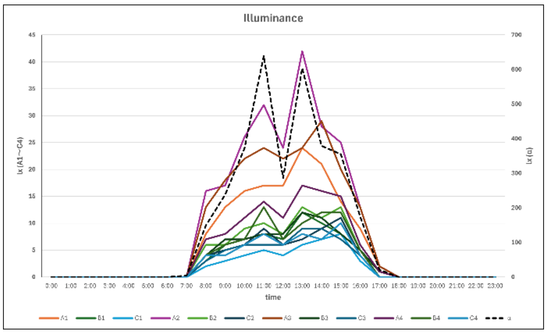


Fig .6 Illuminance change of sensor

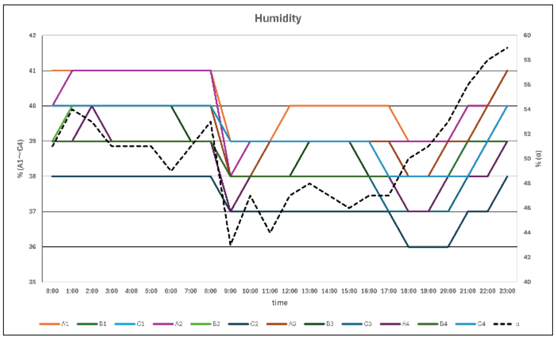


Fig .7 Humidity change of sensor

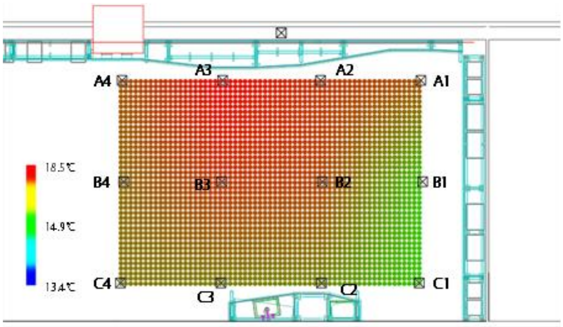


Fig. 8 emperature environment map at 9:00 a.m.

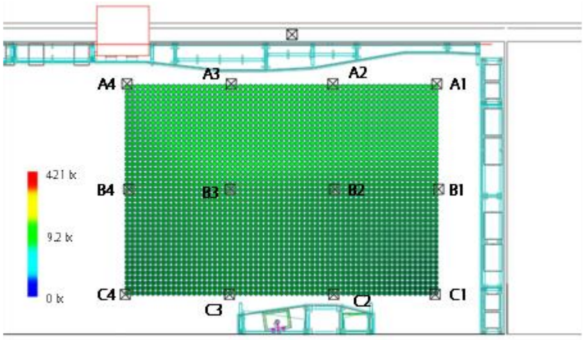


Fig. 9 Illuminance environment map at 9:00 a.m.

4. Discussion and Conclusion

In this study, we developed a system for the autonomous movement of houseplants using transport robots. While Plantroid focuses on plant growth, our robot responds to voice commands, allowing people to interact with plants. This enhances user convenience and promotes coexistence between humans and plants. Besides supporting healthy plant growth, the system also serves as an interior decorator and contributes to visual landscape design.

The environmental gradient maps utilize sensor data on illuminance, temperature, and humidity. Changes in these factors can help design optimal routes for plants to move along. The collected data indicates when and where plants should move throughout the day. However, continuous data collection is essential due to seasonal variations and human activities. Machine learning can effectively estimate conditions with fewer sensors, optimizing placement and reducing costs. Additionally, considering vertical data alongside horizontal data will enhance insights into plant growth and better optimization of plant movement in indoor environments.

Future research will focus on developing specific routes for relocating houseplants, referring to environmental gradient maps, to improve plant health and management efficiency. An integrated system will be established to continuously update these gradient maps by utilizing data collected from sensors on plants during their movement, enhancing the accuracy of environmental gradient maps and enabling houseplants' autonomous management. These results present new possibilities in biophilic design practices and houseplant maintenance methods. In addition, how a space composed of autonomously moving plants affects human space use and spatial perception will be further studied. This research is expected to propose new environmental and spatial design methods using AI robots, while new perspectives, such as plants as sensory organs in AI robots, are also introduced.

5. Acknowledgement

This project was supported by a crowdfunding project, “Create a Future Space Where Humans and Robots Live Together”.

6. References

1. Wilson, Edward O. (1985). Biophilia. Harvard University Press.
2. S. R. Kellert, J. Heerwagen, and M. Mador, "Biophilic Design: The Theory, Science, and Practice of Bringing Buildings to Life," John Wiley & Sons, Inc. 2008.
3. A.Elantary, Unleashing the Potential: The Impact of Biophilic Office Design on Enhancing Employee Productivity, Mansoura Engineering Journal. Vol.49, 2024.
4. U.Watwani, Investigating the Occupant’s Perception of Biophilia on the Health and WellBeing in a Hospital Setting, CATE2023, pp.89-98.
5. M.Yuasa, S.Nishiki, I.Mizuuchi, Development of Autonomous Movable Fruit Growing Plantroid, Proceedings Of the 2013 JSME Conference on Robotics and Mechatronics, 2013.
6. M.Yuasa, I.Mizuuchi, A Control Method of a Swarm of Plant Pot Robots for Utilization of

Sunlight Based on Photosynthetic Property and Preliminary Experiment by Using Real Robots, Proceedings Of the 2013 JSME Conference on Robotics and Mechatronics, 2013.

7. S.Sugano, R.Nitta, K.Shindo, M.Ishii, S.Tanabe, Application of Spectral Irradiance Simulation to Biophilic Design Part.1 Calculation Method of Plant Growth Evaluation Metrics and Case Study Using Office Model, Online Technical Papers of Annual Meeting SHASE of Japan, 2020.
8. R.Nitta, S.Sugano, K.Shindo, M.Ishii, S.Tanabe, Application of Spectral Irradiance Simulation to Biophilic Design Part.2 Validation of Simulation and Evaluation of Light Environments based on Plant Growth Evaluation Metrics, Online Technical Papers of Annual Meeting SHASE of Japan, 2020.
9. https://readyfor.jp/projects/hma_wakamatsu
10. Kachaka, Preferred Robotics, https://events.kachaka.life/, Accessed: 2024-12-15.
11. Koji Terada, "Kachaka: Revolutionizing Home and Business Environments with Advanced Autonomous Service Robotics," The 5th International Symposium on Neuromorphic AI Hardware, 2024.

Authors Introduction

Mr. Kairi Manabe

He received the civil engineering degree from Kyushu Institute of Technology, Japan, in 2024. His research interests include Water purification and utilization in planting basins using charcoal and bamboo charcoal.

Mr. Ryo Miyazono

He received the civil engineering degree from Kyushu Institute of Technology, Japan, in 2024. His research interests include management of the school biotope from the perspective of ecosystem conservation and children's use.

Dr. Keitaro Ito

He received his Bachelor's, Master's, and Ph.D. degrees in Agricultural Science from Kyushu University in 1991, 1996, and 1998 respectively. As a professor in the Department of Civil Engineering and Architecture, he focuses on landscape ecology and design. His experiences include research and working in Japan, the UK, Germany, Norway, the US, and Thailand, and he was a Fulbright scholar at the University of Florida in 2020. His primary interest lies in ecological landscape design. Over the past 25 years, he has focused on designing nature restoration in urban areas, river restoration, primary school gardens, forest parks, and urban parks.

Dr.Tomomi Sudo

She received her B.E., M.E., and Ph.D. degrees from the Kyushu Institute of Technology in 2013, 2016, and 2019, respectively, and has been an assistant professor at the Department of Civil Engineering and Architecture since 2019. Her interest is in designing and managing natural environments from the perspective of people's use, experience, ecosystem services, and biodiversity.

Dr. Naoto Ishizuka

He received the B.E., M.E., and Ph.D. degrees from the Yokohama National University, in 2010, 2013, and 2019, respectively. He was an Assistant Professor with the Kyushu Institute of Technology, from 2018 to 2024. His research interests include architectural design. He is a member of AIJ and JIA.

Mr. Akinobu Mizutani

He received his B.Eng. and M.Eng. degrees from Kyushu Institute of Technology, Japan in 2020 and 2022, respectively. He is currently a doctoral course student at Kyushu Institute of Technology, Japan. His research interests include brain-inspired artificial intelligence and environmental design for home service robots. He is a student member of IEEE, IEICE, and RSJ.

Dr. Hakaru Tamukoh

He received the B.Eng. degree from Miyazaki University, Japan, in 2001, and the M.Eng. and Ph.D. degrees from the Kyushu Institute of technology, from April 2006 to September 2007. He was an Assistant and technology, from October 2007 to January 2013. He is currently a Professor with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include digital hardware design, neural networks, and home service robots.

Dr. Yuichiro Tanaka

He received the B.E., M.E., and Ph.D. degrees from the Kyushu Institute of Technology, in 2016, 2018, and 2021, respectively. He was a research fellow with the Japan Society for the Promotion of Science（JSPS）, from 2019 to 2021. He was an assistant professor at the Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology, from 2021 to 2024. He has been an associate professor with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology since 2024. His research interests include neural networks, digital hardware implementation. And home service robots. He is a member of IEEE, IEICE and JNNS.

Mr. Anamizu Yuki

He received the civil engineering degree from Kyushu Institute of Technology, Japan, in 2023. His research interests include people's behavior and awareness of environmental conservation.

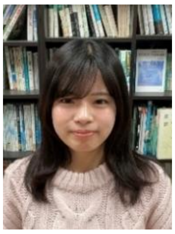
Mr. Ueda Etsushi

He received the civil engineering degree from Kyushu Institute of Technology, Japan, in 2023. His research interests include management of Satoyama.

Mr. Leon Furuya

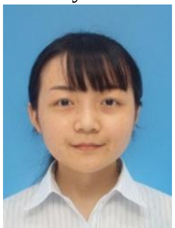
He is currently an undergraduate student at Kyushu Institute of Technology, Japan. His research interests include proposal of measures against Inland Flooding in urban areas.

Ms. Honoka Tamai



She is currently an undergraduate student at Kyushu Institute of Technology, Japan. Her research interests include proposal for outdoor open space design based on spatial characteristics and user behavior.

Ms. Saya Nakano



She is currently an undergraduate student at Kyushu Institute of Technology, Japan. Her research interests include proposal of pleasant walking paths to promote health.

Dr. Hirofumi Tanaka

He completed his doctorate in materials science by studying the structural and magnetic properties of ferromagnetic nanoalloys at Osaka University in 1999. Then, he moved to a national laboratory, RIKEN, as a postdoc researcher and then joined the Research Center for Molecular-Scale Nanoscience at the Institute for Molecular Science in 2003 as an assistant professor. He moved to the Department of Human Intelligence Systems, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology (Kyutech) as a full professor in 2014. He concurrently became a director of the Research Center for Neuromorphic AI Hardware, Kyutech, from 2020. He was awarded an honorary degree from Suranaree University of Technology, Thailand, in 2021, and received the degree certificate directly from Thai HRH Princess Sirindhorn. He also became a visiting professor of Universiti Putra Malaysia in 2024. His broad knowledge of materials, from metals and inorganic materials to organic materials, and techniques for measurement and fabrication helped lead efforts in molecular electronics and in combining many kinds of nanomaterials to realize a new world of "materials intelligence".

Basic Research on the Development of Space Standards for the Use of Service Robots in Housing Using the Urban Renaissance Agency's Housing Complex

Ren Matsuoka
Graduate School of Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata, Kitakyushu, Fukuoka, 804-8550, Japan

Kanon Nonoshita,
School of Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata, Kitakyushu, Fukuoka, 804-8550, Japan

Naoto Ishizuka
Graduate School of Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata, Kitakyushu, Fukuoka, 804-8550, Japan

Ryohei Kobayashi, Akinobu Mizutani, Hakaru Tamukoh, Hirofumi Tanaka
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology
Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan

E-mail: matsuoka.ren711@mail.kyutech.jp, nonoshita.kanon733@mail.kyutech.jp

Abstract

This study conducts basic research on developing space standards for integrating home service robots in residences, using a room in an housing complex from Urban Renaissance Agency as a model. The study compares these spaces with existing robot mobility standards to identify issues. The robot-friendly level of the room was evaluated using RFA standard. Results showed that most rooms fell into Level C, the lowest of three levels, in areas such as "fixture width" and "steps." Furthermore, when operating two types of home service robots, the study uncovered unique robot behavior issues in the housing that were not anticipated by RFA standard.

Keywords: Living space, Home service robot, Steps, Aisle widths, RFA standards, Urban Renaissance Agency

1. Introduction

In Japan, not only industrial robots but also service robots with independent mobility have begun to spread in recent years. Examples include guidance robots, transport robots, and cleaning robots for offices, public facilities, and commercial facilities. Cleaning robots, monitoring robots, pet robots, and small transport robots that can operate in the home are also commercially available. The safe and comfortable operation of these service robots can be considered mainly from the perspective of controlling the robot and maintaining the environment in which the robot operates.

This study examines the safe and comfortable operation of service robots from the viewpoint of environmental arrangements related to the robot's operation, i.e., the performance of the space in which the robot operates.

In Japan, the Robot Friendly Asset Promotion Association standard (RFA standard) (Fig.1) established by the Robot Friendly Asset Promotion Association is a standard for the safe and comfortable operation of robots in buildings. RFA standard is intended for public facilities, commercial facilities, and offices. It sets a Robot Friendly Level for each item, such as the size of steps, based on Japanese barrier-free laws and several service robot operation experiments. However, RFA standard does not assume the interior of homes and residences as the target space in its establishment process. We consider the difference between a house and a public facility. Generally, a house consists of private rooms such as living room, dining room, kitchen, and bedroom, and a corridor connecting them. Rooms are separated from each other by doors, and the width of these doors is usually around 900 mm. The width of corridors is generally around 780 mm based on architectural design modules, since there is little need to consider passing each other.

In that corridor, for example, consider the case of operating the Human Support Robot (HSR), which is being developed as a home service robot. [2] The footprint of HSR is 430 mm, so when a person and a robot pass each other, the passage width for the person would be 350 mm, which is quite narrow and would interfere with daily life. In addition, it is impossible for robots to pass each other. On the other hand, in commercial spaces that are public facilities, doors and other fittings that limit the entrance are not provided, and the frontage is generally large, in anticipation of the passage of an unspecified number of users. In addition, according to the Building Standard Law, corridors with rooms on both sides must be at least 1.6 m wide in floors of 200 m^2 or more. Thus, there is a large difference in standard dimensions between residences and public facilities due to the difference in the number of people using the space. In addition, many houses with many fittings have steps on the floor due to the sashes of the fittings, footwear slides, and guttering materials. On the other hand, the Barrier-Free Law requires public facilities to have as few steps as possible, and ramps to eliminate steps are easier to install in commercial facilities with a larger floor area than in residential facilities.

Therefore, environmental conditions for the safe and comfortable operation of home service robots may differ from those in public facilities, commercial facilities, and offices. Consequently, this study will organize issues on whether RFA standard can be used as an appropriate standard when aiming for the diffusion of home service robots. This is basic research on the standardization of space for the diffusion of home service robots in residences.

2. Outline and subject of this study

2.1. *Outline of this study*

In this study, we used a room in an housing complex of Urban Renaissance Agency (UR) as a model and conducted a measurement survey in the area where the thresholds are specified in RFA standard. Through this investigation, we verified where the modeled room was placed in RFA standard's Robot Friendly Level. This survey verified the relationship between robot friendliness level and actual home service robot behavior.

2.2. *Subject of this study (1)-RFA standard*

RFA standard was established in 2024 by the Robot Friendly Asset Promotion Association (RFA). This standard evaluates environmental factors that affect robot operations and quantifies them into three “Robot Friendly Levels” (Levels A through C). In developing the standard, environmental factors affecting robot operation were first identified through a literature review and on-site survey. In the literature review, factors such as passage widths, slopes, steps, and illumination were extracted based on building codes, barrier-free codes, and fire codes, and specific numerical standards were established. Next, a demonstration experiment was conducted using three meal delivery robots sold by Softbank Robotics. This experiment confirmed the presence or absence of problems caused by environmental factors and verified the validity of the robot-friendly level hypothetically defined in the literature survey.

RFA Standard (Facilitation of Service Robot Mobility - Classification and Indicators of the Physical Environment - Buildings and their Premises RFA B 0003: 2024)

Annex G “Robot-Friendly Level Definition List”

Item	summary	Level **A** The robot’s mobility is even greater than in Level B, Desired level	Level **B** Level of robot mobility without facility renovation	Level **C** Refurbishment of facilities, review of operations, etc. Level at which the robot becomes mobile	**Rationale and Considerations**
a) Steps	Excluding door sill of fittings.	6mm or less	-	More than 6mm	Refer to JIS T9251 (Regulations concerning the shape and dimensions of projections of guide blocks for the visually impaired and their arrangement) and “Guidelines for the Design of Housing for the Elderly”.
b) Groove	The width should be sufficiently wider than the robot, and should be set for the width and depth directions, assuming wheel slip-off.	Width: less than 10mm Depth: -	Width: 10mm or more but less than 20mm or Depth: 5mm or less	Width: 10mm or more but less than 20mm or Depth: 5mm or less	The standard is based on whether the casters fit or fall 5mm. The minimum caster width is 20 mm in JIS S 1038:1994 (casters for office chairs) and JIS B 8923:2015 (casters for industrial use). Since the curved ground surface, the caster may fit in the groove direction at robot-friendly level B. Robot-friendly level A is set regarding fine grating.
c) Aisle width	Set both sides of the aisle as vertical walls, about the width direction.	1.8m or more	1.2m or more but less than 1.8m	Less than 1.2m	1.2m is assumed to be the width necessary for a service robot and a person to pass each other, based on the Barrier-Free Act's facilitation standard. 1.8m is assumed to be the width necessary for robots to pass each other based on the Barrier-Free Law's facilitation and guidance standard. The induction standard has an effective dimension of 900 mm for entrances, exits, and flapper gates.
d) Fittings width and specifications	Set the effective width of doors and other openings for the robot to pass through, and the type of door, such as whether it can open and close automatically.	Width of fittings: 0.9m or more and Specifications: Automatic opening and closing door	Fittings width: 0.8m or more but less than 0.9m and Specifications: Automatic opening and closing door (Server integration functionality is required when security is involved.)	Fittings width: Less than 0.8m or Specifications: Not automatic opening and closing door	The effective width is the narrowest width including doorknobs and other protrusions. The 0.8m width of the fittings is based on the Barrier-Free Law's facilitation standard, and the 0.9m width is based on the same guidance standard.

Fig. 1 Excerpt of sizes that define the friendly level of the RFA standard [1]

2.3. *Subject of this study (2)-UR*

UR was established in 1955 as the Japan Housing Corporation to solve the severe housing shortage immediately after World War II. After its establishment, the demand for housing was for quantity for 20 years, including a period of high economic growth. [3] However, it changed from quantity to quality from around 1975, and from around 1995 to the revitalization and utilization of estate stock. And now, UR owns approximately 1,400 estates with a total housing stock of 700,000 units (as of March 2024). [4] However, due to the semi-public nature of these units, UR housing is currently difficult to replace. In addition, UR's management policy is to utilize housing complexes built after 1979 without demolishing them in principle. A survey by MLIT [5] is a comprehensive statistical survey of housing planning and spatial conditions in Japan. However, there is no statistical survey on details such as the existence and size of steps, which is necessary to verify consistency with RFA standards.

In UR housing, spatial conditions such as the existence and size of steps are different from one room to another. However, there exist specifications for UR housing that are referred to as models or standard designs. There are also standard specifications for materials and construction methods. Therefore, there is a possibility that some statistical analysis can be performed in the future. Because of these peculiarities of UR housing, we selected UR housing as the subject of this case study.

3. Field survey at Edamitsu estate

The subject of this survey is a room in the Edamitsu estate, a housing complex operated by UR (address: 4-14 Edamitsu, Yahatahigashi Ward, Kitakyushu City, Fukuoka Prefecture). The estate is a five-story reinforced concrete structure, and the subject room is located on the fourth floor of Building No. 14. The layout of this room is 3DK with an area of 65 m^2. In this survey, measurements were made for the items of “steps,” “gutters,” “passage width,” and “fixture width” among the environmental factors specified in RFA standard. The results are summarized in Fig. 2 below.

In addition, Table 1 shows the results of applying these values to the Robot Friendly Levels of RFA standard. Items such as “fixture width,” “steps,” and “groove depth” fell into Level C, the lowest of the three levels, in the majority of the houses.

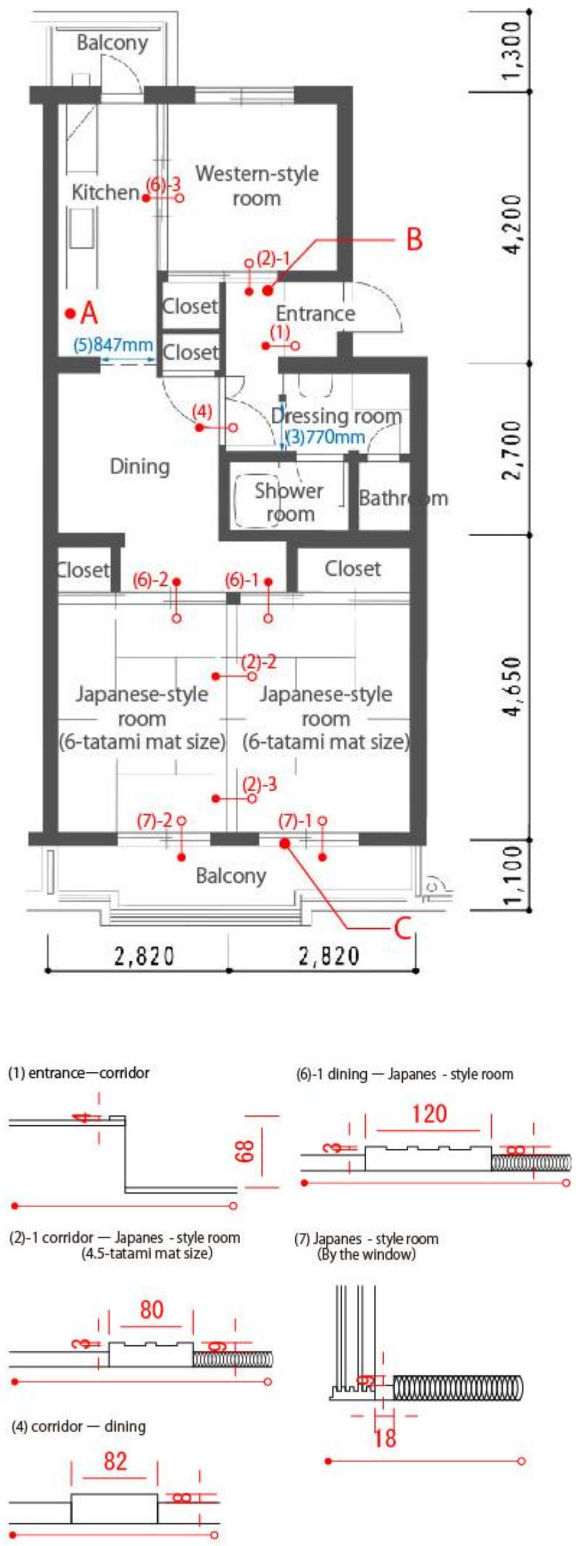


Fig. 2 Measurements of the physical environment [6]

Table1. Color-coded by level of RFA standard

Level A	Level B	Level C
The robot's mobility is even greater than in Level B, Desired level	Level of robot mobility without facility renovation	Refurbishment of facilities, review of operations, etc. Level at which the robot becomes mobile

		Step width	Step height	Groove	Fittings width / Aisle width
(1)	entrance — corridor		Outside → Inside: 68mm		
			Inside → Outside: 4mm		
(2)-1	corridor — Japanes - style room (4.5-tatami mat size)	80mm	9mm	3mm	880mm
(3)	corridor — dressing room				770mm
(4)	corridor — dining	82mm	8mm		1020mm
(5)	kitchen — dining				847mm
(6)-1	dining — Japanes - style room	120mm	8mm	3mm	838mm
(7)-1	Japanes - style room (By the window)	34mm	9mm		
		18mm	40mm		

4. Robot operation experiment

We also conducted operational experiments with two types of home service robots, Luna (Loona Blue, A34A0) and Roomba (Roomba Combo 10max, RCA-Y2), in the same room. For Roomba, we set up a station at Fig. 2 position A and used the mapping function to move the robot around the entire room. Roomba was unable to enter the shower room, which had a large step up. Roomba did not stop at any of the steps or widths we measured, except for the step at the entrance. At the step at the entrance, the step detection function was activated and Roomba did not fall. However, at a slight difference between the entrance and the Western-style room (Fig. 2 position B), Roomba was unable to escape and had to be moved by our hand.

Luna can perform any of its pre-defined actions, such as "summoning" and "returning to the charging station," by voice input. However, due to the characteristics of the AI used, it does not always perform as instructed. For this reason, we used a controller to make the robot perform arbitrary actions in this operation experiment. Luna was unable to enter the shower room, which had a large step up. Luna did not stop at any of the steps or widths we measured, except for Fig. 2 position C. At the step at the entrance, it did not detect the step and fell to the entrance dirt floor side. It can be pointed out that if the front door is released, Luna could exit the dwelling unit. In the situation where the opening on the terrace side was open and the screen door was closed (Fig. 2 position C), a wheel fell into the groove and Luna was unable to escape and had to be moved by our hand.

One issue related to movement that RFA standard does not seem to anticipate, which is unique to the house and unique to Japan, is the removal of shoes. For example, Luna can descend to the entrance earthen floor. When the difference in level is lowered due to a barrier-free dwelling, for example, it can be assumed that Luna will come up again from the entrance earthen floor to the flooring. It can be pointed out that dirt from the entrance dirt floor may be brought into the flooring.

5. Conclusion and Prospect

In this study, using one room in the Edamitsu estate, a estate of UR, we confirmed the Robot Friendly Level in RFA standard. The results showed that the majority of the rooms in the housing fell into Level C, the lowest of the three levels, for items such as "fixture width," "step height," and "groove depth. By running two different types of home service robots, we also discovered a challenge related to the robot's movements unique to the not believe is contemplated by RFA standard. UR owns many housing units with specifications similar to those of the Edamitsu estate. In other words, they represent a large proportion of the housing stock. The results of this paper showed that physical barriers at level C of RFA standard may not impede the movement of home service robots. If the development standards for home service robots are set to level A, the design would be over-specified for robots used in homes. What are the appropriate specifications for a home service robot? We believe that indicators for home service robots can be set by subdividing Level C of RFA standard.

As for prospects based on the results of this study, it is necessary to clarify the physical conditions under which the home service robots experience operational difficulties. We plan to conduct such movement experiments in the future.

Acknowledgment

The experiment at the Edamitsu Housing Complex is being conducted based on a collaboration agreement between UR Kyushu Branch and Kyushu Institute of Technology's Faculty of Engineering.

References

1. RFA, "Robot Friendly Asset Promotion Association web site",https://robot-friendly.org/publication/physical_definition_2024/, 2024.1.26 (In Japanese)
2. T. Yamamoto et al. "Development of Human Support Robot as the research platform of a domestic mobile manipulator", ROBOMECH Journal volume 6, Article number:4, 2019
3. "Japan Housing Corporation 20 Year History", Japan Housing Corporation 20 Year History Publication Committee, 1975 (In Japanese)
4. Urban Renaissance Agency, "Stock housing complex type (draft) list of UR rental housing", https://www.ur-net.go.jp/chintai_portal/stock/ ,2024.3 (In Japanese)
5. MLIT Japan, "Reiwa 5th year housing economy related data", https://www.mlit.go.jp/statistics/details/t-jutaku-2_tk_000002.html, 2023.7.30 (In Japanese)
6. Urban Renaissance Agency, "UR rental housing web site", https://www.ur-net.go.jp/chintai/kyushu/fukuoka/90_1260.html, 2024 (In Japanese)

Authors Introduction

Mr. Ren Matsuoka

He received his bachelor's degree in Engineering in 2024 from the Faculty of Engineering, Kyushu Institute of Technology in Japan. He is currently a master student in Kyushu Institute of Technology, Japan.

Ms. Kanon Nonoshita,

She is currently a bachelor's student in school of Engineering, Kyushu Institute of Technology, Japan.

D r. Naoto Ishizuka

He received the B.E., M.E., and Ph.D. degrees from the Yokohama National University, in 2010, 2013, and 2019, respectively. He was an Assistant Professor with the Kyushu Institute of Technology, from 2018 to 2024. His research interests include architectural design. He is a member of the Architectural Institute of Japan and The Japan Institute of Architects.

Mr. Ryohei Kobayashi

He received his Bachelor's degree in Engineering in 2024 from the College of Engineering, Tamagawa University in Japan. He is currently a master student in Kyushu Institute of Technology, Japan.

Mr. Akinobu Mizutani

He received his B.Eng. and M.Eng. degrees from Kyushu Institute of Technology, Japan in 2020 and 2022, respectively. He is currently a doctoral course student at Kyushu Institute of Technology, Japan. His research interests include brain-inspired artificial intelligence and environmental design for home service robots. He is a student member of IEEE, IEICE, and RSJ.

Dr. Hakaru Tamukoh

He received the B.Eng. degree from Miyazaki University, Japan, in 2001, and the M.Eng. and Ph.D. degrees from the Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a Postdoctoral Research Fellow at the Kyushu Institute of Technology, from April 2006 to September 2007. He was an Assistant Professor with the Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He is currently a Professor with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include digital hardware design, neural networks, and home service robots.

Dr. Hirofumi Tanaka

He completed his doctorate in materials science by studying the structural and magnetic properties of ferromagnetic nanoalloys at Osaka University in 1999. Then, he moved to a national laboratory, RIKEN, as a postdoc researcher and then joined the Research Center for Molecular-Scale Nanoscience at the Institute for Molecular Science in 2003 as an assistant professor. He moved to the Department of Human Intelligence Systems, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology (Kyutech) as a full professor in 2014. He concurrently became a director of the Research Center for Neuromorphic AI Hardware, Kyutech, from 2020. He was awarded an honorary degree from Suranaree University of Technology, Thailand, in 2021, and received the degree certificate directly from Thai HRH Princess Sirindhorn. He also became a visiting professor of Universiti Putra Malaysia in 2024. His broad knowledge of materials, from metals and inorganic materials to organic materials, and techniques for measurement and fabrication helped lead efforts in molecular electronics and in combining many kinds of nanomaterials to realize a new world of "materials intelligence".

An Exhibition Environment with 2D Markers for Guide Robot

Akinobu Mizutani
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan

Yui Hattori
Graduate School of Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata, Kitakyushu, Fukuoka, 804-8550, Japan

Naoto Ishizuka
Faculty of Engineering, Kyushu Institute of Technology
1-1 Sensui-cho, Tobata, Kitakyushu, Fukuoka, 804-8550, Japan

Yuichiro Tanaka, Hirofumi Tanaka, Hakaru Tamukoh
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan

Email: mizutani.akinobu515@mail.kyutech.jp, tamukoh@brain.kyutech.ac.jp
https://www.brain.kyutech.ac.jp/~neuro/

Abstract

The guide robot in an exhibition environment is expected to entertain visitors and reduce the maintenance cost of updating the robot database according to the changes in exhibition contents. The exhibition comprises printed panels for humans and a 2D marker printed with ultraviolet ink for robots. 2D markers are attached to the bottom of the exhibition furniture to make the markers invisible to visitors. The exhibition point may change depending on the contents of the exhibition. By searching for 2D markers autonomously in an exhibition space, the robot can update its internal database. This enables us to keep both the printed panel and the robot database without changing the robot database by users. The experiment is conducted in an exhibition environment, and the success rate in finding the exhibition point is evaluated.

Keywords: Guide robot, Kachaka, 2D marker, ArUco marker, Invisible printing, Exhibition environment

1. Introduction

Service robots are being developed for commercial environments. These robots typically have a camera and laser sensor for autonomous movement and a microphone and speaker for communicating with humans [1][2][3][4]. The guidance of facilities is one of the tasks of service robots. These robots must move autonomously to guide the environment. Two-dimensional markers and/or laser sensors are typically used for the localization of the robot.

Fancγ is the base of the Research Center for Neuromorphic AI Hardware, which has an exhibition, meeting, work, and relaxation area. This study focuses on an exhibition area of Fancγ and an environment for a guide robot. The robot in Fancγ is required to guide the guests of the research center and explain each display item to the guests. The display items and their explanation boards are located on a white exhibition platform. The update procedure for the robot database with changes in the display content should be simple. Non-experts in robots will change the display items; therefore, the internal robot program should not be modified each time.

This paper summarizes the environmental design concept of Fancγ and a system with a two-dimensional marker that matches the environment to locate the position of the display item for a robot.

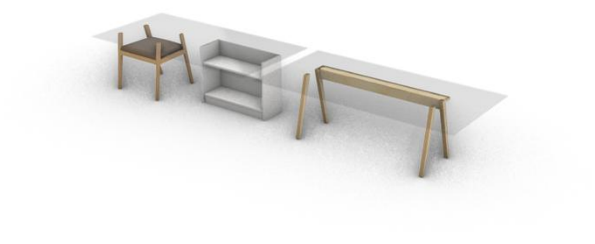


Fig. 1 The design concept of the exhibition platform

2. The Design Concept for the Exhibition Environment

The exhibition platform of Fancγ is composed of furniture in a hall, bedroom, kitchen, and living room in the home environment. The Furniture was cut to a height of 60 [cm] and an acrylic board was placed on it. The design concept of Fancγ is inspired by the idea of robot sensing range: home service robots only detect a part of the space. In particular, a robot with 2D LiDAR detects the bottom of the space. The area that the robot can detect is represented by the white part of the furniture, and the acrylic surface is the limit of the robot's perception (Fig. 1).

The Display items are typically located under the acrylic surface, and the explanation panel is located on top of the acrylic surface. The position and content of the display items change; therefore, the destination position and contents for explanation are also updated. The surface layout of each exhibition furniture is different. However, the variance of the foot in each exhibition furniture is less than that of the surface. The requirements of the robots in Fancγ are as follows:

(i) The robot should not collide with the acrylic surface at a height of 60 cm.
(ii) An update of the display items is included in the guide.
(iii) Modification of the exhibition platform should be minimal.

We propose an environment in which ArUco [5] markers are printed with invisible ink and attached to the foot of the furniture. The robot can identify its position with a serial number using an ArUco marker. By linking each display item with the serial number and its position, the robot can reach the destination position for guidance. We used invisible ink that reacts with ultraviolet (UV) light. Under daylight, the two-dimensional marker is invisible to humans; however, under UV light, it becomes visible. This contributes to maintaining the design concept of the room and conveying necessary information to the robot (requirement (iii)).

2.1. The operation flow of updating display items

The flow for updating the display items is as follows:

(i) Users update the positions and/or contents of the display items.
(ii) Users attach a two-dimensional marker to the foot of each exhibition furniture under the display item.
(iii) Users link the display item and marker serial number on a database

Here, humans simply link the display items and serial numbers of the two-dimensional marker, and users are not required to set the coordinates of each exhibition position.

3. Guide Robot

The guide robot has two operation modes: guide mode and update mode.

3.1. Guide mode

The robot uses the 2D LiDAR information for localization to move to specified positions. To provide an explanation, the robot moved to the destination position specified in the database.

3.2. Update mode

In this mode, the robot updates its internal database, serial number for each exhibition, and position by scanning a two-dimensional marker under UV light during the night. Guests are not in the exhibition environment and external light does not exist; therefore, it is suitable for the robot to scan two-dimensional code under UV light at night. The robot travels to a predefined scanning position, and if the robot detects markers at different locations in the previous scan, it updates the destination position for explanation.

4. Marker detection method

The ArUco marker is printed using special ink that is visible under UV light (blacklight ink). The ink emits light when it is exposed to UV light. The area where the ink is printed appears white, unlike a normal printed area. The inverted marker is printed using blacklight ink, as shown in Fig. 2, where (a) is the marker size, (b) is the

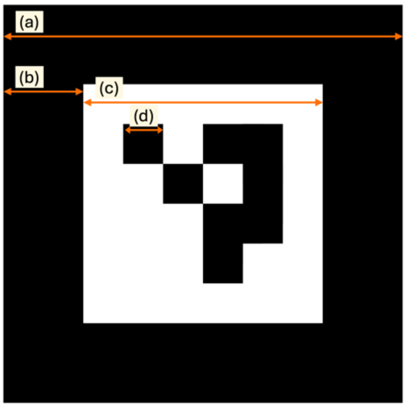


Fig. 2 ArUco marker for black light ink printing.

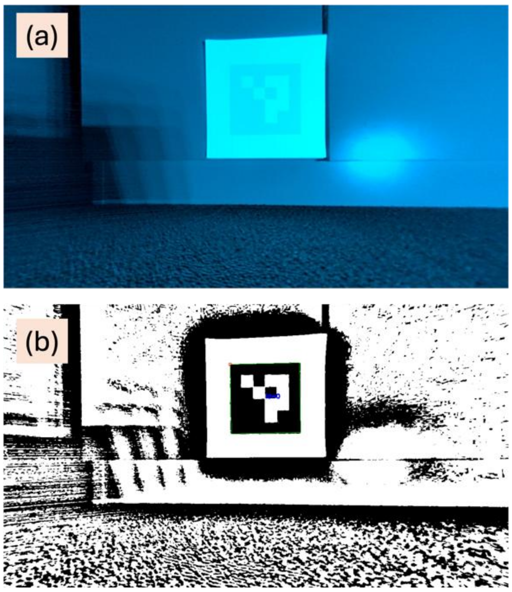


Fig. 3 Marker detection (a) Distortion-corrected input image (b) Binarized image with detection result (The detected bounding box and the ID of the marker is displayed in the image)

margin size, (c) is the module size, and (d) is the cell size. In addition, the robot maintains its distance from obstacles, and the robot cannot reach too close to the marker. Thus, the marker size is important for stable detection.

The distortion-corrected image from the camera was converted into a binarized image (Fig. 3 (a)). The adaptive Gaussian thresholding method [6] was applied to binarize the images (Fig. 3 (b)). Finally, the coordinates of the marker in the map were calculated and stored for guidance.

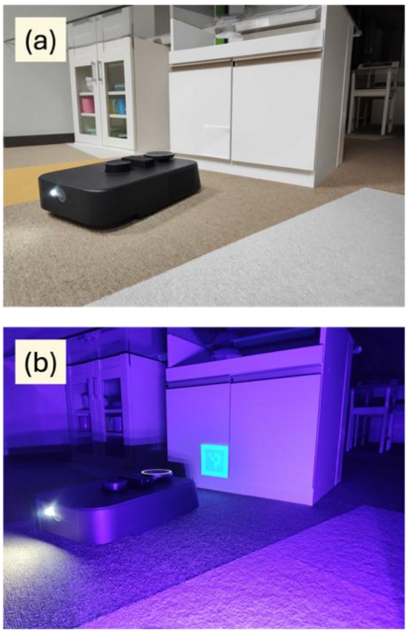


Fig. 4 Experiment Environment and marker (a) Under the daylight (b) Under the UV light

5. Experiment

The experiment setup is illustrated in Fig. 4. The ArUco marker was printed in blacklight ink and attached to the bottom of a white exhibition platform. Kachaka [1][2] was used as a guide robot, which is much lower than that an of acrylic surface. Kachaka has a camera in front and LiDAR sensor for navigation, and a speaker for audio

Table 1. Detection results by margin size

Marker ID	A1	A2	A3	A4
Margin size (b) [cell]	2.0	1.5	1.0	0.5
Module size (c) [pixel]	180			
Is detected	Yes	Yes	Yes	No

Table 2. Summary of the result.

Marker ID	B1	B2	B3
Marker size (a) [mm]	88	112	105
Margin size (b) [cell]	2	2	1.5
Module size (c) [mm]	66	84	84
N_d / N_t (ratio)	10/15 (0.666)	10/10 (1.000)	10/118 (0.085)
b [pixel]	151	191	191
$\bar{z}$ [mm]	0.23576	0.22813	0.27732
s_z [mm]	0.00075	0.00033	0.00620

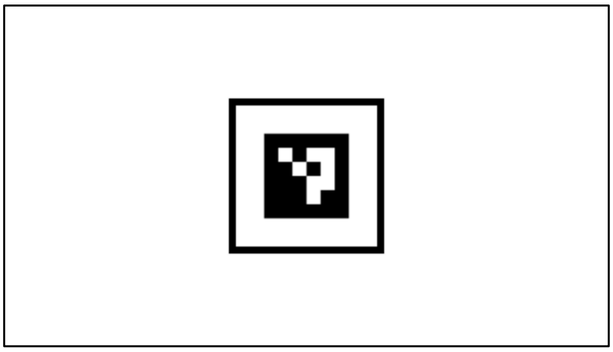


Fig. 5 ArUco marker used in a simulation experiment (marker A1)

guidance; thus, it meets the requirements (i) and (ii) of the robot in Fancγ and is suitable for a guide robot.

5.1. Margin around the module (simulation)

The binarized image contains a black frame around the marker. The necessary margin size around the module is confirmed using a simulated detection image, as shown in Fig. 5. A simulated image module with 180 [pixel] and 15 [pixel] of black frame was prepared. Four different proportions of cell margin size were evaluated. The results of the experiments are shown in Table 1. We confirmed that at least 1.0 [cell] of margin is required to detect the marker.

5.2. Marker size

We evaluated three different markers: B1, B2, and B3. Kachaka was navigated 0.23 [m] in front of the exhibition platform. The detection was performed every second. The number of detected markers (N_d) and the total number of trials (N_t) were counted, and the detection ratio was calculated. The distance from the front camera of Kachaka to the marker (z), the average ($\bar{z}$), and the standard deviation (s_z) were evaluated.

The experiment results are shown in Table 2. The detection rate of B3 was lower than that of the other markers. The standard deviation of B3 was also larger than those of the other two markers. Despite the simulation results, 2.0 [cell] of the margin was required for reliable detection. Markers smaller than B1 were not detected.

6. Discussion

In a real-world environment, a larger margin was required for marker detection. We think this is because the property of detection, such as noise, is not sufficiently reflected in a simulation environment. However, the margin size might also affect the binarization of the image; therefore, the detection accuracy is not only a problem of the ArUco marker itself. The detection performance of the marker increased with module size, and the margin size was larger. Markers B1 and B2 had sufficient accuracy for guiding robot applications. We can conclude that marker B1 is suitable for harmonizing with the design, and marker B2 is suitable for more stable detection.

7. Conclusion

We summarized the design concept of Fancγ and explained the guide robot system for locating each exhibition item. The special ink reacted to UV light and was applied to create an invisible marker in the daylight. The invisible marker has sufficient ability to locate exhibition contents and position them in a darker environment with UV light. Marker size and detection accuracy have a tradeoff, and two suitable marker sizes were found in the experiment. The implementation of the guide function and the evaluation of human acceptance will be planned in the future.

Acknowledgments

This project was supported by a crowdfunding project, “Create a Future Space Where Humans and Robots Live Together”.

References

1. Kachaka, Preferred Robotics, https://events.kachaka.life/, Accessed: 2024-12-15.
2. Koji Terada, “Kachaka: Revolutionizing Home and Business Environments with Advanced Autonomous Service Robotics,” The 5th International Symposium on Neuromorphic AI Hardware, 2024.
3. Temi, Temi, https://www.robotemi.com/, Accessed: 2024-12-15.
4. A. K. Pandey and R. Gelin, “A Mass-Produced Sociable Humanoid Robot: Pepper: The First Machine of Its Kind,” IEEE Robotics & Automation Magazine, Vol. 25, No. 3, pp. 40-48, 2018.
5. S. Garrido-Jurado, R. Muñoz-Salinas, F.J. Madrid-Cuevas, M.J. Marín-Jiménez, “Automatic generation and detection of highly reliable fiducial markers under occlusion,” Pattern Recognition, Vol. 47, No. 6, 2280-2292, 2014.
6. OpenCV, Adaptive Thresholding, https://docs.opencv.org/4.x/d7/d4d/tutorial_py_thresholding.html, Accessed: 2024-12-15.

Authors Introduction

Mr. Akinobu Mizutani

He received his B.Eng. and M.Eng. degrees from Kyushu Institute of Technology, Japan in 2020 and 2022, respectively. He is currently a doctoral course student at Kyushu Institute of Technology, Japan. His research interests include brain-inspired artificial intelligence and environmental design for home service robots. He is a student member of IEEE, IEICE, and RSJ.

Ms. Yui Hattori

She received her bachelor's degree in Engineering in 2024 from the Faculty of Engineering, Kyushu Institute of Technology in Japan. She is currently a master student in Kyushu Institute of Technology, Japan.

Dr. Naoto Ishizuka

He received the B.E., M.E., and Ph.D. degrees from the Yokohama National University, in 2010, 2013, and 2018, respectively. He was an Assistant Professor with the Kyushu Institute of Technology, from 2018 to 2024. His research interests include architectural design. He is a member of the Architectural Institute of Japan and The Japan Institute of Architects.

Dr. Yuichiro Tanaka

He received the B.E., M.E., and Ph.D. degrees from the Kyushu Institute of Technology, in 2016, 2018, and 2021, respectively. He was a research fellow with the Japan Society for the Promotion of Science (JSPS), from 2019 to 2021. He was an assistant professor at the Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology, from 2021 to 2024. He has been an associate professor with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology since 2024. His research interests include neural networks, digital hardware implementation, and home service robots. He is a member of IEEE, IEICE, and JNNS.

Dr. Hirofumi Tanaka

He completed his doctorate in materials science by studying the structural and magnetic properties of ferromagnetic nanoalloys at Osaka University in 1999. Then, he moved to a national laboratory, RIKEN, as a postdoc researcher and then joined the Research Center for Molecular-Scale Nanoscience at the Institute for Molecular Science in 2003 as an assistant professor. He moved to the Department of Human Intelligence Systems, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology (Kyutech) as a full professor in 2014. He concurrently became a director of the Research Center for Neuromorphic AI Hardware, Kyutech, from 2020. He was awarded an honorary degree from Suranaree University of Technology, Thailand, in 2021, and received the degree certificate directly from Thai HRH Princess Sirindhorn. He also became a visiting professor of Universiti Putra Malaysia in 2024. His broad knowledge of materials, from metals and inorganic materials to organic materials, and techniques for measurement and fabrication helped lead efforts in molecular electronics and in combining many kinds of nanomaterials to realize a new world of "materials intelligence".

Dr. Hakaru Tamukoh

He received the B.Eng. degree from Miyazaki University, Japan, in 2001, and the M.Eng. and Ph.D. degrees from the Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a Postdoctoral Research Fellow at the Kyushu Institute of Technology, from April 2006 to September 2007. He was an Assistant Professor with the Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He is currently a Professor with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include digital hardware design, neural networks, and home service robots.

Prediction of Timing and Amount of Houseplants Watering by an Echo State Network on Jetson

Wataru Yoshimura, Koshun Arimura, Ryohei Kobayashi, Akinobu Mizutani, Tomoaki Fujino, Yuichiro Tanaka
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Tomomi Sudo, Naoto Ishizuka, Keitaro Ito
Graduate School of Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata-ku, Kitakyushu, 804-8550, Japan

Hakaru Tamukoh
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Email: yoshimura.wataru281@mail.kyutech.jp, arimura.koshun523@mail.kyutech.jp,
kobayashi.ryohei621@mail.kyutech.jp, mizutani.akinobu515@mail.kyutech.jp, fujino.tomoaki404@mail.kyutech.jp,
tanaka-yuichiro@brain.kyutech.ac.jp, sudo.tomomi313@mail.kyutech.jp, ishizuka@civil.kyutech.ac.jp,
ito.keitaro230@mail.kyutech.jp, tamukoh@brain.kyutech.ac.jp

Abstract

Cultivating of houseplants in biophilic designed spaces requires appropriate timing and amount of watering. However, determining them is challenging, as fluctuations in ambient temperature can influence these factors. We develop a system capable of predicting ambient temperature changes and determining the appropriate timing and amount of watering. The system acquires ambient data using sensors connected to a Jetson Nano and processes the data using a neural network for the prediction and determination. We adopt an echo state network, a lightweight neural network, enabling a power-efficient system capable of running on edge devices. Additionally, we implement a function to notify the user of the timing and amount of watering via a chat service whenever the soil moisture content drops below a predefined threshold.

Keywords: Biophilic-design, Echo State Network, Edge device

1. Introduction

Houseplants have various effects on our lives [1]. For example, they clean the air and give relaxation as interior decorations. Recently, biophilic design has gained attention. This is a methodology for designing building environments that achieve long-term sustainability by restoring and enhancing people's positive relationship with nature [2]. It is wide-ranging, for example, nature restoration design in artificial environments [3][4][5][6], biomimicry design, and greening buildings to improve people's well-being. Using houseplants is a key component of biophilic design for indoor building environments, and its function and benefits are reported [7][8]. To maintain a biophilic design space, it is important to properly manage the environment for houseplants, such as lightning and watering conditions.

However, managing the environment to grow houseplants is difficult because plants placed indoors depend on sunlight from outside and artificial lighting for their light supply. Sunlight from outside is affected by factors, such as the building structure, the sun angle depending on the region, seasonal changes, and weather. Additionally, artificial lighting often results in uneven light distribution indoors when buildings are designed based only on human needs. Lighting schedules are also controlled by human activities, with lights being turned on and off, making it difficult to manage optimal light conditions for houseplants. For water supply, the required amount of water changes depending on the plant's photosynthesis, transpiration, growth, and the indoor temperature and humidity. Even in indoor environments, the temperature and humidity can become uneven due to heating, cooling, ventilation, and humidity control from air conditioning systems. To address these issues, many studies focus on selecting plant types suitable for indoor greenery, mainly from the perspective of light supply. However, managing houseplants manually requires a lot of effort fundamentary. Also, not all caretakers have knowledge about plants, making it hard to provide the right amount of light and water at the right time.

Therefore, we develop a system to manage houseplants using AI that senses and predicts environmental data and notifies the user of the best watering time for the plants. For the hardware to run this system, we adopt the Jetson Nano [9], a device with an interface to connect sensors. Additionally, Jetson Nano is equipped with a graphics processing unit (GPU) and has a compact size, making it suitable for implementing edge AI. Since Jetson Nano has limited processing power, we cannot implement AI models with high computational load on it. Based on this, we build a system using a lightweight AI model, reservoir computing (RC).

2. Technologies and Equipments

This study adopts RC as a prediction model for environment data because RC is lightweight and suitable for time-series processing, such as environment data. We implement an echo state network (ESN) [10] for the RC model.

The proposed system collects various environmental data such as temperature and soil moisture in real-time, analyzes them, and makes watering decisions. In this experiment, we use BME280 sensor [11], which measures temperature, humidity, and pressure, and Adafruit STEMMA Soil Sensor [12], which measures soil moisture.

3. Proposed Method

The proposed system consists of an environmental prediction part and a user notification part. This section describes these parts, respectively.

3.1. *Environmental prediction*

Generally, the optimal watering frequency for houseplants is approximately once every 4-5 days in spring and autumn, every 2-3 days in summer, and once every 1-2 weeks in winter. Therefore, we aim to implement a system that can predict environmental data a few days ahead.

For the environmental data prediction, we focus on temperature data among the environmental data. The proposed system feeds the temperature data into the ESN and trains it so that the ESN predicts one step ahead data. After the training, sequential forecasting, where the ESN output is fed to the ESN again to predict future steps, is available.

3.2. *User notification system*

When soil moisture remains below the proper range for a long time, plants may show mild stress signs, such as wilting leaves. Therefore, we focus on soil moisture here and develop the user notification system via LINE [13] when soil moisture decreases.

In this system, when the value from the soil moisture sensor falls below the threshold of 600, which was experimentally determined by obtaining the value when the soil becomes dry, the ESN predicts the temperature 10 minutes later. The value from the soil moisture sensor is a capacitance, with a range from approximately 200 (very dry) to 2000 (very wet). If the predicted temperature is higher than the current temperature, the system notifies the user to water the plants more. If the predicted temperature is lower than the current temperature, the system notifies the user to water the

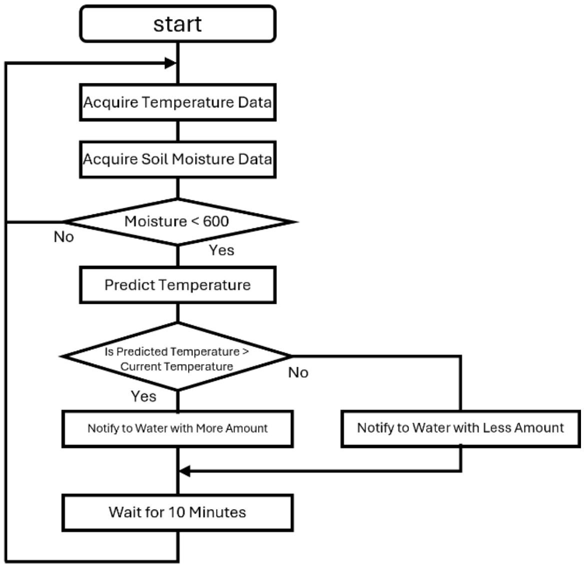


Fig 1. Flowchart of the system

plants less. These notifications act as a guideline for users to determine the appropriate amount of watering, helping them avoid overwatering or underwatering. Figure 1 shows the flowchart of the system.

4. Experimental Method

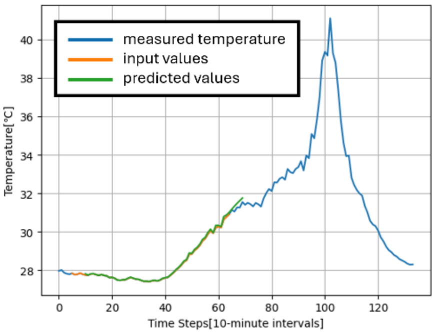
4.1. *Environmental prediction*

First, we created the training data using long-term observation data [14] published by the Japan Meteorological Agency. The long-term observation data consists of temperature records at 10-minute intervals. We created the training data by selecting 10 consecutive steps (100 minutes) from the 800 step (8000 minute) long-term data, resulting in a total of 790 sets of 10 steps.

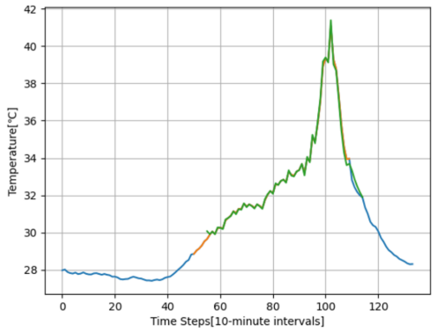
Next, we trained the ESN using the created training data by the ridge regression. The reservoir layer in the ESN consisted of 100 nodes. The connection weights for each node were randomly set values between -1 and 1, and the connection density was set to 0.5. Then, the connection weights were scaled so that the spectral radius became 0.99. The input layer had one node, and the connection weights from input to reservoir layer are randomly set between -0.3 and 0.3. The output layer had one node, and the connection weights from output to reservoir layer are randomly set between -0.01 and 0.01. The leak rate was set to 0.75. After the training, we verified the performance of the trained the ESN by giving untrained environmental dataset to it and comparing the predicted values with the actual values. Specifically, we fed 10 steps of untrained data into the ESN and compared prediction and the actual value.

4.2. *Sequential environmental prediction*

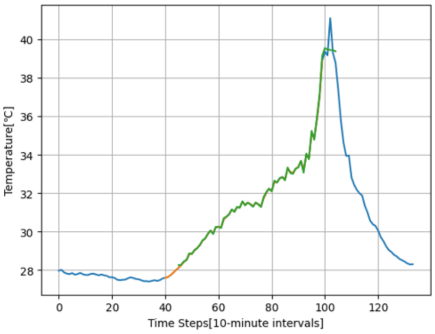
In this experiment, we used the model from Section 4.1 and temperature data observed on a specific day to evaluate the performance of sequential forecasting up to 10 steps ahead. First, we fed the actual values into the



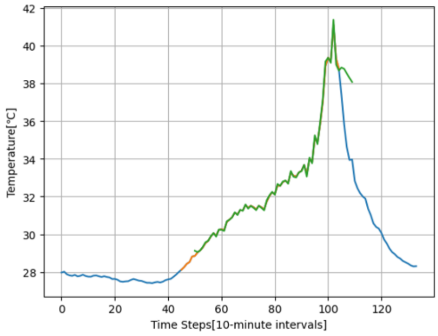
(a) During a smooth increase



(b) During a smooth decrease



(c) During a rapid increase



(d) During a rapid decrease

Fig 3. Evaluation results

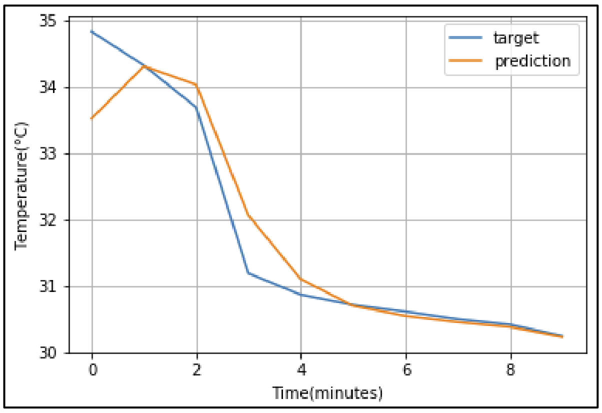


Fig 2. Temperature prediction using ESN

ESN up to a certain time. Then, starting from a specific time, we fed the predicted value from the ESN into the ESN for the next time as input and continue updating the reservoir state. A specific time refers to points where the data shows clear trends, such as smooth increases, smooth decreases, sharp rises, and sharp drops. This allows us to evaluate the performance of the prediction model under various conditions.

4.3. *Comparison of power consumption*

We measured the total power consumption in the system using a wattmeter [15]. We measured power consumption in two cases: when the system was in idle state (with no prediction model running) and when it run the temperature prediction using the ESN.

5. Results

5.1. *Environmental prediction*

We input measured temperature data for 10 steps into the trained ESN to check which step allows accurate prediction of the temperature data 10 minutes ahead.

Figure 2 shows an example of temperature prediction results. The blue line in the graph represents the data from the sensor, and the orange line represents the predicted data. In this example, from the 5th step, the sensor data and predicted data are almost the same, confirming that the temperature can be roughly predicted 10 minutes ahead.

5.2. *Sequential environmental prediction*

Figure 3(a) shows the prediction results for a smooth increase in temperature, Figure 3(b) shows the results for a smooth decrease, Figure 3(c) shows the results for a rapid increase, and Figure 3(d) shows the results for a rapid decrease. The blue line represents the measured temperature of the day, the orange line represents the input values to the ESN, and the green line represents the predicted values. Between time steps 65 and 70 in Figure 3(a) and between time steps 110 and 115 in Figure 3(b), the green line roughly follows the blue line. In contrast, between time steps 100 and 105 in Figure 3(c) and between time steps 105 and 110 in Figure 3(d), the green line does not follow the blue line and shows significant deviations. These results suggest that sequential

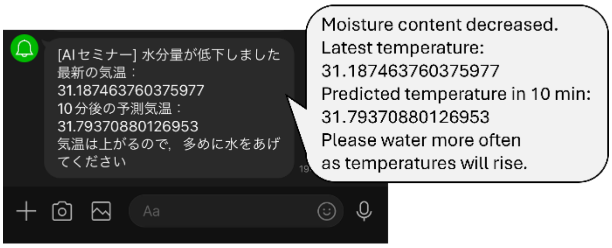


Fig 4. Execution results of the user notification system

forecasting can predict smooth changes, but it may be difficult to predict sudden changes.

5.3. *Comparison of power consumption*

The average power consumption in idle state was 35.85 [W], and the average power consumption during temperature prediction with ESN was 43.95 [W]. From this result, it was confirmed that when predictions using the ESN were made at short intervals of 1 second, the power consumption difference for each state was 8.1 [W]. In the system developed in this study, predictions using the ESN are made at 10-minute intervals, so the impact on power consumption by the ESN is small.

5.4. *Development of the user notification system*

Figure 4 shows the screenshot of the user notification system. The notification includes information about the decreased moisture level, the current temperature, the predicted temperature after 10 minutes, and the suggested watering amount. From this result, we confirmed that the notification system for the user worked correctly.

6. Conclusion

This study developed a system that predicts the timing and amount of watering for houseplants, which can run on an edge device using Jetson Nano. Specifically, the system uses an ESN for lightweight environmental prediction and sends appropriate notifications to the user based on data from sensors. Experimental results showed that this system accurately collects environmental data and can respond flexibly to changes in soil moisture. Additionally, predictions using the ESN can be performed with low power consumption on the edge device.

In the future, we introduce a function to predict sensor values for environmental factors other than temperature, such as humidity, light levels, and CO2 concentration. Additionally, we will analyze the growth of plants, including leaf color and shape, using image processing. This information be used as training data for a model that automatically adjusts the timing and amount of watering, aiming to further improve system accuracy. Moreover, we will enhance the system by predicting the optimal watering timing for each plant individually. As an application in biophilic design spaces, we develop robots with functions such as moving plants to watering stations based on the predicted environmental data.

References

1. V.-I. Lohr and C. H. Pearson-Mims, "Effects of Indoor Plants on Human Functions: A Systematic Review with Meta-Analyses," International Journal of Environmental Research and Public Health, vol. 19, no. 12, 2022.
2. S. R. Kellert, J. Heerwagen, and M. Mador, "Biophilic Design: The Theory, Science, and Practice of Bringing Buildings to Life," John Wiley & Sons, Inc. 2008.
3. K. Ito, T. Sudo, K. Ishimatsu, and H. Hasegawa, "Design and Ecological Management Process of Fishway and Surroundings." In: K. Ito (Eds.) Urban Biodiversity and Ecological Design for Sustainable Cities. Springer, 2021, pp.105-121.
4. K. Ito, T. Sudo, and I. Fjørtoft, "Ecological design: collaborative landscape design with school children." In Murnaghan, A.M.F. and Shillington, L.J. (Eds.) Children, Nature, Cities. UK: Routledge, 2016, pp. 95-209.
5. K. Ito, I. Fjørtoft, T. Manabe, and M. Kamada, "Landscape Design for Urban Biodiversity and Ecological Education in Japan: Approach from Process Planning and Multifunctional Landscape Planning," In: N. Nakagoshi, and J.A. Mabuhay, (Eds) Design Low Carbon Societies in Landscapes, Ecological Research Monographs. Springer, 2014, pp. 73-86.
6. K. Ito, I. Fjørtoft, T. Manabe, K. Masuda, M. Kamada, and K. Fujiwara, "Landscape Design and Children's Participation in a Japanese Primary School – Planning Process of School Biotope for 5 Years". In: Müller et al., (Eds.) Urban Biodiversity and Design. Willey-Blackwell, 2010, pp441-453.
7. F. Liu, L. Yan, X. Meng, C. Zhang, "A review on indoor green plants employed to improve indoor environment," J. Build. Eng., 53, 2022, 104542.
8. P. Chatakul, and S. Janpathompong, "Interior plants: Trends, species, and their benefits," Building and Environment, 2022, 222, 109325.
9. NVIDIA Corporation, "Jetson Nano Developer Kit," [Online]. Available: https://developer.nvidia.com/embedded/jetson-nano-developer-kit. Accessed: Dec. 8, 2024.
10. H. Jaeger, "The 'echo state' approach to analysing and training recurrent neural networks-with an erratum note," GMD Technical Report 148, German National Research Center for Information Technology, 2001.
11. Bosch Sensortec, "BME280 Combined Humidity and Pressure Sensor," [Online]. Available: https://www.bosch-sensortec.com/products/environmental-sensors/humidity-sensors-bme280/. Accessed: Dec. 8, 2024.
12. Adafruit Industries, "Adafruit STEMMA Soil Sensor - I2C Capacitive Moisture Sensor," [Online]. Available: https://www.adafruit.com/product/4026. Accessed: Dec. 8, 2024.
13. LINE Corporation, "LINE," [Online]. Available: https://line.me. Accessed: Dec. 8, 2024.
14. Japan Meteorological Agency, "Historical Weather Data Download," [Online]. Available: https://www.data.jma.go.jp/gmd/risk/obsdl/. Accessed: Dec. 8, 2024.
15. Sanwa Supply Co., Ltd., "TAP-TST10N Watt Monitor USB (2P, 1 Socket, White)," [Online]. Available: https://www.sanwa.co.jp/product/syohin?code=TAP-TST10N. Accessed: Dec. 8, 2024.

Authors Introduction

Mr. Wataru Yoshimura

He received his Bachelor's degree in Engineering in 2024 from the Advanced Course of Production Systems Engineering at the National Institute of Technology, Ube College in Japan. He is currently a master student in Kyushu Institute of Technology, Japan

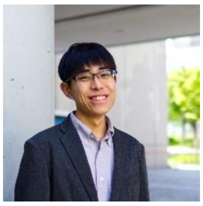
Mr. Koshun Arimura

He received his B.Eng. degree from Kyushu Institute of Technology, Japan, in 2024. He is currently a master's student at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include home service robots.

Mr. Ryohei Kobayashi

He received his Bachelor's degree in Engineering in 2024 from the College of Engineering, Tamagawa University in Japan. He is currently a master student in Kyushu Institute of Technology, Japan

Mr. Akinobu Mizutani



He received his B.Eng. and M.Eng. degrees from Kyushu Institute of Technology, Japan in 2020 and 2022, respectively. He is currently a doctoral course student at Kyushu Institute of Technology, Japan. His research interests include brain-inspired artificial intelligence and environmental design for home service robots. He is a student member of IEEE, IEICE, and RSJ.

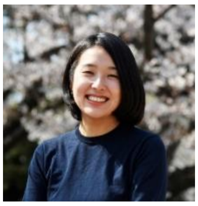
Mr. Tomoaki Fujino

He received his Bachelor's degree in Engineering in 2023 from the Advanced Course of Production Systems Engineering at the National Institute of Technology, Ube College in Japan. He is currently a master student in Kyushu Institute of Technology, Japan

Dr. Yuichiro Tanaka

He received the B.E., M.E., and Ph.D. degrees from the Kyushu Institute of Technology, in 2016, 2018, and 2021, respectively. He was a research fellow with the Japan Society for the Promotion of Science (JSPS), from 2019 to 2021. He was an assistant professor at the Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology, from 2021 to 2024. He has been an associate professor with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology since 2024. His research interests include neural networks, digital hardware implementation, and home service robots. He is a member of IEEE, IEICE, and JNNS.

Dr. Tomomi Sudo



She received her B.E., M.E., and Ph.D. degrees from the Kyushu Institute of Technology in 2013, 2016, and 2019, respectively, and has been an assistant professor at the Department of Civil Engineering and Architecture since 2019. Her interest is in designing and managing natural environments from the perspective of people's use, experience, ecosystem services, and biodiversity.

Dr. Naoto Ishizuka

He received the B.E., M.E., and Ph.D. degrees from the Yokohama National University, in 2010, 2013, and 2018, respectively. He was an Assistant Professor with the Kyushu Institute of Technology, from 2018 to 2024. His research interests include architectural design. He is member of Architectural Institute of Japan and The Japan Institute of Architects.

Dr. Keitaro Ito

He received his Bachelor's, Master's, and Ph.D. degrees in Agricultural Science from Kyushu University in 1991, 1996, and 1998. As a professor in the Department of Civil Engineering and Architecture, he focuses on landscape ecology and design. His experiences include research and working in Japan, the UK, Germany, Norway, the US, and Thailand, and he was a Fulbright scholar at the University of Florida in 2020. His primary interest lies in ecological landscape design. Over the past 25 years, he has focused on designing nature restoration in urban areas, river restoration, primary school gardens, forest parks, and urban parks.

Dr. Hakaru Tamukoh

He received the B.Eng. degree from Miyazaki University, Japan, in 2001, and the M.Eng. and Ph.D. degrees from the Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a Postdoctoral Research Fellow at the Kyushu Institute of Technology, from April 2006 to September 2007. He was an Assistant Professor with the Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He is currently a Professor with the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology. His research interests include digital hardware design, neural networks, and home service robots.

Practical Linearization Control of Nonholonomic Unicycles

Lixia Yan
The Seventh Research Division and the Center for Information and Control, School of Automation Science and Electrical Engineering, Beihang University (BUAA), Beijing 100191, China

Yingmin Jia
The Seventh Research Division and the Center for Information and Control, School of Automation Science and Electrical Engineering, Beihang University (BUAA), Beijing 100191, China
Email: yanlixia@buaa.edu.cn, ymjia@buaa.edu.cn

Abstract

Due to underactuation, the states of nonholonomic systems cannot be steered toward any direction of the state space. This note takes unicycles as an example and demonstrates a new idea of control design for nonholonomic systems. More precisely, we apply state transformation technique and external dynamic oscillator, and convert an underactuated nonholonomic unicycle into a fully-actuated and linearizable one. A control law, capable of tracking and stabilization uses, is then constructed. The tradeoff therein is that the tracking/stabilization errors can only be steered into the neighborhood of the origin rather than converging to zero. Numerical simulations are carried out to validate the proposed control scheme.

Keywords: Practical linearization, Practical stabilization, Nonholonomic unicycle.

1. Introduction

The states of unicycles cannot be arbitrarily steered in the state space due to nonholonomic constraints caused by the lack of lateral control input [1]. The control laws developed for unicycles can be divided into two types: full-state (position + orientation) [2] [3] [4] [5] and position-only controllers [6] [7], both of which focus on either tracking or stabilization purposes.

Time-varying [2] or discontinuous technique [3] is generally applied to develop the stabilizers for unicycles. In [2], a time-related oscillating function is fused into the visual control law that achieves pose stabilization with the unknown focal length of the camera. Via combing the σ-process and switching mechanism, the discontinuous controller in [3] ensures exponential convergence of pose stabilization errors. For tracking uses in torque level, the backstepping control law proposed in [4] obtains uniformly asymptotical convergence of position and orientation errors. In [5], a simple switching control law with an invariant set capable of simultaneous tracking and stabilization is proposed. Additionally, the position-only control schemes in [6] [7] choose a point in front of the body center as a position coordinate and convert the position derivative into a feedback linearizable form, making linear control approaches applicable. However, the feedback linearization control laws like those in [6] [7] do not consider the zero dynamics associated with uncontrolled orientation. As a result, how orientation behaves in the steady state remains unknown. Therefore, it is significant to investigate the full-state feedback linearization schemes for unicycles.

Motivated by the discussion above, we generalize the state transformation technique used in our previous work [5] and convert the original kinematic model of the unicycle into a new feedback linearizable form by fusing pose states with external oscillators and introducing an additional control input. A simple proportional control law in terms of feedback linearization is then proposed. We prove that the pose error for either time-invariant or time-varying desired signals can be stabilized into a small ball enclosing the origin.

The remainder is organized as follows. Section 2 includes the problem formulation and control design. Numerical simulation results are presented in Section 3. Section 4 concludes the work briefly.

2. Main Results

2.1. Problem Formulation

The typical kinematic model of a unicycle can be described by [5],

$$\dot{x} = u\cos\theta, \dot{y} = u\sin\theta, \dot{\theta} = r \qquad (1)$$

where $[x, y]^T$ is the position, θ denotes the orientation, u stands for the linear velocity, and r represents the angular velocity. Suppose that the desired pose signal $[x_d, y_d, \theta_d]^T$ obeys the same kinematics as (1) and satisfies the assumptions below,

Assumption 1. The time derivatives $[\dot{x}_d, \dot{y}_d, \dot{\theta}_d]^T \in L_\infty$.

Assumption 2. There exists a positive constant γ so that

$$\|[x_d, y_d, \theta_d]\|_2 \le \gamma \qquad (2)$$

Then, the control objective can be stated as: under Assumptions 1-2, find a control law (u, r) such that

$$\lim \|[x, y, \theta] - [x_d, y_d, \theta_d]\| \le \delta \qquad (3)$$

with a positive constant δ .

2.2. *Control Design*

First of all, we would like to convert the unicycle model (1) into a feedback linearizable form via state transformations. Define

$$\begin{cases} x_1 = x\cos\theta + y\sin\theta \\ y_1 = -x\sin\theta + y\cos\theta \\ \theta_1 = \theta \end{cases} \tag{4}$$

and calculate its derivative as below,

$$\dot{x}_1 = u + ry_1, \dot{y}_1 = -rx_1, \dot{\theta}_1 = r \tag{5}$$

For the purpose of practical linearization, an external dynamic oscillator in terms of auxiliary variables is needed. Let ϕ be an auxiliary variable and ε a small positive constant, we define the following new states

$$\begin{cases} z_1 = \theta_1 - \varepsilon\cos\phi \\ z_2 = x_1 - \varepsilon\sin\phi \\ z_3 = 2y_1 + \left(z_1 + \varepsilon\cos\phi\right)\left(z_2 + \varepsilon\sin\phi\right) \end{cases} \tag{6}$$

Calculating the time derivative of (6) leads to

$$\begin{cases} \dot{z}_1 = r + \varepsilon\dot{\phi}\sin\phi \\ \dot{z}_2 = u + ry_1 - \varepsilon\dot{\phi}\cos\phi \\ \dot{z}_3 = \varepsilon^2\dot{\phi} - \dot{z}_1\left(z_2 + 2\varepsilon\sin\phi\right) + \dot{z}_2\left(z_1 + 2\varepsilon\cos\phi\right) \end{cases} \tag{7}$$

Define new control inputs $w_1 \triangleq \dot{z}_1, w_2 \triangleq \dot{z}_2, w_3 \triangleq \varepsilon^2\dot{\phi}$, we convert (7) into

$$\left[\dot{z}_1, \dot{z}_2, \dot{z}_3\right]^T = A\left[w_1, w_2, w_3\right]^T \tag{8}$$

where

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -z_2 - 2\varepsilon\sin\phi & z_1 + 2\varepsilon\cos\phi & 1 \end{bmatrix} \tag{9}$$

Obviously, the matrix A is invertible and the new model (8) is feedback linearizable. A more favorable feature is that the (8) is fully-actuated.

Following the (4) and (6), we perform some calculations on the desired signal and obtain,

$$\begin{cases} x_{1d} = x_d\cos\theta_d + y_d\sin\theta_d \\ y_{1d} = -x_d\sin\theta_d + y_d\cos\theta_d \\ \theta_{1d} = \theta_d \\ \dot{x}_{1d} = u_d + r_d y_{1d}, \dot{y}_{1d} = -r_d x_{1d}, \dot{\theta}_{1d} = r_d \end{cases} \tag{10}$$

Let

$$z_{1d} = \theta_{1d}, z_{2d} = x_{1d}, z_{3d} = 2y_{1d} + z_{1d}z_{2d} \tag{11}$$

and define errors,

$$e_1 = z_1 - z_{1d}, e_2 = z_2 - z_{2d}, e_3 = z_3 - z_{3d} \tag{12}$$

Before moving on, we propose the following lemma,

Lemma 1. There exists a positive constant δ so that (3) establishes, if $e_1 \to 0, e_2 \to 0, e_3 \to 0$ as $t \to +\infty$.

Proof. Using the conditions in the lemma and (6), one gets,

$$\begin{aligned} &\theta_1 - \theta_{1d} = \theta - \theta_d \to \varepsilon\cos\phi \\ &x_1 - x_{1d} \to \varepsilon\sin\phi \\ &y_1 - y_{1d} \to -\varepsilon\frac{z_{1d}\sin\phi + z_{2d}\cos\phi + \varepsilon\cos\phi\sin\phi}{2} \end{aligned} \tag{13}$$

For the term $y_1 - y_{1d}$, it is obvious that

$$\lim_{t\to\infty}\left\|y_1 - y_{1d}\right\| \le 0.5\varepsilon\left(\sqrt{z_{1d}^2 + z_{2d}^2} + \varepsilon\right) \tag{14}$$

which, together with the fact that $z_{2d}^2 \le x_d^2 + y_d^2$ and Assumption 2, leads to

$$\lim_{t\to\infty}\left\|y_1 - y_{1d}\right\| \le 0.5\varepsilon\left(\gamma + \varepsilon\right) \tag{15}$$

Therefore,

$$\left\|\left[x_1 - x_{1d}, y_1 - y_{1d}\right]\right\| \le 0.5\varepsilon\left(\gamma + \varepsilon + 2\right) \tag{16}$$

Let $S(*) = \begin{bmatrix} \cos* & \sin* \\ -\sin* & \cos* \end{bmatrix}$, we know that

$$\begin{aligned} \begin{bmatrix} x_1 - x_{1d} \\ y_1 - y_{1d} \end{bmatrix} &= S(\theta)\begin{bmatrix} x \\ y \end{bmatrix} - S(\theta_d)\begin{bmatrix} x_d \\ y_d \end{bmatrix} \\ &= \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x - x_d \\ y - y_d \end{bmatrix} \\ &+ \left[S(\theta) - S(\theta_d)\right]\begin{bmatrix} x_d \\ y_d \end{bmatrix} \end{aligned} \tag{17}$$

and

$$\begin{aligned} S(\theta) - S(\theta_d) &= \begin{bmatrix} \cos\theta - \cos\theta_d & \sin\theta - \sin\theta_d \\ -\sin\theta + \sin\theta_d & \cos\theta - \cos\theta_d \end{bmatrix} \\ &= 2\sin\frac{\theta - \theta_d}{2}\begin{bmatrix} -\sin\frac{\theta + \theta_d}{2} & \cos\frac{\theta + \theta_d}{2} \\ -\cos\frac{\theta + \theta_d}{2} & -\sin\frac{\theta + \theta_d}{2} \end{bmatrix} \end{aligned}$$

where the sum-to-product formula is used. Henceforth,

$$\left\|\left[S(\theta) - S(\theta_d)\right]\begin{bmatrix} x_d \\ y_d \end{bmatrix}\right\| \le \left\|\theta - \theta_d\right\|\left\|\left[x_d, y_d\right]\right\| \tag{18}$$

By Assumption 2 and the fact $\left\|\left[x_d, y_d\right]\right\| \le \left\|\left[x_d, y_d, \theta_d\right]\right\|$, we use (17) and obtain

$$\lim_{t\to+\infty}\left\|\left[x - x_d, y - y_d\right]\right\| \le 0.5\varepsilon\left(3\gamma + \varepsilon + 2\right) \tag{19}$$

The ultimate bound δ can then be estimated by,

$$\delta = 0.5\varepsilon\left(3\gamma + 3\varepsilon + 2\right) \tag{20}$$

The claims in the lemma hence establish. □

Let $K = \mathrm{diag}\left\{k_1, k_2, k_3\right\}$ be a positive definite matrix, we design the control inputs as,

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = A^{-1}\left\{\begin{bmatrix} \dot{z}_{1d} \\ \dot{z}_{2d} \\ \dot{z}_{3d} \end{bmatrix} - K\begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}\right\} \tag{21}$$

The original control inputs can be recovered via the following procedures,

$$\dot{\phi} = \frac{w_3}{\varepsilon^2}, r = w_1 - \varepsilon\dot{\phi}\sin\phi, u = w_2 - ry_1 + \varepsilon\dot{\phi}\cos\phi \quad (22)$$

Then, we summarize the main results of the current work in the theorem below,

Theorem 1. Given Assumptions 1-2, the application of the control law (21)(22) on the unicycle (1) achieves the control objective (3).

Proof. Using (22) into (21), we obtain,

$$\dot{e}_1 = -k_1 e_1, \dot{e}_2 = -k_2 e_2, \dot{e}_3 = -k_3 e_3 \quad (23)$$

which demonstrates that $[e_1, e_2, e_3]^T$ converges to zero exponentially [8]. Using Lemma 1, we conclude that the control objective (3) is achieved. □

Remark 1. The pose errors in steady state are not stabilized to zero and the actual pose states would converge into the neighborhood around the desired signal, see Fig.1.

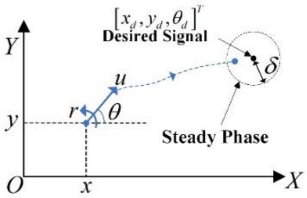


Fig.1 Pose trajectories in steady state.

3. Numerical Simulation

In this section, we perform the simulation with three different desired signals to verify the control scheme. The control coefficients for all cases are selected as $\varepsilon = 0.2, k_1 = k_2 = k_3 = 5.5$ identically. We initialize the unicycle states by $x(0) = -4, y(0) = -4.5, \theta(0) = -0.5\pi$. We choose the following desired signals:

Case 1. Circular trajectory

$\dot{x}_d = 0.2\cos\theta_d, \dot{y}_d = 0.2\sin\theta_d, \dot{\theta}_d = 0.025$

$x_d(0) = 0, y_d(0) = -8, \theta_d(0) = 0$

Case 2. Straight line

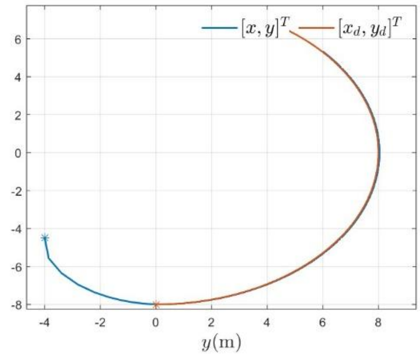
$\dot{x}_d = 0.2\cos\theta_d, \dot{y}_d = 0.2\sin\theta_d, \dot{\theta}_d = 0$

$x_d(0) = -3.5, y_d(0) = -3, \theta_d(0) = 0.25\pi$

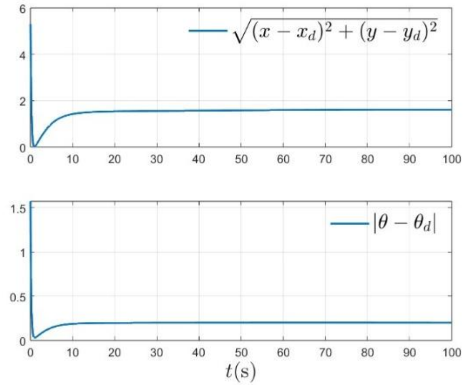
Case 3. Fixed point

$x_d \equiv 0, y_d \equiv 0, \theta_d \equiv 0$

The results are depicted in Fig. 2-4. As can be seen, all results accord with the theoretical analysis and the unicycle can stabilize itself to the neighborhood of the desired signal that can be either time-varying or time-invariant. The position trajectories seem to be zig-zag during transient phase, due to the auxiliary variable ϕ is oscillating for compensating the underactuation. Note also that the ultimate bounds of pose tracking errors can be reduced by decreasing ε, which can be drawn from Lemma 1 and (20).

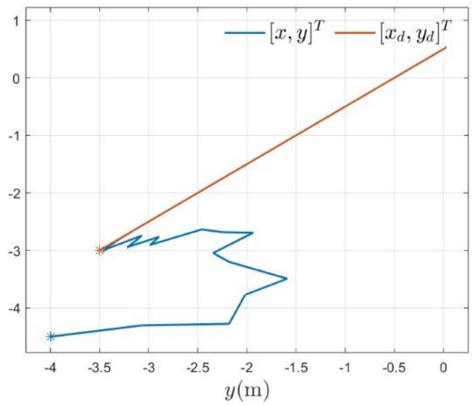


a .Position trajectories (*: starting point)

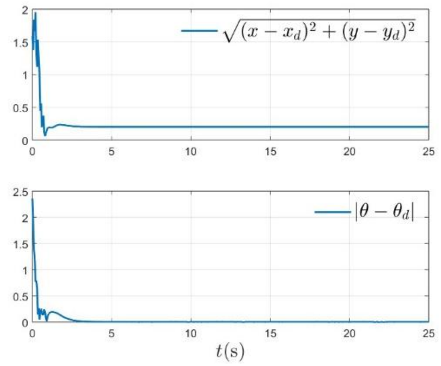


b . Pose error trajectories

Fig.2 Simulation results of Case 1.

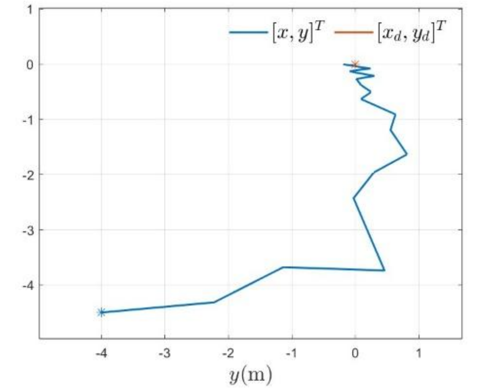


a . Position trajectories (*: starting point)

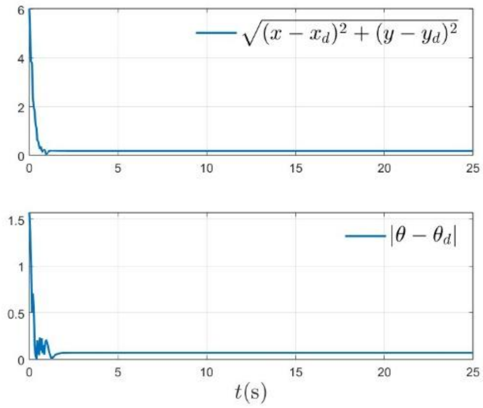


b . Pose error trajectories

Fig.3 Simulation results of Case 2.



a . Position trajectories (*: starting point)



b . Pose error trajectories

Fig.4 Simulation results of Case 3.

4. Conclusion

This work proposes a practical linearization control law for nonholonomic unicycles. The underactuated property is compensated by an additional control input fused with external oscillators after state transformations. The feedback linearizable form of the unicycle system is also illustrated. It of worthy noting that the errors with respect to the desired pose signal can converge to a small ball enclosing the origin. In the future, we will generalize the presented control law to the coordinated control of networked unicycles.

Acknowledgements

This work was supported in part by the NSFC (62133001, 62227810, 62403037) and the National Basic Research Program of China (973 Program: 2012CB821200, 2012CB821201).

References

1. L. Yan and B. Ma, Adaptive practical leader-following formation control of multiple nonholonomic wheeled mobile robots, International Journal of Robust Nonlinear Control, Vol.30(17), 2020, pp.7216-7237.
2. X. Zhang, Y. Fang, and N. Sun, Visual servoing of mobile robots for posture stabilization: from theory to experiments, International Journal of Robust and Nonlinear Control, Vol. 25(1), pp.1-15, 2015.
3. A. Astolfi, Discontinuous control of nonholonomic systems, Systems and Control Letters, Vol.27(1), pp.37-45, 1996.
4. R. Fierro and F. L. Lewis, Control of a nonholonomic mobile robot: Backstepping kinematics into dynamics, Journal of Robotic Systems, Vol.14(3), pp.149-163, 1998.
5. L. Yan and B. Ma, Universal Control for Both Rendezvous and Tracking of Multiple Nonholonomic Unicycles, IEEE Transactions on Control of Network Systems, Vol.11(1), 2024, pp. 439-449.
6. W. Xie and B. Ma, Position centroid rendezvous and centroid formation of multiple unicycle agents, IET Control Theory and Applications, Vol.8(17), pp.2055–2061, 2014.
7. M. A. Maghenem, A. Loria, and E. Panteley, Cascades-based leaderfollower formation tracking and stabilization of multiple nonholonomic vehicles, IEEE Transactions on Automatic Control, Vol.65(8), pp. 3639–3646, 2019.
8. H.K., Khalil, Nonlinear Systems (3rd Edition), Patience Hall, 2002.

Authors Introduction

Dr. Lixia Yan

He received B.S. degree from Beijing Jiaotong University, Beijing, China in 2013, and the M.S. and Ph.D degrees from Beihang University, Beijing, China in 2016 and 2021, respectively. He is currently a postdoctor with the Seventh Research Division and the Center for Information and Control at Beihang University. His research interests include nonlinear control theory, multiagent systems and embedded system applications.

Prof. Yingmin Jia

He received B.S. degree from Shandong University, Jinan, China, in 1982, and the M.S. and Ph.D. degrees from Beihang University, Beijing, China, in 1990 and 1993, respectively.

He is currently a professor with the Seventh Research Division and the director of the Center for Information and Control at Beihang University. His research interests include robust control, robotic systems, spacecraft coordination and on-orbit servicing.

Task-space Tracking Control for Dual-arm Free-floating Space Manipulators with Disturbances and Uncertainties

Qian Sun
The Seventh Research Division and the Center for Information and Control, School of Automation Science and Electrical Engineering, Beihang University (BUAA), Beijing, 100191, P.R.China

Yingmin Jia*
The Seventh Research Division and the Center for Information and Control, School of Automation Science and Electrical Engineering, Beihang University (BUAA), Beijing, 100191, P.R.China
E-mail: mssunqian@buaa.edu.cn, ymjia@buaa.edu.cn

Abstract

This paper investigates task-space tracking control for dual-arm free-floating space manipulators (DFFSM) subject to unknown disturbances, kinematic and dynamic uncertainties. First, we design a sliding mode disturbance observer to compensate for the lumped disturbances, including unknown disturbances and dynamic uncertainties. Then, an adaptive dynamic surface controller (ADSC) is proposed for DFFSM system, where the uncertain kinematics is estimated by an adaptive algorithm. It is validated through Lyapunov analysis that the tracking errors of the end-effectors are uniformly ultimately bounded under the proposed control scheme. Numerical simulations validate the effectiveness of the proposed control scheme.

Keywords: Dual-arm free-floating space manipulator, Task space, Disturbance observer, Dynamic surface control

1. Introduction

In recent years, space manipulators have increasingly replaced or assisted astronauts in performing various on-orbit servicing. Compared to single-arm space manipulators, DFFSM hold enhanced flexibility and operational capability, making them widely utilized in complex space missions. Therefore, they have attracted significant attention in the field of aerospace technology [1].

The task-space tracking control for DFFSM aims to drive the end-effector of each arm to track a desired trajectory. However, DFFSM are subject to disturbances and uncertainties in space environment, requiring the designed controller with strong robustness. Some researchers have focused on robust tracking control for dual-arm space manipulators [2], [3], [4], [5], [6]. Shi et al. [2], [3] investigated the coordinated control of the base attitude and the manipulator motion in task space, and proposed a sliding mode controller subject to system uncertainties. In [4], an adaptive fuzzy control scheme was developed for grasped dual-arm robots. An approximate Jacobian matrix and a decentralized fuzzy logic system were employed to handle uncertain kinematics and dynamics, and a novel parameter adaptation technique was applied to achieve practical finite-time convergence in the estimation of kinematic parameters and fuzzy logic weights. Aiming to capture a spinning spacecraft with a dual-arm space robot, Cheng et al. [5] constructed a recurrent fuzzy neural network to approximate the uncertain inertial parameters. Then, a coordinated stabilization control strategy with H_∞ tracking characteristics was then proposed for the post-capture phase. Based on predefined-time stability theory, Liu et al. [6] designed a general nonsingular terminal sliding mode control law for DFFSM subject to persistent disturbances. Despite these achievements, robust tracking control for DFFSM in task space remains limited, which requires further investigation.

Motivated by the discussions above, this paper investigates a disturbance observer-based ADSC for DFFSM. A sliding mode disturbance observer is designed to compensate for unknown disturbances and dynamic uncertainties, while an adaptive law is established to estimate the kinematic uncertainties. The proposed control scheme ensures the uniformly ultimate boundedness of the tracking errors in task space.

2. Main Results

2.1. *Kinematics and Dynamics for DFFSM*

The DFFSM system includes a base spacecraft and two *n*-degrees-of-freedom (*n*-DOF) manipulators, as shown in Fig.1. The kinematic model of the DFFSM is given by

$$\begin{bmatrix} \dot{\boldsymbol{x}}_e^a \\ \dot{\boldsymbol{x}}_e^b \end{bmatrix} = \boldsymbol{J}_g\left(\boldsymbol{x}_b, \boldsymbol{q}^i, m, \boldsymbol{I}\right)\begin{bmatrix} \dot{\boldsymbol{q}}^a \\ \dot{\boldsymbol{q}}^b \end{bmatrix}, \tag{1}$$

*Corresponding author.

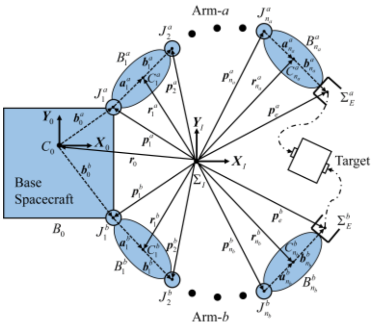


Fig.1 General Structure of DFFSM

where $\dot{\boldsymbol{x}}_e^i \in \boldsymbol{R}^{m_i}$ ($i = a, b$) denotes the velocity and angular velocity of the end-effector of Arm-*i*, $\boldsymbol{q}^i \in \boldsymbol{R}^{n_i}$ is the generalized coordinate vector of Arm-*i* with n_i -DOF, $\boldsymbol{J}_g \in \boldsymbol{R}^{(m_a+m_b)\times(n_a+n_b)}$ is the generalized Jacobian matrix of the DFFSM. Using Lagrange method, a reduce-order form dynamic model of DFFSM is derived as

$$\boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{q}} = \boldsymbol{\tau} + \boldsymbol{d}, \tag{2}$$

where $\boldsymbol{q} = [(\boldsymbol{q}^a)^{\mathrm{T}}, (\boldsymbol{q}^b)^{\mathrm{T}}]^{\mathrm{T}}$, $\boldsymbol{\tau} = [(\boldsymbol{\tau}^a)^{\mathrm{T}}, (\boldsymbol{\tau}^b)^{\mathrm{T}}]^{\mathrm{T}} \in \boldsymbol{R}^{n_a+n_b}$, $\boldsymbol{\tau}^i \in \boldsymbol{R}^{n_i}$ is the control input torque, $\boldsymbol{d} \in \boldsymbol{R}^{n_a+n_b}$ is the unknown disturbances, and $\boldsymbol{M}(\boldsymbol{q})$, $\boldsymbol{C}(\boldsymbol{q},\dot{\boldsymbol{q}})$ are the inertia matrix and the centrifugal and Coriolis matrix, where

$$\boldsymbol{C}(\boldsymbol{q},\dot{\boldsymbol{q}}) = \dot{\boldsymbol{M}}(\boldsymbol{q})\dot{\boldsymbol{q}} - \frac{\partial}{\partial \boldsymbol{q}}\left(\frac{1}{2}\dot{\boldsymbol{q}}^{\mathrm{T}}\boldsymbol{M}(\boldsymbol{q})\dot{\boldsymbol{q}}\right).$$

In general, the kinematics and dynamics of DFFSM are uncertain due to the complicated structures. The dynamic uncertainties can be described as

$$\begin{cases} \boldsymbol{M}(\boldsymbol{q}) = \boldsymbol{M}_0(\boldsymbol{q}) + \Delta\boldsymbol{M}(\boldsymbol{q}), \\ \boldsymbol{C}(\boldsymbol{q},\dot{\boldsymbol{q}}) = \boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}}) + \Delta\boldsymbol{C}(\boldsymbol{q},\dot{\boldsymbol{q}}), \end{cases} \tag{3}$$

where $\boldsymbol{M}_0(\boldsymbol{q})$ and $\boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}})$ represent the nominal parts, $\Delta\boldsymbol{M}(\boldsymbol{q})$ and $\Delta\boldsymbol{C}(\boldsymbol{q},\dot{\boldsymbol{q}})$ are the unknown perturbed parts. Substituting (3) into (2), one has

$$\boldsymbol{M}_0(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{q}} = \boldsymbol{\tau} + \boldsymbol{\delta}, \tag{4}$$

where $\boldsymbol{\delta} = \boldsymbol{d} - \Delta\boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} - \Delta\boldsymbol{C}(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{q}}$ denotes the lumped disturbances, which satisfies $\|\boldsymbol{\delta}\| \le \beta$ and $\|\dot{\boldsymbol{\delta}}\| \le \beta_d$ with β and β_d being known positive constants.

Property 1. $\boldsymbol{J}_g\dot{\boldsymbol{q}}$ can be linearly parameterized to $\boldsymbol{\theta}_k$:

$$\dot{\boldsymbol{x}}_e = \boldsymbol{J}_g\dot{\boldsymbol{q}} = \boldsymbol{Y}_k(\boldsymbol{q},\dot{\boldsymbol{q}})\boldsymbol{\theta}_k, \tag{5}$$

where $\dot{\boldsymbol{x}}_e = [(\dot{\boldsymbol{x}}_e^a)^{\mathrm{T}}, (\dot{\boldsymbol{x}}_e^b)^{\mathrm{T}}]^{\mathrm{T}}$, $\boldsymbol{\theta}_k = [\theta_{k1}, \theta_{k2}, ..., \theta_{kq}]^{\mathrm{T}} \in \boldsymbol{R}^q$ is a constant parameter, $\boldsymbol{Y}_k(\cdot) \in \boldsymbol{R}^{(m_a+m_b)\times q}$ denotes the kinematic regressor matrix.

Property 2. The matrix $\dot{\boldsymbol{M}}_0(\boldsymbol{q}) - 2\boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}})$ is skew-symmetric, i.e., $\boldsymbol{v}^{\mathrm{T}}(\dot{\boldsymbol{M}}_0(\boldsymbol{q}) - 2\boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}}))\boldsymbol{v} = 0$ for all $\boldsymbol{q}$, $\dot{\boldsymbol{q}}$ and $\boldsymbol{v} \in \boldsymbol{R}^{n_a+n_b}$.

Property 3. For a given known positive constant c_0 , the matrix $\boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}})$ satisfies $\boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}}) \le c_0\|\dot{\boldsymbol{q}}\|$.

This paper aims at designing a control algorithm $\boldsymbol{\tau}$ for DFFSM system (1) and (2) to track the desired trajectory of the end-effectors $\boldsymbol{x}_{ed} \in \boldsymbol{R}^{m_a+m_b}$ in task space subject to unknown disturbances, kinematic and dynamic uncertainties.

2.2. *Sliding Mode Disturbance Observer Design*

Define a sliding mode surface $\boldsymbol{s} = \dot{\boldsymbol{q}} - \boldsymbol{\xi}$ with an auxiliary variable $\boldsymbol{\xi} \in \boldsymbol{R}^{n_a+n_b}$, where $\boldsymbol{\xi}$ is obtained by

$$\boldsymbol{M}_0(\boldsymbol{q})\dot{\boldsymbol{\xi}} + \boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}})\dot{\boldsymbol{q}} = \boldsymbol{\tau} + \lambda_0\boldsymbol{s}, \tag{6}$$

where $\lambda_0 > 0$ is a gain of system (6). The sliding mode disturbance observer is proposed as

$$\begin{cases} \boldsymbol{M}_0(\boldsymbol{q})\dot{\hat{\boldsymbol{s}}} = -\lambda_0\hat{\boldsymbol{s}} + \lambda_1\,\mathrm{sgn}(\tilde{\boldsymbol{s}}) + \hat{\boldsymbol{\delta}}, \\ \dot{\hat{\boldsymbol{\delta}}} = \lambda_2\tilde{\boldsymbol{s}} + \lambda_3\boldsymbol{\sigma} + (\beta_d + \lambda_4)\,\mathrm{sgn}(\boldsymbol{\sigma}), \end{cases} \tag{7}$$

where $\hat{\boldsymbol{s}}$ and $\hat{\boldsymbol{\delta}}$ denote the estimation of $\boldsymbol{s}$ and $\boldsymbol{\delta}$, respectively, $\tilde{\boldsymbol{s}} = \boldsymbol{s} - \hat{\boldsymbol{s}}$ is the estimation error of $\boldsymbol{s}$, the notation $\boldsymbol{\sigma}$ is defined as $\boldsymbol{\sigma} = \lambda_1\,\mathrm{sgn}(\tilde{\boldsymbol{s}})$, λ_j ($j = 2,3,4$) is a positive constant to be determined. Combining (2), (6) and (7), the dynamics of the estimation error $\tilde{\boldsymbol{s}}$ and $\tilde{\boldsymbol{\delta}}$ can be derived into

$$\begin{cases} \boldsymbol{M}_0(\boldsymbol{q})\dot{\tilde{\boldsymbol{s}}} = -\lambda_0\tilde{\boldsymbol{s}} - \lambda_1\,\mathrm{sgn}(\tilde{\boldsymbol{s}}) + \tilde{\boldsymbol{\delta}}, \\ \dot{\tilde{\boldsymbol{\delta}}} = -\lambda_2\tilde{\boldsymbol{s}} - \lambda_3\boldsymbol{\sigma} - (\beta_d + \lambda_4)\,\mathrm{sgn}(\boldsymbol{\sigma}) + \dot{\boldsymbol{\delta}}, \end{cases} \tag{8}$$

where $\tilde{\boldsymbol{\delta}} = \boldsymbol{\delta} - \hat{\boldsymbol{\delta}}$ is the estimation error of $\boldsymbol{\delta}$.

Theorem 1. Considering the dynamic model (2) and the designed sliding mode disturbance observer (7), if the observer gains are set as $\lambda_0 = c_0\|\dot{\boldsymbol{q}}\|$, $\lambda_1 = \beta - \|\hat{\boldsymbol{\delta}}\| + \bar{\lambda}_1$, $\bar{\lambda}_1 > 0$, and $\lambda_2 < 2\lambda_3$, the estimation error $\tilde{\boldsymbol{\delta}}$ could converge to zero in finite time.

Proof. The proof is divided into the following two steps:
Step 1: We will prove the finite-time convergence of $\tilde{\boldsymbol{s}}$. Choosing V_1 as a Lyapunov function candidate

$$V_1 = \frac{1}{2}\tilde{\boldsymbol{s}}^{\mathrm{T}}\boldsymbol{M}_0(\boldsymbol{q})\tilde{\boldsymbol{s}}. \tag{9}$$

The time derivative of V_1 is given as

$$\begin{aligned} \dot{V}_1 &= \tilde{\boldsymbol{s}}^{\mathrm{T}}\boldsymbol{M}_0(\boldsymbol{q})\dot{\tilde{\boldsymbol{s}}} + \frac{1}{2}\tilde{\boldsymbol{s}}^{\mathrm{T}}\dot{\boldsymbol{M}}_0(\boldsymbol{q})\tilde{\boldsymbol{s}} \\ &= \tilde{\boldsymbol{s}}^{\mathrm{T}}\left[-\lambda_0\tilde{\boldsymbol{s}} - \lambda_1\,\mathrm{sgn}(\tilde{\boldsymbol{s}}) + \tilde{\boldsymbol{\delta}}\right] + \tilde{\boldsymbol{s}}^{\mathrm{T}}\boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}})\tilde{\boldsymbol{s}} \\ &\le -(\lambda_0 - c_0\|\dot{\boldsymbol{q}}\|)\tilde{\boldsymbol{s}}^{\mathrm{T}}\tilde{\boldsymbol{s}} - (\lambda_1 - \|\tilde{\boldsymbol{\delta}}\|)\|\tilde{\boldsymbol{s}}\| \\ &\le -\bar{\lambda}_1\|\tilde{\boldsymbol{s}}\| \le -\frac{\sqrt{2}\bar{\lambda}_1}{\sqrt{\lambda_{\max}(\boldsymbol{M}_0(\boldsymbol{q}))}}V_1^{\frac{1}{2}}. \end{aligned} \tag{10}$$

From (10), it can be derived that $\tilde{\boldsymbol{s}} = \dot{\tilde{\boldsymbol{s}}} = \boldsymbol{0}$ within a finite time $t_{d1} \le \sqrt{2\lambda_{\max}(\boldsymbol{M}_0(\boldsymbol{q}))}V_1^{1/2}(0)/\bar{\lambda}_1$,where $V_1(0)$ is the initial value of V_1 . Based on the equivalent output injection theory, $\tilde{\boldsymbol{\delta}}$ is equivalent to the term $\lambda_1\,\mathrm{sgn}(\tilde{\boldsymbol{s}})$, i.e., $(\tilde{\boldsymbol{\delta}})_{\mathrm{eq}} = \boldsymbol{\sigma} = \lambda_1\,\mathrm{sgn}(\tilde{\boldsymbol{s}})$.
Step 2: We will prove the convergence of $\tilde{\boldsymbol{\delta}}$ when $\tilde{\boldsymbol{s}} = \boldsymbol{0}$. Considering the following Lyapunov function candidate

$$V_2 = \frac{1}{2}\tilde{\boldsymbol{\delta}}^{\mathrm{T}}\tilde{\boldsymbol{\delta}}. \tag{11}$$

Differentiating V_2 and using (8) yield

$$\begin{aligned} \dot{V}_2 &= \tilde{\boldsymbol{\delta}}^{\mathrm{T}}\dot{\tilde{\boldsymbol{\delta}}} \\ &= -\tilde{\boldsymbol{\delta}}^{\mathrm{T}}\left[\lambda_2\tilde{\boldsymbol{s}} + \lambda_3\boldsymbol{\sigma} + (\beta_d + \lambda_4)\,\mathrm{sgn}(\boldsymbol{\sigma}) - \dot{\boldsymbol{\delta}}\right] \\ &\le -\lambda_2\tilde{\boldsymbol{\delta}}^{\mathrm{T}}\tilde{\boldsymbol{s}} - \lambda_3\tilde{\boldsymbol{\delta}}^{\mathrm{T}}\tilde{\boldsymbol{\delta}} - (\beta_d + \lambda_4)\|\tilde{\boldsymbol{\delta}}\| + \|\dot{\boldsymbol{\delta}}\|\|\tilde{\boldsymbol{\delta}}\| \\ &\le -\lambda_4\|\tilde{\boldsymbol{\delta}}\| + \frac{\lambda_2}{2}\tilde{\boldsymbol{s}}^{\mathrm{T}}\tilde{\boldsymbol{s}} \le -\lambda_4\|\tilde{\boldsymbol{\delta}}\|. \end{aligned} \tag{12}$$

Eq.(12) indicates $\dot{V}_2 = -\sqrt{2}\lambda_4 V_2^{1/2}$. Therefore, the estimation error $\tilde{\boldsymbol{\delta}}$ finally converges to the origin in finite time

$T_d \leq t_{d1} + \sqrt{2}V_2^{1/2}(0)/\lambda_4$, where $V_2(0)$ is the initial value of V_2. The proof is completed.

2.3. *ADSC Design*

In this section, an ADSC is designed for DFFSM system (1) and (2). The control design procedure can be divided into the following two steps:

Step 1: Define the tracking error variable as

$$\boldsymbol{z}_1 = \boldsymbol{x}_e - \boldsymbol{x}_{ed}. \quad (13)$$

Differentiating $\boldsymbol{z}_1$ and using Property 1 yield

$$\dot{\boldsymbol{z}}_1 = \dot{\boldsymbol{x}}_e - \dot{\boldsymbol{x}}_{ed} = \boldsymbol{J}_g\dot{\boldsymbol{q}} - \dot{\boldsymbol{x}}_{ed} = \boldsymbol{Y}_k(\boldsymbol{q},\dot{\boldsymbol{q}})\boldsymbol{\theta}_k - \dot{\boldsymbol{x}}_{ed}. \quad (14)$$

The parameter $\boldsymbol{\theta}_k$ is unknown owing to kinematic uncertainties. We employ its estimation $\hat{\boldsymbol{\theta}}_k$ to replace $\boldsymbol{\theta}_k$, that is, $\hat{\dot{\boldsymbol{x}}}_e = \hat{\boldsymbol{J}}_g\dot{\boldsymbol{q}} = \boldsymbol{Y}_k(\boldsymbol{q},\dot{\boldsymbol{q}})\hat{\boldsymbol{\theta}}_k$. Rewrite (14) as

$$\dot{\boldsymbol{z}}_1 = \hat{\boldsymbol{J}}_g^{-1}\dot{\boldsymbol{q}} + \boldsymbol{Y}_k(\boldsymbol{q},\dot{\boldsymbol{q}})\tilde{\boldsymbol{\theta}}_k - \dot{\boldsymbol{x}}_{ed} \quad (15)$$

where $\tilde{\boldsymbol{\theta}}_k = \boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k$ represents the estimation error of $\boldsymbol{\theta}_k$. The virtual control $\boldsymbol{\alpha}$ and the adaptive law of $\hat{\boldsymbol{\theta}}_k$ are proposed as

$$\boldsymbol{\alpha} = \hat{\boldsymbol{J}}_g^{-1}(-c_1\boldsymbol{z}_1 + \dot{\boldsymbol{x}}_{ed}), \quad (16)$$

$$\dot{\hat{\boldsymbol{\theta}}}_k = \gamma_k\left[\boldsymbol{Y}_k^{\mathrm{T}}(\boldsymbol{q},\dot{\boldsymbol{q}})\boldsymbol{z}_1 - \mu_k\hat{\boldsymbol{\theta}}_k\right], \quad (17)$$

where c_1, γ_k, μ_k are positive control gains. Substituting (16) into (15) yields

$$\begin{aligned}\dot{\boldsymbol{z}}_1 &= \hat{\boldsymbol{J}}_g\boldsymbol{\alpha} + \boldsymbol{Y}_k(\boldsymbol{q},\dot{\boldsymbol{q}})\tilde{\boldsymbol{\theta}}_k - \dot{\boldsymbol{x}}_{ed} + \hat{\boldsymbol{J}}_g(\dot{\boldsymbol{q}} - \boldsymbol{\alpha}) \\ &= -c_1\boldsymbol{z}_1 + \boldsymbol{Y}_k(\boldsymbol{q},\dot{\boldsymbol{q}})\tilde{\boldsymbol{\theta}}_k + \hat{\boldsymbol{J}}_g(\dot{\boldsymbol{q}} - \boldsymbol{\alpha}).\end{aligned} \quad (18)$$

Selecting a Lyapunov function candidate as

$$V_3 = \frac{1}{2}\boldsymbol{z}_1^{\mathrm{T}}\boldsymbol{z}_1 + \frac{1}{2\gamma_k}\tilde{\boldsymbol{\theta}}_k^{\mathrm{T}}\tilde{\boldsymbol{\theta}}_k, \quad (19)$$

The time derivative of V_3 is derived as

$$\begin{aligned}\dot{V}_3 &= \boldsymbol{z}_1^{\mathrm{T}}\dot{\boldsymbol{z}}_1 + \frac{1}{\gamma_k}\tilde{\boldsymbol{\theta}}_k^{\mathrm{T}}\dot{\tilde{\boldsymbol{\theta}}}_k \\ &= \boldsymbol{z}_1^{\mathrm{T}}\left[-c_1\boldsymbol{z}_1 + \boldsymbol{Y}_k(\boldsymbol{q},\dot{\boldsymbol{q}})\tilde{\boldsymbol{\theta}}_k + \hat{\boldsymbol{J}}_g(\dot{\boldsymbol{q}} - \boldsymbol{\alpha})\right] \\ &\quad -\tilde{\boldsymbol{\theta}}_k^{\mathrm{T}}\left[\boldsymbol{Y}_k^{\mathrm{T}}(\boldsymbol{q},\dot{\boldsymbol{q}})\boldsymbol{z}_1 - \mu_k\hat{\boldsymbol{\theta}}_k\right] \\ &= -c_1\boldsymbol{z}_1^{\mathrm{T}}\boldsymbol{z}_1 + \boldsymbol{z}_1^{\mathrm{T}}\hat{\boldsymbol{J}}_g(\dot{\boldsymbol{q}} - \boldsymbol{\alpha}) + \mu_k\tilde{\boldsymbol{\theta}}_k^{\mathrm{T}}\hat{\boldsymbol{\theta}}_k.\end{aligned} \quad (20)$$

To avoid using the time derivative of $\boldsymbol{\alpha}$ in the following step, a first-order low-pass filter is given by

$$\gamma_\alpha\dot{\bar{\boldsymbol{\alpha}}} + \bar{\boldsymbol{\alpha}} = \boldsymbol{\alpha}, \quad \bar{\boldsymbol{\alpha}}(0) = \boldsymbol{\alpha}(0) \quad (21)$$

where γ_α is the time constant.

Step 2: Introduce another error variable

$$\boldsymbol{z}_2 = \dot{\boldsymbol{q}} - \bar{\boldsymbol{\alpha}}. \quad (22)$$

Define the filtering error as $\boldsymbol{\varepsilon}_\alpha = \bar{\boldsymbol{\alpha}} - \boldsymbol{\alpha}$, the time derivative of $\boldsymbol{\varepsilon}_\alpha$ can be expressed as

$$\dot{\boldsymbol{\varepsilon}}_\alpha = \dot{\bar{\boldsymbol{\alpha}}} - \dot{\boldsymbol{\alpha}} = -\frac{1}{\gamma_\alpha}\boldsymbol{\varepsilon}_\alpha + \boldsymbol{\phi}(\cdot), \quad (23)$$

where $\boldsymbol{\phi}(\cdot)$ is a continuous function consisting $\dot{\boldsymbol{\alpha}}$, which is bounded by a positive constant, i.e., $\|\boldsymbol{\phi}(\cdot)\| \leq \phi_M$. Using (4), the dynamics of $\boldsymbol{z}_2$ is obtained as

$$\begin{aligned}&\boldsymbol{M}_0(\boldsymbol{q})\dot{\boldsymbol{z}}_2 + \boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}})\boldsymbol{z}_2 \\ &= -\boldsymbol{M}_0(\boldsymbol{q})\dot{\bar{\boldsymbol{\alpha}}} - \boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}})\bar{\boldsymbol{\alpha}} + \boldsymbol{\tau} + \boldsymbol{\delta}.\end{aligned} \quad (24)$$

Design the control torque $\boldsymbol{\tau}$ as

$$\boldsymbol{\tau} = \boldsymbol{M}_0(\dot{\boldsymbol{q}})\dot{\bar{\boldsymbol{\alpha}}} + \boldsymbol{C}_0(\boldsymbol{q},\dot{\boldsymbol{q}})\bar{\boldsymbol{\alpha}} - \hat{\boldsymbol{J}}_g\boldsymbol{z}_1 - c_2\boldsymbol{z}_2 - \hat{\boldsymbol{\delta}}, \quad (25)$$

Selecting another positive definite function

$$V_4 = V_3 + \frac{1}{2}\boldsymbol{z}_2^{\mathrm{T}}\boldsymbol{M}_0(\boldsymbol{q})\boldsymbol{z}_2 + \frac{1}{2}\boldsymbol{\varepsilon}_\alpha^{\mathrm{T}}\boldsymbol{\varepsilon}_\alpha, \quad (26)$$

The time derivative of V_4 is calculated as

$$\begin{aligned}\dot{V}_4 &= \dot{V}_3 + \boldsymbol{z}_2^{\mathrm{T}}\boldsymbol{M}_0(\boldsymbol{q})\dot{\boldsymbol{z}}_2 + \frac{1}{2}\boldsymbol{z}_2^{\mathrm{T}}\dot{\boldsymbol{M}}_0(\boldsymbol{q})\boldsymbol{z}_2 + \boldsymbol{\varepsilon}_\alpha^{\mathrm{T}}\dot{\boldsymbol{\varepsilon}}_\alpha \\ &\equiv -c_1\boldsymbol{z}_1^{\mathrm{T}}\boldsymbol{z}_1 + \boldsymbol{z}_1^{\mathrm{T}}\hat{\boldsymbol{J}}_g(\boldsymbol{z}_2 + \boldsymbol{\varepsilon}_\alpha) + \mu_k\tilde{\boldsymbol{\theta}}_k^{\mathrm{T}}\hat{\boldsymbol{\theta}}_k + \boldsymbol{z}_2^{\mathrm{T}}\boldsymbol{M}_0(\boldsymbol{q})\dot{\boldsymbol{z}}_2 \\ &\quad + \frac{1}{2}\boldsymbol{z}_2^{\mathrm{T}}\dot{\boldsymbol{M}}_0(\boldsymbol{q})\boldsymbol{z}_2 + \boldsymbol{\varepsilon}_\alpha^{\mathrm{T}}\dot{\boldsymbol{\varepsilon}}_\alpha.\end{aligned} \quad (27)$$

Substituting (24) and (25) into (27) yields

$$\begin{aligned}\dot{V}_4 &\equiv -c_1\boldsymbol{z}_1^{\mathrm{T}}\boldsymbol{z}_1 - c_2\boldsymbol{z}_2^{\mathrm{T}}\boldsymbol{z}_2 - \frac{1}{\gamma_\alpha}\boldsymbol{\varepsilon}_\alpha^{\mathrm{T}}\boldsymbol{\varepsilon}_\alpha + \boldsymbol{z}_1^{\mathrm{T}}\hat{\boldsymbol{J}}_g\boldsymbol{\varepsilon}_\alpha \\ &\quad + \boldsymbol{z}_2^{\mathrm{T}}\tilde{\boldsymbol{\delta}} + \mu_k\tilde{\boldsymbol{\theta}}_k^{\mathrm{T}}\hat{\boldsymbol{\theta}}_k + \boldsymbol{\varepsilon}_\alpha^{\mathrm{T}}\boldsymbol{\phi}(\cdot).\end{aligned} \quad (28)$$

According to Young's inequality, one has

$$\begin{aligned}\dot{V}_4 &\leq -\left[c_1 - \frac{1}{2}\lambda_{\max}^2(\hat{\boldsymbol{J}}_g)\right]\boldsymbol{z}_1^{\mathrm{T}}\boldsymbol{z}_1 - \left(c_2 - \frac{1}{2}\right)\boldsymbol{z}_2^{\mathrm{T}}\boldsymbol{z}_2 \\ &\quad -\left(\frac{1}{\gamma_\alpha} - 1\right)\boldsymbol{\varepsilon}_\alpha^{\mathrm{T}}\boldsymbol{\varepsilon}_\alpha - \frac{\mu_k}{2}\tilde{\boldsymbol{\theta}}_k^{\mathrm{T}}\tilde{\boldsymbol{\theta}}_k + \frac{1}{2}\tilde{\boldsymbol{\delta}}^{\mathrm{T}}\tilde{\boldsymbol{\delta}} + \eta.\end{aligned} \quad (29)$$

where $\eta = 0.5\mu_k\boldsymbol{\theta}_k^{\mathrm{T}}\boldsymbol{\theta}_k + 0.5\phi_M^2$.

Theorem 2. For the DFFSM system (1) and (2), the tracking error $\boldsymbol{z}_1$ in task space is uniformly ultimately bounded with the sliding mode disturbance observer (7), the virtual control (16), the adaptive law (17), and the control law (25).

Proof. According to Theorem 1, the estimation error $\tilde{\boldsymbol{\delta}}$ will converge to zero after T_d, it is obtained that

$$\begin{aligned}\dot{V}_4 &\leq -\left[c_1 - \frac{1}{2}\lambda_{\max}^2(\hat{\boldsymbol{J}}_g)\right]\boldsymbol{z}_1^{\mathrm{T}}\boldsymbol{z}_1 - \left(c_2 - \frac{1}{2}\right)\boldsymbol{z}_2^{\mathrm{T}}\boldsymbol{z}_2 \\ &\quad -\frac{\mu_k}{2}\tilde{\boldsymbol{\theta}}_k^{\mathrm{T}}\tilde{\boldsymbol{\theta}}_k - \left(\frac{1}{\gamma_\alpha} - 1\right)\boldsymbol{\varepsilon}_\alpha^{\mathrm{T}}\boldsymbol{\varepsilon}_\alpha + \eta.\end{aligned} \quad (30)$$

By appropriate choosing gains such that $c_1 \geq 0.5\lambda_{\max}^2(\hat{\boldsymbol{J}}_g)$, $c_2 > 0.5$ and $\gamma_\alpha < 1$, Eq.(30) can be derived as

$$\dot{V}_4 \leq -\varsigma V_4 + \eta \quad (31)$$

with $\varsigma = \min\{\varsigma_0, \varsigma_1, \mu_k\gamma_k, 2\gamma_\alpha^{-1} - 2\}$, $\varsigma_0 = 2c_1 - \lambda_{\max}^2(\hat{\boldsymbol{J}}_g)$, $\varsigma_1 = \lambda_{\max}^{-1}(\boldsymbol{M}_0(\boldsymbol{q}))(2c_2 - 1)$. From (31), the uniformly ultimate boundedness of $\boldsymbol{z}_1$, $\boldsymbol{z}_2$, $\tilde{\boldsymbol{\theta}}_k$ and $\boldsymbol{\varepsilon}_\alpha$ is verified. This completes the proof.

3. Numerical Simulations

In this section, numerical simulations with a planar 3-DOF DFFSM are constructed to verify the effectiveness of the proposed control scheme. The system parameters of the DFFSM are set as: $m_0 = 100\mathrm{kg}$, $m_2^i = m_3^i = 2m_1^i = 10\mathrm{kg}$, $\boldsymbol{I}_0 = 10\mathrm{kg\cdot m^2}$, $\boldsymbol{I}_2^i = \boldsymbol{I}_3^i = 2\boldsymbol{I}_1^i = 0.1\mathrm{kg\cdot m^2}$, $l_1^i = 0.2\mathrm{m}$, $l_2^i = l_3^i = 0.75\mathrm{m}$. The perturbed parts of inertial parameters are chosen as: $\Delta m_0 = 0.1m_0$, $\Delta m_j^i = 0.1m_j^i$, $\Delta\boldsymbol{I}_0 = 0.1\boldsymbol{I}_0$, $\Delta\boldsymbol{I}_j^i = 0.1\boldsymbol{I}_j^i$ ($i = a,b$, $j = 1,2,3$). The external disturbances exerted on the DFFSM system are:

$$\boldsymbol{d} = [0.05\sin(0.05\mathrm{t}); 0.06\cos(0.05\mathrm{t}); 0.04\sin(0.05\mathrm{t})]$$

The desired trajectories of each arm are (m):

$$x_{ed}^{a} = 0.1\sin(0.1\pi t) + 0.1,\ y_{ed}^{a} = 0.1\cos(0.1\pi t) + 2.2,$$
$$x_{ed}^{b} = 0.1\sin(0.1\pi t) - 0.1,\ y_{ed}^{b} = 0.1\cos(0.1\pi t) + 2.1.$$

The control parameters of the disturbance observer (7), the virtual control (16), the adaptive law (17) and the control law (25) are chosen as: $\bar{\lambda}_1 = 0.01$, $\lambda_2 = 1.5\times10^{-3}$, $\lambda_4 = 2\lambda_3 = 2\times10^{-3}$, $\gamma_k = \mu_k = 0.01$, $\gamma_\alpha = 0.1$, $c_1 = 5$, $c_2 = 1$. Damped least square method is utilized in (16) to avoid the possible dynamic singularity, that is, $\hat{\boldsymbol{J}}_g^{-1}$ is replaced by $\hat{\boldsymbol{J}}_g^{\#} = (\lambda \boldsymbol{E} + \hat{\boldsymbol{J}}_g^{\mathrm{T}} \hat{\boldsymbol{J}}_g)^{-1} \hat{\boldsymbol{J}}_g^{\mathrm{T}}$, where $\lambda = 1\times10^{-3}$, $\boldsymbol{E}$ is the identity matrix.

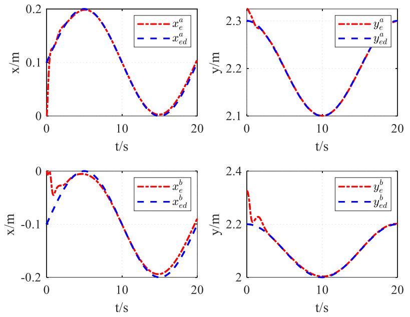


Fig.2 Time Response of $\boldsymbol{x}_e^i$ and $\boldsymbol{x}_{ed}^i$

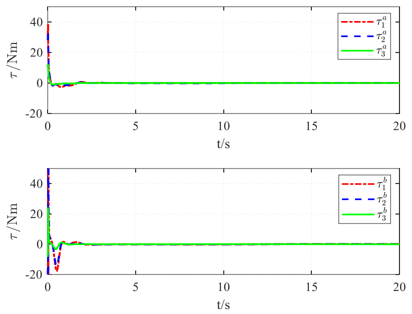


Fig.3 Time Response of $\boldsymbol{\tau}^i$

The simulation results are presented in Fig.2 and Fig.3. Fig.2 depicts the comparisons between the actual and desired trajectories of each arm. It is indicated that the actual trajectories could finally track the desired ones. The time vibrations of control input torque of each arm are shown in Fig.3. Thus, the robustness of the proposed ADSC with the sliding mode disturbance observer under disturbances and uncertainties is demonstrated.

4. Conclusions

This paper proposes a disturbance observer-based ADSC for DFFSM in task space subject to unknown disturbances, kinematic and dynamic uncertainties. It is proven that the uniformly ultimate boundedness of tracking errors is achieved with the proposed control scheme. In the future, the authors will study the coordinated control for DFFSM in on-orbit servicing.

Acknowledgements

This work was supported in part by the NSFC (62227810, 62133001) and the National Basic Research Program of China (973 Program: 2012CB821200, 2012CB821201).

References

1. L. Yan, H. Yuan, W. Xu, et al, Generalized relative Jacobian matrix of space robot for dual-arm coordinated capture, Journal of Guidance, Control, and Dynamics, Vol. 41(5), 2018, pp.1202-1208.
2. L. Shi, S. Kayastha, J. Katupitiya, Robust coordinated control of a dual-arm space robot, Acta Astronautica, Vol. 138, 2017, pp.475-489.
3. L. Shi, H. Jayakody, J. Katupitiya, et al, Coordinated control of a dual-arm space robot: Novel models and simulations for robotic control methods, IEEE Robotics & Automation Magazine, Vol. 25(4), 2018, pp.86-95.
4. C. Yang, Y. Jiang, J. Na, et al, Finite-time convergence adaptive fuzzy control for dual-arm robot with unknown kinematics and dynamics, IEEE Transactions on Fuzzy Systems, Vol. 27(3), 2019, pp.574-588.
5. J. Cheng, L. Chen, The fuzzy neural network control scheme with H∞ tracking characteristic of space robot system with dual-arm after capturing a spin spacecraft, IEEE/CAA Journal of Automatica Sinica, Vol. 7(5), 2020, pp.1417-1424.
6. Y. Liu, W. Yan, T. Zhang, et al, Trajectory tracking for a dual-arm free-floating space robot with a class of general nonsingular predefined-time terminal sliding mode, IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 52(5), 2021, pp.3273-3286.

Authors Introduction

Ms. Qian Sun

She received the B.S. degree in Automation from Chongqing University, Chongqing, China, in 2020. She is cur-rently working toward the Ph.D. degree in control theory and control engineer-ing from Beihang University, Beijing, China. Her research interests include coordinated control of on-orbit servicing.

Prof. Yingmin Jia

He received B.S. degree from Shandong University, Jinan, China, in 1982, and the M.S. and Ph.D. degrees from Beihang University, Beijing, China, in 1990 and 1993, respectively.
He is currently a professor with the Seventh Research Division and the director of the Center for Information and Control at Beihang University. His research interests include robust control, robotic systems, spacecraft coordination and on-orbit servicing.

Manipulability Optimization for Redundant Dual-Arm Robots at the Acceleration Level

Yang Zhang
School of Energy and Power Engineering, Beihang University (BUAA), Beijing 100191, China.

Yingmin Jia*
The Seventh Research Division and the Center for Information and Control, School of Automation Science and Electrical Engineering, Beihang University (BUAA), Beijing, 100191, China
E-mail: zhyang2019@foxmail.com, ymjia@buaa.edu.cn

Abstract

Existing manipulability optimization schemes typically solve at the velocity level, which cannot consider joint acceleration limits and are unsuitable for torque control of robotic arms. Therefore, this paper constructs a cost function that considers both joint torque constraints and manipulability optimization of the manipulator, and equivalently transforms it into a convex quadratic function. The proposed scheme addresses the non-convexity issue of manipulability with respect to the robotic arm joint acceleration and the inversion problem of the generalized Jacobian matrix. Simulation results show that the proposed method can maximize the manipulability of redundant dual-arm robots at the acceleration level, verifying the effectiveness of the scheme.

Keywords: Manipulability optimization, Dual-arm robot, Quadratic programming, Redundancy resolution

1. Introduction

In the context of rapid advancements in robotics, dual-arm robots have emerged as a significant research focus in fields such as industrial automation, surgical operations, and service robotics due to their flexibility and capability in handling complex tasks [1], [2], [3]. Dual-arm robots are capable of completing intricate operational tasks through coordinated movements, such as assembly, transportation, and precision handling, which are crucial for enhancing production efficiency and quality [4]. However, the key to achieving these complex tasks lies in improving the robot's manipulability, which refers to its motion capability and responsiveness in the task space [5].

Manipulability was initially proposed and studied to address singularity issues in manipulator motion planning and control. High manipulability allows a manipulator to achieve the desired end-effector motion with smaller joint velocities, making it an essential performance indicator for evaluating manipulator joint configurations and trajectory planning algorithms. Current research predominantly focuses on optimizing manipulability for single-arm systems, mainly relying on velocity-level analytical methods [6], [7], [8]. These methods are typically based on the characteristics of the velocity Jacobian matrix, such as condition numbers and singular values, to enhance the movement capabilities of robotic end effectors. However, velocity-level optimization has significant limitations: firstly, it fails to adequately consider the dynamic characteristics of the robotic system, especially when inertia and torques are involved; secondly, for the collaborative operations of dual-arm systems, velocity-level optimization struggles to address the complex coupling relationships and synchronized coordination between the two arms. These shortcomings restrict the applicability and efficiency of dual-arm robots in complex tasks.

To overcome these limitations, this paper proposes an acceleration-level optimization method for the manipulability of dual-arm robots. By introducing the acceleration Jacobian matrix and a detailed dynamic model, we can more comprehensively analyze the dynamic performance of dual-arm robots, particularly the impact of inertial coupling and torque distribution on system operational capability. This approach not only considers the dynamic coupling characteristics in dual-arm collaboration but also enhances the robot's response speed and stability in dynamic environments by optimizing acceleration-level indicators.

2. System Model

This section will provide a detailed introduction to the mathematical model of the dual-arm robot. The dual-arm system under study is shown in Fig. 1, where both the left and right arms of the robot are redundant manipulators with 7 degrees of freedom.

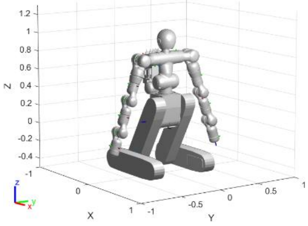


Fig. 1 The dual-arm robot platform

*Corresponding author.

2.1. *Kinematic model of the manipulator*

The mapping relationship between the position and orientation vector $\chi \in \mathbb{R}^6$ of the end-effector of the manipulator in the task space and the joint space variables $\Theta \in \mathbb{R}^7$ is as follows:

$$\chi = f(\Theta) \tag{1}$$

where $f(\cdot): \mathbb{R}^7 \to \mathbb{R}^6$ represents a nonlinear forward kinematics function. From the above kinematic equations, we can derive the differential-level kinematic equations.

$$\dot{\chi} = J\dot{\Theta} \tag{2}$$

$$\ddot{\chi} = \dot{J}\dot{\Theta} + J\ddot{\Theta} \tag{3}$$

where, $\dot{\chi} \in \mathbb{R}^6$ and $\ddot{\chi} \in \mathbb{R}^6$ are the generalized velocity and acceleration of the end-effector, $\dot{\Theta} \in \mathbb{R}^7$ and $\ddot{\Theta} \in \mathbb{R}^7$ are the joint angular velocity and angular acceleration, J is the corresponding Jacobian matrix, and $\dot{J} = dJ/dt$.

2.2. *Task-Oriented Coordination Operation equations*

During the coordinated operation of dual-arm robots, it is essential to maintain certain kinematic constraints between the robotic arms. Task-oriented coordination operation equations define the workspace and interaction dynamics in a manner that optimizes the performance of the cooperative task. First, a set of geometrically clear manipulation variables are defined to describe the coordinated operation tasks of the dual arms, namely the absolute motion variable χ_a, which describes the motion state of the manipulated object, and the relative motion variable χ_r, which describes the motion states of the two robotic arms.

By differentiating the aforementioned variables, one can obtain the absolute and relative motion variables in terms of velocity and acceleration, specifically described as:

$$\begin{cases} \dot{\chi}_a = (\dot{\chi}_R + \dot{\chi}_L)/2 \\ \dot{\chi}_r = \dot{\chi}_R - \dot{\chi}_L \end{cases} \tag{4}$$

$$\begin{cases} \ddot{\chi}_a = (\ddot{\chi}_R + \ddot{\chi}_L)/2 \\ \ddot{\chi}_r = \ddot{\chi}_R - \ddot{\chi}_L \end{cases} \tag{5}$$

Define $\dot{\chi}_G = [\dot{\chi}_a^T, \dot{\chi}_r^T]^T \in \mathbb{R}^{12}$ as the generalized variable describing the state of the manipulated object. In conjunction with Eq. (2), the task-oriented coordination operation equations for the velocity level can be expressed as follows:

$$\dot{\chi}_G = \begin{bmatrix} J_a \\ J_r \end{bmatrix} \begin{bmatrix} \dot{\Theta}_R \\ \dot{\Theta}_L \end{bmatrix} = J_G \dot{\Theta}_G \tag{6}$$

where, $J_G \in \mathbb{R}^{12\times 14}$ is the generalized Jacobian matrix of the coordination operation equation, which represents the mapping relationship between the generalized variables $\dot{\chi}_G$ and the joint space variables , and $\dot{\Theta}_G \in [\dot{\Theta}_R^T, \dot{\Theta}_L^T]^T$, $J_a \triangleq [J_R/2 \quad J_L/2] \in R^{6\times 14}$, $J_r \triangleq [-J_R \quad J_L] \in R^{6\times 14}$.

Differentiating Eq. (6), one can obtain the coordination operation equation at the acceleration level, specifically described as follows:

$$\ddot{\chi}_G = \dot{J}_G \dot{\Theta}_G + J_G \ddot{\Theta}_G \tag{7}$$

where , $\dot{J}_G$ is the is the time derivative of the generalized Jacobian matrix. Eqs. (6) and (7) describe the task-oriented coordination operation equations for dual-arm robots. Since the manipulator has redundant degrees of freedom, meaning the generalized Jacobian matrix is not a square matrix, the solution to this equation is infinite.

For the aforementioned kinematic equations, the solution can be found using the pseudoinverse method, which is specifically described as follows:

$$\dot{\Theta}_G = J_G^\dagger \dot{\chi}_G + (I - J_G^\dagger J_G)\Pi_1 \tag{8}$$

$$\ddot{\Theta}_G = J_G^\dagger (\ddot{\chi}_G - \dot{J}_G \dot{\Theta}_G) + (I - J_G^\dagger J_G)\Pi_2 \tag{9}$$

where $J_G^\dagger \in \mathbb{R}^{14\times 12}$ is the generalized inverse of J_G, assuming that J_G is full row rank, $J_G^\dagger = J_G^T (J_G J_G^T)^{-1}$. Π_1 and Π_2 are arbitrary vectors representing the gradients of certain selected performance metrics.

3. Acceleration-level Manipulability Optimization

In this section, a scheme for optimizing manipulability at the acceleration level is developed to address the redundancy resolution of dual-arm robots.

3.1. *Optimization metrics*

Manipulability is a crucial performance metric for assessing the configuration of robotic arm joints and trajectory planning algorithms. According to [6], the manipulability of a dual-arm system is defined as:

$$M = \sqrt{\det(J_G J_G^T)} \tag{10}$$

The larger of M, the stronger the manipulability of the robotic arm. When $M = 0$, the robotic arm will be in a singular configuration. The derivative of maneuverability M with respect to time is

$$\begin{aligned} \frac{dM}{dt} &= \sqrt{\det(J_G J_G^T)} Tr\left((J_G J_G^T)^{-1} (\dot{J}_G J_G^T + J_G \dot{J}_G^T)\right) \\ &= \sqrt{\det(J_G J_G^T)} Tr(\dot{J}_G J_G^\dagger) \end{aligned} \tag{11}$$

where, $Tr(\cdot)$ represents the trace of the matrix. The derivative of the generalized Jacobian matrix can be expressed as

$$\dot{J}_G = \sum_{i=1}^{N} \frac{\partial J_G}{\partial \Theta_{Gi}} \dot{\Theta}_{Gi} \tag{12}$$

From this point, we can obtain the optimization vector in terms of velocity .

$$\Pi_1 = -\frac{dM}{dt} \tag{13}$$

The vector representation for optimization at the acceleration level is given by

$$\Pi_2 = \frac{d\Pi_1}{dt} = -\frac{d}{dt}\left(\sqrt{\det(J_G J_G^T)}Tr(\dot{J}_G J_G^{\dagger})\right)$$
$$= -\frac{d\sqrt{\det(J_G J_G^T)}}{dt}Tr(\dot{J}_G J_G^{\dagger}) - \sqrt{\det(J_G J_G^T)}\frac{d}{dt}\left(Tr(\dot{J}_G J_G^{\dagger})\right)$$
(14)

with which, $d\left(Tr(\dot{J}_G J_G^{\dagger})\right)/dt$ is expressed as

$$\frac{d}{dt}\left(Tr(\dot{J}_G J_G^{\dagger})\right) = Tr\left(\frac{d}{dt}\left(\dot{J}_G\right)J_G^{\dagger} + \dot{J}_G \dot{J}_G^{\dagger}\right) \tag{15}$$

As to $\dot{J}_G^{\dagger}$, which can be calculated as follows:

$$\dot{J}_G^{\dagger} = (I - J_G^{\dagger}J_G)\dot{J}_G^T(J_G J_G^T)^{-1} - J_G^{\dagger}\dot{J}_G J_G^{\dagger} \tag{16}$$

Additionally, to reduce control efforts, we further considered the corresponding optimization criteria, that is Minimum acceleration norm

$$\min \quad (\ddot{\Theta}_G + k\dot{\Theta}_G)^T(\ddot{\Theta}_G + k\dot{\Theta}_G) \tag{17}$$

At this point, the optimization criteria for the dual-arm robot in acceleration level can be expressed as

$$\min_{\dot{\Theta}_G,\ddot{\Theta}_G} \quad \Upsilon_1 = \frac{1}{2}\ddot{\Theta}_G^T\ddot{\Theta}_G - (a_1\dot{\Theta}_G^T + a_2\Pi_2^T)\ddot{\Theta}_G \tag{18}$$

where, the design parameters $a_1 \in (0,1)$ and $a_2 \in (0,1)$ are used to adjust the weights of each criterion, respectively.

3.2. *Joint physical constraints*

In engineering applications, virtually all robotic arms have physical limitations on joint angles and joint angular velocities. Therefore, it is essential to consider the actual range of angles and angular velocities that can be achieved by each joint of the robotic arm. This constraint is specifically described as:

$$\Theta_G^{\min} \le \Theta_G \le \Theta_G^{\max} \tag{19}$$
$$\dot{\Theta}_G^{\min} \le \dot{\Theta}_G \le \dot{\Theta}_G^{\max} \tag{20}$$
$$\ddot{\Theta}_G^{\min} \le \ddot{\Theta}_G \le \ddot{\Theta}_G^{\max} \tag{21}$$

The aforementioned optimization scheme is based on acceleration level solution; therefore, the joint angle and angular velocity constraints need to be converted into descriptions based on joint angular acceleration. According to [8], the physical limits can be expressed as:

$$\Theta_G^- \le \ddot{\Theta}_G \le \Theta_G^+ \tag{22}$$

The boundary constraints Θ_G^- and Θ_G^+ are defined as follows:

$$\begin{cases} \Theta_G^- = \max\left\{\eta_1\eta_2(\Theta_G^{\min} - \Theta_G), \eta_1(\dot{\Theta}_G^{\min} - \dot{\Theta}_G), \ddot{\Theta}_G^{\min}\right\} \\ \Theta_G^+ = \min\left\{\eta_1\eta_2(\Theta_G^{\max} - \Theta_G), \eta_1(\dot{\Theta}_G^{\max} - \dot{\Theta}_G), \ddot{\Theta}_G^{\max}\right\} \end{cases} \tag{23}$$

where, $\eta_1 \ge \eta_2 > 0$ are design parameters used to adjust the feasible domains of joint acceleration and velocity, respectively.

3.3. *Acceleration-level optimization scheme*

By incorporating the end-effector trajectory tracking constraints Eq. (7) and joint acceleration constraints Eq. (22) into the optimization criterion Eq. (18), the redundancy resolution problem of a dual-arm robot can be formulated as an optimization scheme in the following form.

$$\begin{aligned} \min_{\dot{\Theta}_G,\ddot{\Theta}_G} \quad & \Upsilon_1 = \frac{1}{2}\left\|\ddot{\Theta}_G\right\|_2^2 - (a_1\dot{\Theta}_G^T + a_2\Pi_2^T)\ddot{\Theta}_G \\ s.t. \quad & \ddot{\chi}_G - J_G\ddot{\Theta}_G - \dot{J}_G\dot{\Theta}_G = 0 \\ & \Theta_G^- \le \ddot{\Theta}_G \le \Theta_G^+ \end{aligned} \tag{24}$$

4. Simulation Results and Discussion

In order to evaluate the performance of the proposed manipulability optimization scheme, we applied the scheme to the dual-arm robot as shown in Fig. 1.
Simulation experiments were conducted in the MATLAB software environment. To demonstrate the effectiveness of the scheme, we designed the following working scenario: Both arms collaboratively move a circular component, rotating it 180 degrees in the world coordinate system. The simulation results are presented in Figs. 2-4.

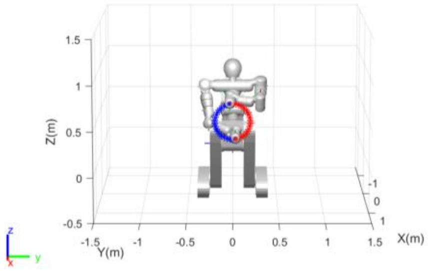


Fig. 2. The simulation process of the operational task.

First, Fig. 2 displays the motion trajectory of the robotic arm's end-effector from a front view.

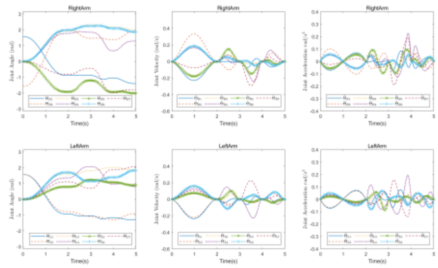


Fig. 3 Joint angles, velocities, and accelerations

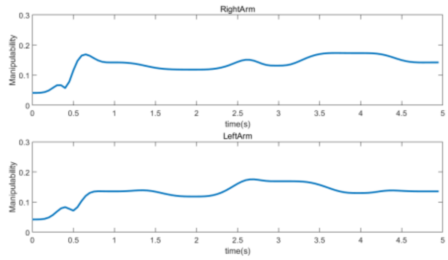


Fig. 4. Manipulability measures of the dual-arm robot

Fig. 3 shows that the joint angles, angular velocities, and angular acceleration of the robotic arm consistently remain within the constraint limits, demonstrating the effectiveness of boundary constraints. It can be observed that the coordinated movement of the dual arms allows the posture tracking task to be successfully accomplished.

Moreover, the joint trajectory planning results show smooth and continuous changes over time, meeting the needs of practical applications. Fig. 4 presents the variability in manipulability of each operational arm during the process.

From the above simulation results, it can be clearly seen that the manipulability optimization scheme proposed in this paper effectively addresses the problem of redundancy resolution at the acceleration level for the dual-arm robot.

5. Conclusion

The proposed scheme effectively addresses the limitations of previous methods that optimize at the velocity level without considering joint acceleration limits. It does so by constructing a cost function that simultaneously considers joint torque constraints and manipulability optimization. This method successfully tackles the issue of non-convexity between manipulator joint angle acceleration and manipulability by converting the non-convex problem into an equivalent convex quadratic function. This method enables manipulability optimization at the acceleration level, thereby enhancing the algorithm's applicability and robustness. Simulation results demonstrate that the proposed method can maximize the manipulability of redundant dual-arm robots at the acceleration level, verifying its effectiveness. This offers a novel solution for robotic control tasks demanding high precision and efficiency in practical applications.

Acknowledgements

This work was supported in part by the NSFC under Grant 62227810, and Grant 62133001, and in part by the National Basic Research Program of China (973 Program) under Grant 2012CB821200 and Grant 2012CB821201.

References

1. A. Dietrich, T. Wimbock, A. Albu-Schaffer, and G. Hirzinger, "Integration of Reactive, Torque-Based Self-Collision Avoidance Into a Task Hierarchy," IEEE Transactions on Robotics, vol. 28, pp. 1278-1293, 2012.
2. M. Garabini, D. Caporale, V. Tincani, A. Palleschi, C. Gabellieri, M. Gugliotta, A. Settimi, M. G. Catalano, G. Grioli, and L. Pallottino, "WRAPP-up: A Dual-Arm Robot for Intralogistics," IEEE Robotics & Automation Magazine, vol. 28, pp. 50-66, 2021.
3. Z. Sun, H. Yang, Y. Ma, X. Wang, Y. Mo, H. Li, and Z. Jiang, "BIT-DMR: A Humanoid Dual-Arm Mobile Robot for Complex Rescue Operations," IEEE Robotics and Automation Letters, vol. 7, pp. 802-809, 2022.
4. Z. Zhang, Y. Lin, S. Li, Y. Li, Z. Yu, and Y. Luo, "Tricriteria Optimization-Coordination Motion of Dual-Redundant-Robot Manipulators for Complex Path Planning," IEEE Transactions on Control Systems Technology, vol. 26, pp. 1345-1357, 2018.
5. Z. Jia, S. Chen, Z. Zhang, N. Zhong, P. Zhang, X. Qu, J. Xie, and F. Ouyang, "Tri-Criteria Optimization Motion Planning at Acceleration-Level of Dual Redundant Manipulators," Robotica, vol. 38, pp. 983-999, 2020-1-1 2020.
6. Y. Zhang, X. Yan, D. Chen, D. Guo, and W. Li, "QP-based refined manipulability-maximizing scheme for coordinated motion planning and control of physically constrained wheeled mobile redundant manipulators," Nonlinear Dynamics, vol. 85, pp. 245-261, 2016.
7. L. Jin, S. Li, H. M. La, and X. Luo, "Manipulability Optimization of Redundant Manipulators Using Dynamic Neural Networks," IEEE Transactions on Industrial Electronics, vol. 64, pp. 4710-4720, 2017.
8. X. Lu and Y. Jia, "Trajectory Planning of Free-Floating Space Manipulators With Spacecraft Attitude Stabilization and Manipulability Optimization," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 51, pp. 7346-7362, 2021.

Authors Introduction

Mr. Yang Zhang

He received the B.S degree in aircraft design and engineering from North-western Polytechnical University, Xian, China, in 2016, and the M.S degree in Aerospace Engineering from National University of Defense Technology, Changsha, China, in 2019. He is currently working toward a Ph.D. with the School of Energy and Power Engineering, Beihang University. His main research focusses on motion planning and control of collaborative robot, robust and nonlinear control.

Prof. Yingmin Jia

He received B.S. degree from Shandong University, Jinan, China, in 1982, and the M.S. and Ph.D. degrees from Beihang University, Beijing, China, in 1990 and 1993, respectively.

He is currently a professor with the Seventh Research Division and the director of the Center for Information and Control at Beihang University. His research interests include robust control, robotic systems, spacecraft coordination and on-orbit servicing.

Flocking Control for Multiple Convex Polygonal Agents with Obstacle Avoidance

Yaxin Li
The Seventh Research Division and the Center for Information and Control, School of Automation Science and Electrical Engineering, Beihang University (BUAA), Beijing, 100191, China

Yingmin Jia
The Seventh Research Division and the Center for Information and Control, School of Automation Science and Electrical Engineering, Beihang University (BUAA), Beijing, 100191, China
Email: lyx2018@buaa.edu.cn, ymjia@buaa.edu.cn

Abstract

This paper addresses the distributed flocking control for second-order convex polygonal multiagent systems with obstacle avoidance. Typically, agent shapes are reduced to mere points or circles in existing research. Overlooking the actual geometries of entities such as autonomous ships and vehicles may lead to suboptimal utilization of spatial resources. To rectify this, the paper introduces an approach for calculating the relative distance between two convex polygonal agents. A potential function, which balances attractive and repulsive forces, is designed based on these calculated distances. A flocking trajectory guides collective motion, and a separate obstacle-avoidance trajectory is activated when the agent's proximity to an obstacle falls below a specified safety threshold. The proposed control strategy integrates potential function and reference trajectory is designed to achieve flocking behavior and obstacle avoidance. Stability analysis proves the effectiveness of the algorithm.

Keywords: Flocking control, Multi-agent systems, Convex polygonal, Obstacle avoidance

1. Introduction

Natural phenomena such as the collective behavior of birds, fish, and ants have attracted researchers from various fields [1]. Inspired by these, Minsky introduced the concept of an agent. Advances in artificial intelligence, big data, and networking have led to the development of multi-agent systems (MASs) with autonomous and intelligent capabilities. MASs have been successfully applied in areas such as autonomous driving, drones, robotics, and sensor networks [2]. They excel at distributed cooperative control, tackling complex tasks that single agents cannot. Flocking, a key aspect of collective behavior, is a major research focus within MAS cooperative control [3].

In 1987, Reynolds introduced the Boid model [4], which describes the behavior of groups with rules for cohesion, alignment, and separation. Vicsek et al. refined this in 1995 [5] with a distributed alignment rule. Tanner et al. addressed control of varying topology in MASs [6], Subsequent flocking research addressed real-world challenges such as obstacle avoidance and limited communication. In particular, Olfati-Saber developed innovative control algorithms for flocking in complex environments, and Jafet addressed communication and input constraints [7].

Many studies simplify agents as particles or circles, ignoring practical shapes such as rectangles or polygons, which are more space efficient, especially in confined spaces such as traffic lanes (Fig. 1). This can lead to inaccurate distance estimates and suboptimal collision avoidance. Early studies [8] modeled agents as ellipses to solve these problems, but issues such as discontinuous control signals limited their applicability. Later studies [9] modeled agents as rectangles, which is good for calculating distances and interactions between agents, reducing the risk of collision and improving safety. However, in practical applications, the shape of the agents may not be monolithic. Therefore, considering the shape of the agents only as a rectangle also has limitations.

Fig.1 Advantages of considering the realities of agents for space utilization.

Motivated by the above discussion, the main contributions of this paper can be summarized as follows: (i) Adoption of convex polygonal shapes for agents, addressing spatial inefficiency in confined spaces and improving motion efficiency. (ii) Introduction of a continuous control input method for calculating relative distances between convex polygonal agents, applicable to shapes with varying numbers of vertices for greater flexibility. (iii) Integrate obstacle avoidance into the flocking process, ensuring that agents follow a reference trajectory and employ avoidance strategies when approaching obstacles, improving safety in complex environments.

The rest of this paper is organized as follows. Section 2 gives some preliminaries. Section 3 describes the problem we studied and presents the distributed flocking algorithm. Simulations are carried out in section 4. Finally, the conclusion is given in Section 5.

2. Preliminaries

2.1. *Convex Polygonal Agent Model*

Suppose that agent i is enclosed by a polygon P_i. Let's define agent i by the vector $f_i(t)=[p_i^T(t),\theta_i(t)]^T$, in which $p_i(t)=[p_{ix}(t),p_{iy}(t)]^T$ indicates the coordinates of the center O_i, and $\theta_i(t)$ is the orientation angle of the global coordinate system OXY. The coordinates of the n-th vertex of agent i are given by $\left\{p_i^k \mid 1,2,\cdots,\Lambda_i\right\}$ in the global frame OXY, and by $\left\{\tilde{p}_i^k \mid 1,2,\cdots,\Lambda_i\right\}$ in the local frame $O_iX_iY_i$, as illustrated in Fig 2.

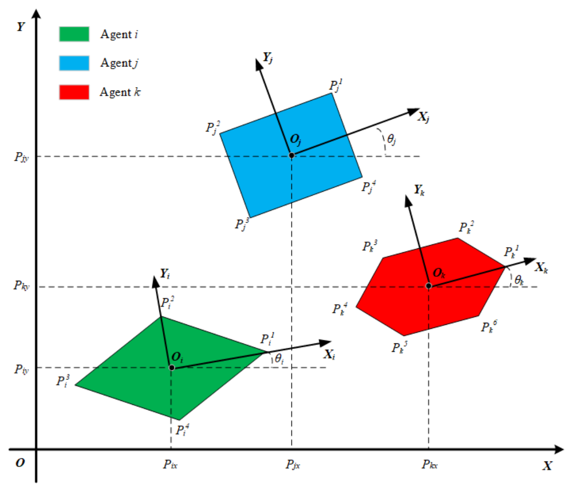


Fig.2 Convex polygonal agents in global-fixed frame and local-fixed frame

The relationship between p_i^k and $\tilde{p}_i^k$ is given as $p_i^k(t)=p_i(t)+R_i(t)\tilde{p}_i^k(t)$, $R_i(t)$ is the rotation matrix $R_i(t)=\begin{bmatrix}\cos(\theta_i(t)) & -\sin(\theta_i(t))\\ \sin(\theta_i(t)) & \cos(\theta_i(t))\end{bmatrix}$.

Agent i is equipped with a circular communication zone, having a radius of r_i centered at O_i. It can transmit its data and receive information from neighboring agents within this range, including position, direction, and speed. The communication radii for agents i and j are as follows:

$$\min\left(r_i,r_j\right)>\max_{1\le k\le\Lambda_i}\left(d\left(O_i,p_i^k\right)\right)+\max_{1\le k\le\Lambda_j}\left(d\left(O_j,p_j^k\right)\right)$$

3. Main results

3.1. *Problem formulation*

Consider a second-order multi-agent system consisting of N convex polygonal agents. We assume that each agent i has the following dynamics:

$$\begin{cases}\dot{f}_i=v_i\\ \dot{v}_i=u_i\end{cases}\tag{1}$$

where $v_i(t)=[v_{ix}(t),v_{iy}(t),\omega_i(t)]^T$ denotes the linear velocity and angular velocity of agent i, respectively. $u_i(t)=[u_{ix}(t),u_{iy}(t),u_{i\theta}(t)]^T$ denotes the control input.

The objective of this paper is to develop a distributed flocking control strategy for the multi-agent system (1) ensuring that all agents flock together with constrained control inputs and obstacles. To meet these goals, we present the following definition.

Definition 1. The multi-agent system (1) is said to achieve flocking if and only if for all agents i, the following conditions are satisfied:

(i) (Separation) The relative distance between any two agents i and j is greater than 0,

(ii) (Cohesion) The relative distance between any two agents i and j is bounded, and the error between the actual and reference trajectories is bounded, $f_r(t)=[p_r(t),\theta_r(t)]^T$, is the motion reference trajectory, $p_r(t)=[p_{rx}(t),p_{ry}(t)]$ and $\theta_r(t)$ denote the desired position the desired heading angle, respectively.

(iii) (Alignment) The velocity direction of all agents is the same.

3.2. *Control design*

Let us define a set of all successive vertices of the agent i: $S_i=\{(1,2),(2,3),\ldots,(\Lambda_i,1)\}$.

The distance between agent i and agent j is defined as:

$$\delta_{ij}=\left(\prod_{a=1}^{\Lambda_i}\prod_{(b,c)\in S_j}\delta_{ij}^{abc}\right)\left(\prod_{a=1}^{\Lambda_i}\prod_{(b,c)\in S_i}\delta_{ji}^{abc}\right)$$
$$\delta_{ij}^{abc}=\sigma\left(d_{ij}^{abc}\right)\tag{2}$$
$$d_{ij}^{abc}=\left\|p_i^a-p_j^b\right\|_2+\left\|p_i^a-p_j^c\right\|_2-\left\|p_j^b-p_j^c\right\|_2$$

Remark 1. From equation (2), it follows that δ_{ij} equals 0 if any δ_{ij}^{abc} is 0, indicating a vertex of agent i lies on the edge of agent j. Consequently, if δ_{ij} remains greater than 0 for all $t\ge 0$, it implies no collisions between agents.

To maintain separation and cohesion in flocks, we use a potential function. This function has a dual role: it generates a repulsive force to prevent collisions when agents are close together, and an attractive force to pull agents closer together when they are too far apart.

The potential function between two agents can be written as

$$\phi_{ij} = \varepsilon_1 Rp(\delta_{ij} \mid \delta_{ijD}, \lambda) + \varepsilon_2 Ar(\delta_{ij} \mid \delta_{ijD}, \delta_{ijM}) \tag{3}$$

where ε_1, ε_2 are positive real numbers, δ_{ijD} is a desirable relative distance, and δ_{ijM} is a further relative distance. $Rp(\delta_{ij} \mid \delta_{ijD}, \lambda)$ and $Ar(\delta_{ij} \mid \delta_{ijD}, \delta_{ijM})$ represents the repulsion and attraction, respectively. Similarly the potential function between the agent and the obstacle can be set as

$$\phi_{ik} = \varepsilon_3 Rp(\delta_{ik} \mid \delta_{ikD}, \lambda) + \varepsilon_4 Ar(\delta_{ik} \mid \delta_{ikD}, \delta_{ikM}) \tag{4}$$

We choose the following repulsive function and attraction function.

$$Rp(\delta_{ij} \mid \delta_{ijD}, \lambda) = \begin{cases} \dfrac{\lambda(\delta_{ijD} - \delta_{ij})^k}{\lambda\delta_{ij} + \delta_{ijD}{}^k}, & if \;\; \delta_{ij} \in \left[0, \delta_{ijD}\right] \\ 0, & otherwise \end{cases}$$

$$Ar(\delta_{ij} \mid \delta_{ijD}, \delta_{ijM}) = \begin{cases} 0, & if \;\; \delta_{ij} \in \left[0, \delta_{ijD}\right] \\ \ln \dfrac{\delta_{ijM}{}^2}{\delta_{ijM}{}^2 - \delta_{ij}{}^2}, & if \;\; \delta_{ij} \in \left[\delta_{ijD}, \delta_{ijM}\right] \\ 1, & if \;\; \delta_{ij} \in \left[\delta_{ijM}, \infty\right) \end{cases}$$

To obtain the control law, we have first to derive the partial derivatives of the potential function.

$$\begin{aligned} \dot{\phi}_{ij} &= E_{ij}\left(\dot{\theta}_i - \dot{\theta}_j\right) + F_{ij}\left(\dot{p}_i - \dot{p}_r - \tilde{R}(p_i - p_r)\dot{\theta}_i\right) + \\ &\quad F_{ji}\left(\dot{p}_j - \dot{p}_r - \tilde{R}(p_j - p_r)\dot{\theta}_j\right) \\ &= E_{ij}\left(\dot{\theta}_i - \dot{\theta}_r\right) + F_{ij}\left(\dot{p}_i - \dot{p}_r - \tilde{R}(p_i - p_r)\dot{\theta}_i\right) \\ &\quad - E_{ij}\left(\dot{\theta}_j - \dot{\theta}_r\right) + F_{ji}\left(\dot{p}_j - \dot{p}_r - \tilde{R}(p_j - p_r)\dot{\theta}_j\right) \end{aligned} \tag{5}$$

$$\begin{aligned} \dot{\phi}_{ik} &= \Big[E_{ik}\left(\dot{\theta}_i - \dot{\theta}_r\right) + F_{ik}\left(\dot{p}_i - \dot{p}_r - \tilde{R}(p_i - p_r)\dot{\theta}_i\right) \\ &\quad - E_{ik}\left(\dot{\theta}_k - \dot{\theta}_r\right) + F_{ki}\left(\dot{p}_k - \dot{p}_r - \tilde{R}(p_k - p_r)\dot{\theta}_k\right)\Big] \end{aligned} \tag{6}$$

where $E_{ij} = \dfrac{\partial \phi_{ij}}{\partial \delta_{ij}}\left[\left(A_{ij}\tilde{R}(p_i - P_r) + B_{ij}\right) - \left(A_{ji}\tilde{R}(p_j - p_r) + B_{ji}\right)\right]$

$$F_{ij} = \frac{\partial \phi_{ij}}{\partial \delta_{ij}}\left(A_{ij} - A_{ji}\right)$$

$$A_{ij} = \sum_{a=1}^{\Lambda_i} \sum_{(b,c)\in S_j} \frac{\partial \delta_{ij}}{\partial \delta_{ij}^{abc}} \frac{\partial \delta_{ij}^{abc}}{\partial d_{ij}^{abc}} M_{ij}^{abc}$$

$$B_{ij} = \sum_{a=1}^{\Lambda_i} \sum_{(b,c)\in S_j} \frac{\partial \delta_{ij}}{\partial \delta_{ij}^{abc}} \frac{\partial \delta_{ij}^{abc}}{\partial d_{ij}^{abc}} N_{ij}^{abc}$$

we take the following Lyapunov function

$$V = \Phi_1 + \Phi_2 + V_1(t) + V_2(t) \tag{7}$$

where

$$\Phi_1 = \sum_{i=1}^{N\text{-}1} \sum_{j=i+1}^{N} \phi_{ij}$$

$$\Phi_2 = \sum_{i=1}^{N} \sum_{k=1}^{M} \phi_{ik}$$

$$V_1(t) = \sum_{i=1}^{N}\left(\varepsilon_{i\theta}\left(\sqrt{(\theta_i - \theta_r)^2 + \eta_\theta{}^2} - \eta_\theta\right) + \varepsilon_{ip}\left(\sqrt{\|p_i - p_r\|_2^2 + \eta_p^2} - \eta_p\right)\right)$$

$$V_2(t) = \frac{1}{2}\sum_{i=1}^{N}\left(\left(\omega_i - \dot{\theta}_r\right)^2 + \left(v_{ix} - \tilde{v}_{ix}\right)^2 + \left(v_{iy} - \tilde{v}_{iy}\right)^2\right)$$

Then the control law of the system (1) is designed as

$$\begin{cases} u_{i\theta} = -\Delta_{i\theta} - \mu_{i\theta}\sigma\left(\omega_i - \dot{\theta}_r\right) \\ u_{ix} = -\Delta_{ix} - \mu_{ix}\sigma\left(v_{ix} - \tilde{v}_{ix}\right) \\ u_{iy} = -\Delta_{iy} - \mu_{iy}\sigma\left(v_{iy} - \tilde{v}_{iy}\right) \end{cases} \tag{8}$$

where

$$\Delta_{i\theta} = D_i + C_{ik} - \ddot{\theta}_r + \frac{\varepsilon_{i\theta}(\theta_i - \theta_r)}{\sqrt{(\theta_i - \theta_r)^2 + \eta_\theta^2}}$$

$$\Delta_{ix} = C_{ix} + D_{ikx} - \ddot{p}_{rx} + \frac{\varepsilon_{ip}(p_{ix} - p_{rx})}{\sqrt{\|p_i - p_r\|_2^2 + \eta_p^2}} + \left(v_{iy} - \dot{p}_{ry}\right)\omega_i + \left(p_{iy} - p_{ry}\right)u_{i\theta}$$

$$\Delta_{iy} = C_{iy} + C_{iky} - \ddot{p}_{ry} + \frac{\varepsilon_{ip}(p_{iy} - p_{ry})}{\sqrt{\|p_i - p_r\|_2^2 + \eta_p^2}} - \left(v_{ix} - \dot{p}_{rx}\right)\omega_i - \left(p_{ix} - p_{rx}\right)u_{i\theta}$$

$$\tilde{v}_{ix} = \dot{p}_{rx} - \left(p_{iy} - p_{ry}\right)\omega_i$$

$$\tilde{v}_{iy} = \dot{p}_{ry} + \left(p_{ix} - p_{rx}\right)\omega_i$$

Theorem 1. Consider a multi-agent system described by the dynamics (1) and governed by the control law (8). If $\lambda > V(0)$, the following results hold:

(i) The system can achieve flocking,

(ii) The agents can avoid obstacles,

(iii) The inputs $u_{i\theta}$, u_{ix}, u_{iy} are continuous and bounded.

Proof.

Consider the Lyapunov function and take the derivative,

When $\delta_{ik} \geq \delta_{ikM}$, $D_{ik} = 0$, $C_{ik} = [0, 0]$. When $\delta_{ik} < \delta_{ikM}$, we design an obstacle-avoidance reference trajectory to make $\dot{\theta}_r = \dot{p}_{rx} = \dot{p}_{ry} = 0$.

$$\begin{aligned} \dot{V} = -\sum_{i=1}^{N}\Big[&\mu_{i\theta}\sigma\left(\omega_i - \dot{\theta}_r\right)\left(\omega_i - \dot{\theta}_r\right) + \mu_{ix}\sigma\left(v_{ix} - \tilde{v}_{ix}\right)\left(v_{ix} - \tilde{v}_{ix}\right) \\ &+ \mu_{iy}\sigma\left(v_{iy} - \tilde{v}_{iy}\right)\left(v_{iy} - \tilde{v}_{iy}\right)\Big] \end{aligned} \tag{9}$$

we get $V \leqslant 0$, so $V(0) \geqslant V(t)$, $\forall t \geqslant 0$. Since $\lambda > V(0) \geqslant V(t)$, it means $\lambda = \phi_{ij}(0) > \phi_{ij}(\delta(t))$, $\forall t \geqslant 0$. The condition "separation" is satisfied.

Since equation (53), $V(t)$ is bounded, so $V_1(t)$ is bounded $\forall t \geqslant 0$, $\|f_i - f_2\|_2$ is bounded. The "cohesion" of flocking is satisfied.

Take the derivative of equation (9)

$$\begin{aligned} \ddot{V} = -\sum_{i=1}^{N}\Big[&\mu_{i\theta}\sigma'\left(\omega_i - \theta_r'\right)\left(\omega_i - \dot{\theta}_r\right) + \mu_{i\theta}\sigma\left(\omega_i - \dot{\theta}_r\right)\left(\mu_{i\theta} - \ddot{\theta}_r\right) \\ &+ \mu_{ix}\sigma'\left(v_{ix} - \tilde{v}_{ix}\right)\left(v_{ix} - \tilde{v}_{ix}\right) + \mu_{ix}\sigma\left(v_{ix} - \tilde{v}_{ix}\right)\left(\mu_{ix} - \dot{\tilde{v}}_{ix}\right) \\ &+ \mu_{iy}\sigma'\left(v_{iy} - \tilde{v}_{iy}\right)\left(v_{iy} - \tilde{v}_{iy}\right) + \mu_{iy}\sigma\left(v_{iy} - \tilde{v}_{iy}\right)\left(\mu_{iy} - \dot{\tilde{v}}_{iy}\right)\Big] \end{aligned} \tag{10}$$

since $\|f_i - f_2\|_2$ and $\|v_i - \dot{f}_r\|_2$ and control inputs (we'll prove it next) are bounded, we can get $\ddot{V}(t)$ is bounded $\forall t \geqslant 0$, hence $\dot{V}(t)$ is uniformly continuous, according to "Lyapunov-Like Lemma", $\dot{V}(t)$ approaches 0 as t approaches ∞ . The condition "alignment" of flocking is satisfied. (i) of the theorem is proved. The proof (ii) of the theorem is similar to the proof (i).

Using trigonometric inequalities, we can show that (iii) holds true.

4. Conclusion

This paper presents a distributed flocking control algorithm for convex polygonal multi-agent systems in an obstacle environment. The optimized relative distance improves space utilization and flocking efficiency. The proposed potential function promotes cohesion and collision avoidance, while an obstacle avoidance strategy ensures safe navigation. The algorithm enables stable flocking in complex scenarios, as confirmed by simulations. In the future, we will consider higher-order and complex systems and the application of these methods to more complex obstacle environments.

Acknowledgements

This work was supported in part by the NSFC (62133001, 62227810) and the National Basic Research Program of China (973 Program: 2012CB821200, 2012CB821201).

References

1. A. Okubo, Dynamical aspects of animal grouping: Swarms, schools, flocks, and herds, Advances in Biophysics 22 (1986) 1–94.
2. C. Yu, X. Wang, X. Xu, M. Zhang, H. Ge, J. Ren, L. Sun, B. Chen, G. Tan, Distributed multiagent coordinated learning for autonomous driving in highways based on dynamic coordination graphs, IEEE Transactions on Intelligent Transportation Systems 21 (2) (2020) 735–748.
3. Y. Sun, Z. Wang, H. Su, T. Geng, A brief overview of flocking control for multi-agent systems, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 10984 LNAI, 2018, pp. 48–58.
4. C. W. Reynolds, Flocks, herds and schools: A distributed behavioral model, in: Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1987, Vol. 21, Association for Computing Machinery, 1987, pp. 25–34.
5. T. Vicsek, A. Czir´ok, E. Ben-Jacob, I. Cohen, O. Shochet, Novel type of phase transition in a system of self-driven particles, Phys. Rev. Lett. 75 (1995) 1226–1229.
6. H. Tanner, A. Jadbabaie, G. Pappas, Stable flocking of mobile agents part i: dynamic topology, in: 42nd IEEE International Conference on Decision and Control (IEEE Cat. No.03CH37475), Vol. 2, 2003, pp. 2016–2021.
7. E. Jafet Avila-Martinez, J. Gonzalo Barajas-Ramirez, Flocking motion in swarms with limited sensing radius and heterogeneous input constraints, Journal Of The Franklin Institute-Engineering And Applied Mathematics 358 (4) (2021) 2346–2366.
8. Y. Liu, P. Shi, H. Yu, C.-C. Lim, Event-triggered probability-driven adaptive formation control for multiple elliptical agents, IEEE Transactions on Systems, Man, and Cybernetics: Systems 52 (1) (2022) 645–654.
9. T. Nguyen, H. M. La, T. D. Le, M. Jafari, Formation control and obstacle avoidance of multiple rectangular agents with limited communication ranges, IEEE Transactions on Control of Network Systems 4 (4) (2016) 680–691.

Authors Introduction

Mr. Yaxin Li

He received B.S. degree in automation from Shandong University, China, in 2018. He is currently pursuing the Ph.D. degree with the Seventh Research Division and the Center for Information and Control, School of Automation Science and Electrical Engineering, Beihang University, China. His research interest is consensus control of multi-agent systems.

Prof. Yingmin Jia

He received B.S. degree from Shandong University, Jinan, China, in 1982, and the M.S. and Ph.D. degrees from Beihang University, Beijing, China, in 1990 and 1993, respectively. He is currently a professor with the Seventh Research Division and the director of the Center for Information and Control at Beihang University. His research interests include robust control, robotic systems, spacecraft coordination and on-orbit servicing.

Low-light Image Enhancement with Color Space (Cielab)

Lee Kok Xiong
Institute of Computer Science and Digital Innovation, UCSI University, Cheras, 56000 Kuala Lumpur, Malaysia

Kasthuri Subaramaniam
Department of Decision Science, Faculty of Business and Economics, University of Malaya, 50603 Kuala Lumpur, Malaysia

Umm E Mariya Shah
Institute of Computer Science and Digital Innovation, UCSI University, Cheras, 56000 Kuala Lumpur, Malaysia

Abdul Samad Bin Shibghatullah
College of Computing & Informatics, Universiti Tenaga Nasional, 43000 Kajang, Selangor, Malaysia

Oras Baker
Faculty of Computing, Ravensbourne University London, SE10 0EW, United Kingdom

Email: 1002060150@ucsiuniversity.edu.my, s_kasthuri@um,edu.my, mariya@ucsiuniversity.edu my, abdul.samad@uniten.edu.my, O.alhassani@rave.ac.uk

Abstract

In this project, we implemented a color transformation from RGB to CIELAB to enhance low-light images. This transformation separates color information from brightness information, which improves contrast and overall quality. We are using a standard color conversion formula and combining it with other techniques, such as histogram equalization and neural networks, for better results. The project will have a user-friendly interface that allows users to upload and download images and compare the original and enhanced versions. The programming language used and the specific details of the implementation process are not mentioned.

Keywords: Low-light, Image Enhancement, Color Space, Cielab

1. Introduction

The history of image processing can be traced back to the 1960s [1], when research institutions began experimenting with enhancing image quality for medical imaging, video telephony, character recognition, and satellite imagery. The goal was to improve the visual impact of low-quality images and generate higher-quality images as a result [2]. Since then, numerous types of image processing have been developed, including image enhancement, restoration, encoding, and compression [3].

Nowadays, the use of digital devices such as smartphones and digital cameras has become ubiquitous 4. However, low light conditions can have a significant impact on image quality, causing images to lose colour information, contrast, and brightness [5]. To address this issue, experts have developed various methods to enhance the quality of low- light images, including histogram equalization, illumination map estimation [6], normalization flow [7], neural networks [8], and dark region-aware low-light image enhancement [9].

To further improve the effectiveness of low-light image enhancement techniques, colour space transformations have been developed to convert RGB colour space to HSI or CIELAB colour space. This project proposes the use of colour transformation from RGB to CIELAB, along with a step-by-step implementation of the transformation and enhancement process. The project also includes a simple interface for uploading low-light images and downloading the enhanced images, as well as a comparison box to show the difference between the original and enhanced images

2. Related Work

2.1. *Histogram Equalization (HE)*

In image processing, histogram equalization is a method used to enhance contrast in images by expanding the intensity range and distributing the most common intensity levels. This allows for regions with low local contrast to receive more contrast, and intensity distributions are changed by histogram equalization. It can be divided into global histogram equalization (GHE) and local histogram equalization (LHE), with LHE being more effective in improving overall contrast. However, histogram equalization may also significantly change the mean brightness of an image, which may not be desired in some cases [10].

2.2. *Illumination Map Estimation (LIME)*

The Illumination Map Estimation (LIME) method is a new approach to enhance low-light images by estimating illumination maps for each RGB color channel separately. Unlike the variational model, LIME proposes additional illumination enhancement processes that further improve the illumination conditions of the image. LIME is considered one of the most advanced works in the field of classical algorithms and builds on the Retinex-based category. However, retinex-based methods produce undesirable results such as halo aberrations, poor contrast, and excessive smoothness in the enhanced images, and certain methods over-compute during the enhancement process.

2.3. *Normalizing Flow (NF)*

A normalizing flow is a method of transforming a simple probability distribution into a more complex distribution using a set of invertible and differentiable mappings. The layers of the network must be constructed carefully to ensure that the inversion and determinant of the Jacbian matrix can be easily captured, limiting the capacity of the generative model. Modifications have been proposed to increase the expressive power of the model, such as 1 x 1 convolution, partitioning and concatenation, permutation, and affine coupling layers. Wang et al. [11] implemented this method to accurately learn local pixel correlations and global image properties by modeling the distributions over the normally exposed image. Although this method provides saturation enrichment and color distortion reductions, it still leads to unnatural image colors [15].

2.4. *Neural Network (CNN)*

Based various image enhancement methods based on neural networks that have emerged in recent years. Inception and LLCNN use convolutional blocks in their pipelines, while ResNet uses residual connections to provide two different channels for illumination and reflection. However, these methods still result in visible visual artifacts. Mehwish et al [12] proposed a color-wise attention network method that balances color on low frequencies using a sigmoid function and preserves contrast while reducing color saturation. Lore et al [13] used LLNet, the first deep learning technique applied to low-light image enhancement, which significantly outperformed conventional techniques using a variation of the stacked-sparse denoising and contrast-enhancement autoencoder. However, a sizable amount of dataset is required for effective training.

2.5. *Dark Region-Aware Low-light Enhancement (DALE)*

The method uses visual attention and enrichment networks to improve images with low light. The attention network creates an attention map to detect dark areas, and the enrichment network enriches the image with low light [14]. The method has been successfully implemented in many images, but requires a large dataset for training.

3. Method

3.1. *Accuracy of numerical integration*

The CIELAB color space is a widely used color model that is designed to be more perceptually uniform than RGB. It separates color information into three components: L for lightness, A for the red-green axis, and B for the blue-yellow axis.

The formula provided in this paper is a way to enhance low-light images by adjusting the lightness component (L) while preserving the chroma and hue information in the image. The steps are as below:

The chroma of the image is calculated by taking the square root of the sum of the squares of the A and B components. This represents the intensity of the color in the image.

$$\text{Chroma} = \text{square root}\,(A*A + B*B) \tag{1}$$

The hue of the image is calculated by taking the arctangent of the B component divided by the A component. This represents the angle of the color in the red-green-blue color space.

$$\text{Hue} = \tan\text{-}1\,(B/A) \tag{2}$$

The overall intensity of the image is calculated by taking the average of the R, G, and B components.

$$I = (R+G+B)/3 \tag{3}$$

An alpha value is calculated based on the intensity of the image, which is used to scale the lightness component later.

$$\text{Alpha} = (255 - I)/255 \tag{4}$$

The new lightness value is calculated by adding 1 to the alpha value and multiplying it by the old lightness value. This increases the brightness of the image.

$$\text{newL} = (\text{Alpha} + 1) \text{ x oldL} \tag{5}$$

The ratio of the new lightness value to the old lightness value is calculated.

$$ratio = newL/oldL \quad (6)$$

The new chroma value is calculated by multiplying the ratio by the old chroma value. This preserves the intensity of the color in the image.

$$new\ chroma = ratio \times Chroma \quad (7)$$

The new A value is calculated by multiplying the new chroma value by the cosine of the hue angle. This adjusts the red-green axis of the image.

$$newA = new\ chroma \times \cos(Hue) \quad (8)$$

The new B value is calculated by multiplying the new chroma value by the sine of the hue angle. This adjusts the blue-yellow axis of the image.

$$newB = new\ chroma \times \sin(Hue) \quad (9)$$

Finally, combine newL, newA and newB together then convert it back to RGB color space. By using this formula, the brightness of the image will improve and it will become clearer while maintaining the color of the image.

3.2. *System Interface*

The interface design of the LLIE system should be simple and consistent to avoid users wasting time learning how to use the system repeatedly. Therefore, the system's design was based on research of similar interfaces with features such as image upload, conversion, comparison, and download. The LLIE system uses Tkinter to build its interface, which is a simple and efficient GUI toolkit for Python.

4. Result

After applying the colour space method Fig. 1, Fig. 2, and Fig. 3 are the result of the low-light image enhancement system together with the interface used to compare the original image with the enhanced image.

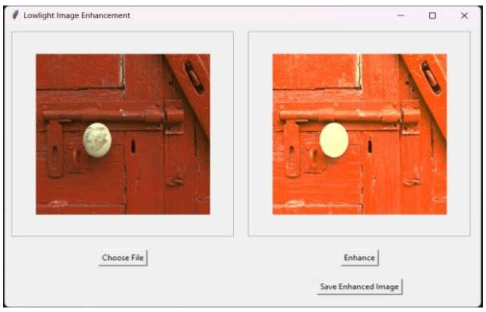


Fig. 1. Door Lock Images

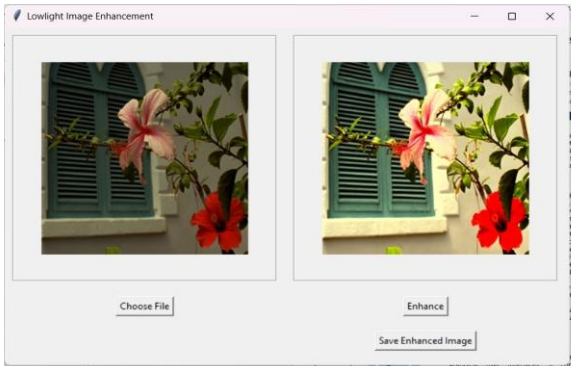


Fig. 2. Flower Images

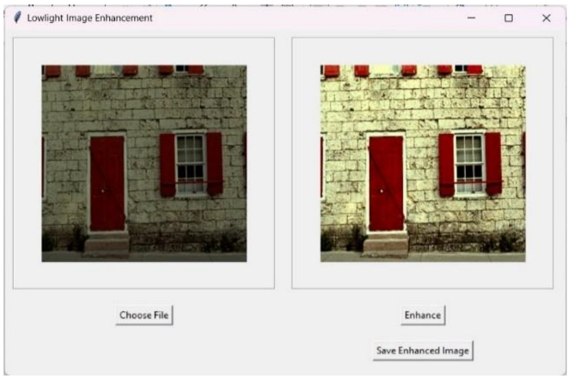


Fig. 3. Wall Images

5. Conclusion

The proposed application aims to improve the overall brightness of the image. However, the proposed application is bound and limited to the certain images and as such, further research will have to be conducted in order to expand the array of viable range of images.

In the conclusion, the result in increasing the brightness of the image helps the user to observe the object in the image clearer with the improvement of the colour but there are limitations that will need to be improved in the future work.

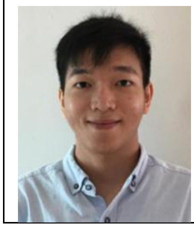
References

1. Pratt, W. K. (2007). Digital image processing: PIKS inside. John Wiley & Sons.
2. Wang, L., Zhao, L., Zhong, T., & Wu, C. (2024). Low-light image enhancement using generative adversarial networks. Scientific Reports, 14(1), 18489.
3. Feng, Y., Hou, S., Lin, H., Zhu, Y., Wu, P., Dong, W., ... & Zhang, Y. (2024). Difflight: integrating content and detail for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 6143-6152).
4. Statista. (2022). Number of mobile phone users worldwide from 2019 to 2026 (in billions).
5. Gonzalez, R. C., & Woods, R. E. (2018). Digital image processing. Pearson Education India.

6. Guo, X., Li, Y., Ling, H., & Wu, J. (2017). LIME: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, 26(2), 982-993.
7. Chen, C., Chen, Q., Xu, J., & Koltun, V. (2018). Learning to see in the dark. In Proceedings of IEEE conference on computer vision and pattern recognition (pp. 3291-3300).
8. Li, C., Ren, W., Fu, D., Tao, D., & Feng, D. (2018). LLNet: A deep autoencoder approach to natural low- light image enhancement. IEEE Transactions on Image Processing, 27(6), 2608-2622.
9. Chen, Z., Yang, J., Li, F., & Feng, Z. (2023). Real-Time Low-Light Image Enhancement Method for Train Driving Scene Based on Improved Zero-DCE. In International Conference on Electrical and Information Technologies for Rail Transportation (pp. 9-18). Singapore: Springer Nature Singapore.
10. Ismael, A. N. (2022). A comparative Study of Image Enhancement Techniques for Natural Images. J. Al-Qadisiyah Comput. Sci. Math, 14(4), 53-65.
11. Wang, Y., Wan, R., Yang, W., Li, H., Chau, L. P., & Kot, A. (2022). Low-light image enhancement with normalizing flow. In Proceedings of the AAAI conference on artificial intelligence (Vol. 36, No. 3, pp. 2604-2612).
12. Atoum, Y., Ye, M., Ren, L., Tai, Y., & Liu, X. (2020). Color-wise attention network for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 506-507).
13. Lore, K. G., Akintayo, A., & Sarkar, S. (2017). LLNet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition, 61, 650-662.
14. Yu, X., Bo, L., & Xin, C. (2022). Low light combining multiscale deep learning networks and image enhancement algorithm. 2(4), 0215-0232.
15. Tan Wan Yin, Kasthuri Subaramaniam, Abdul Samad Bin Shibghatullah and Nur Farraliza Mansor, 2022. Enhancement of Low-Light Image using Homomorphic Filtering, Unsharp Masking, and Gamma Correction. International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 13, No. 9, pp. 53-60.

Authors Introduction

Mr. Lee Kok Xiong



He is currently an undergraduate student, pursuing his study in Bachelor of Science (Hons) in Computing from Institute of Computer Science and Digital Innovation (ICSDI) at UCSI University, Malaysia.

Dr. Kasthuri Subaramaniam

She received her Bachelor's and Master's degrees in Computer Science from University of Malaya. She earned her Ph.D. in Informatics from Malaysia University of Science & Technology. Her research interests include human-computer interaction, human personality types, augmented reality, e-learning, mobile commerce and e-commerce.

Dr. Umm e Mariya Shah

She received the Ph.D. degree in Computer Science from Universiti Malaya, Malaysia and the MS degree in Computer Science from COMSATS University, Islamabad, Pakistan and the BS degree in Computer Science from Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan. Her research areas include usability engineering, human-computer interaction, and healthcare (ICT, interdisciplinary).

Dr. Abdul Samad Bin Shibghatullah

He received his Bachelor's degree in Accounting from Universiti Kebangsaan Malaysia, M.Sc. degree in Computer Science from Universiti Teknologi Malaysia and Ph.D. in Computer Science from Brunel University, UK. He is currently an Associate Professor at the College of Computing & Informatics (CCI), Universiti Tenaga Nasional, Kajang, Malaysia.

Dr. Oras Baker

He is an Associate Professor and Head of Masters in Cyber Security and Cyber Security Management at Ravensbourne University London, UK. With 25 years of distinguished experience spanning academia, research, and industry, he specialises in Artificial Intelligence, Software Engineering, Cyber Security, Data Mining, and Machine Learning.

Integrated AI Voice Assistant News Website for Enhancing User Experience – AI-ReadSmart

Mohammed Mohi Uddin
Institute of Computer Science and Digital Innovation, UCSI University, Cheras, 56000 Kuala Lumpur, Malaysia

Ghassan Saleh ALDharhani
Institute of Computer Science and Digital Innovation, UCSI University, Cheras, 56000 Kuala Lumpur, Malaysia

Kay Hooi Keoy
Center for Business Informatics and Industrial Management, UCSI Graduate Business School, UCSI University, Cheras, 56000 Kuala Lumpur, Malaysia

Chit Su Mon
School of Mathematical and Computer Sciences, Heriot-Watt University Malaysia, 62200 Putrajaya, Malaysia

Kasthuri Subaramaniam
Department of Decision Science, Faculty of Business and Economics, University of Malaya, 50603 Kuala Lumpur, Malaysia

Email: 1001957099@ucsiuniversity.edu.my, ghassan@ucsiuniversity.edu.my, alankeoy@gmail.com, c.mon@hw.ac.uk, s_kasthuri@um,edu.my

Abstract

The offline newspaper sector has been declining for years, and following the epidemic there were further decreases. To Assist Users With the complexity issue of the current online news sector, this study aims to develop a website that is beneficial to both users and the news sector. To ensure a seamless transition to online reading, research was conducted on the existing state of the online news sector. The survey used in this study allowed the researcher to understand how individuals feel about the state of the current online news sector as well as how they feel about voice integration on news websites.

Keywords: Covid-19, Online news, Voice assistant, Artificial intelligence

1. Introduction

In today's world internet has become one of the basic needs for everyone. Internet has transformed people's life in a positive way. An innovative approach to print media was influenced and created by the development of the internet. This encouraged the publication of online content by newspapers and other media. The news industry viewed this as an opportunity to increase revenue since it was the internet era, but there was a percentage of people that preferred to read on paper and did not switch to the online format until COVID-19 came into existence [2]. The use of newspapers was already on the verge of declining, and with the added impact of the COVID-19 outbreak, almost one-fourth of newspaper companies have faded [3].

Machine learning, which has been ingrained in our daily lives without our knowledge, must rank among the greatest technological advancements in recent years. supporting us in doing daily tasks and more [7]. People utilize machine learning and other artificial intelligence technology in one form or another, e.g., image recognition, speech recognition, advancements in healthcare, banking, and many more. Although machine learning has been around since the 1950s, its actual potential has just recently become known due to the tremendous growth in the amount of information that is currently available. Applications and services that are available today in the market are trying to improve their services through artificial intelligence and machine learning. The goal of this project is to support the digital news industry by incorporating a voice assistant into news websites to simplify navigation and improve the user experience as more people switch from traditional paper to digital media.

2. Literature Review

2.1. *The Evolution of Newspapers from Paper to Digital*

Today, technology has become a crucial component of human society. Before the invention of the internet, reporters were required to call in their reports or type them in the newsroom. Later, reporters sent it to their editors, after which the news was released. However, modern technology enables such reporters to address stories from any location using a laptop or a smartphone [4]. enabling businesses to post breaking news on websites minutes after the incident had taken place. Compared to newspapers, which require the incident to be reported first and then printed on paper, this results in delaying the news to be published till the next day.

Digital content is persistent on the web. It cannot be removed from the web [5] [6] [8]. On the other hand, in a situation where the newspaper is lost or torn will require a new purchase of another newspaper. Digital information is far superior in this situation since readers can access news from any given date they pick. Digital content can also be accessed instantly anywhere in the world [7].

2.2. *Importance of News Companies*

The news industry, which is regarded as the go-to source for informing people of daily updates regarding the events taking place all around them, is one of a nation's most important services. An established and well-respected news agency serves as a domestic and international news source. It facilitates the distribution of information, including official alerts. Social awareness and any kind of threat can be simply communicated through a news article. One Article can rally support for causes that are financially out of reach for the average person. It is extremely essential because it briefly covers a variety of topics regarding international affairs [1] [12].

2.3. *What are Voice Assistants*

A voice assistant is a type of virtual assistant that may do tasks or offer services on someone else's behalf. Voice Assistants can help you with a wide range of tasks, from having a nice two-way conversation to guiding you through more challenging tasks like placing an online order for goods.

Voice assistants have been created by several companies, including Amazon Alexa, Google Assistant, and Microsoft Cortana. The true ability of an assistant comes from its potential to offer support, respond swiftly and accurately to a variety of inquiries, grasp context, speak many languages, communicate, and even manage other programs [8].

2.4. *Evolution of Voice Assistants over the Last Decade*

In 2011 Siri was launched as a mobile application, which Apple later purchases. An artificial voice assistant called Siri first appeared as an iPhone app. In 2010, Apple purchased the company and Siri after recognizing its potential. Since the iPhone 4S, Apple has included Siri as its voice assistant in all voice-capable products, including its current ecosystem of smart speakers and wearable devices. As a result of how quickly it spread via the iPhone, Siri was for many people the first voice assistant, and it went on to set standards for how voice assistants should operate [9].

Google Voice Search was introduced to the website in 2012. Although Google began experimenting with voice search on websites in the 2000s, it wasn't until 2012 that Google included a tiny microphone button for voice searches. Use of speech recognition technology by the powerful Google search engine marked a significant advancement for the sector. Although only Chrome browser supported voice search, a significant market suddenly had the chance to experiment with voice-based internet interaction [9].

2014 marked the debut of Microsoft's Cortana voice assistant. Microsoft's entry into the voice assistant business had a very promising start in 2014. In addition to Windows, Xbox, and other innovations from Microsoft, Cortana would also be included in devices created by other companies. However, Microsoft plans to stop offering Cortana's consumption services by the end of the decade and completely redesign it for use in enterprise-level projects [9].

Amazon's efforts to create consumer technology were overshadowed by its dominance in e-commerce and cloud computing until the company revealed a smart speaker with its voice assistant named after the Library of Alexandria. Alexa has surpassed Siri and Cortana in popularity since its debut in 2014, to the point where it is now widely used as a standard for voice assistants.

Google used all it had discovered from its voice app research to develop Google Assistant in 2016, which competes with Alexa for supremacy as a voice assistant option. In addition to offering many of the same features as Amazon's innovations, Google Assistant also has the added benefit of being linked with Google's larger digital environment, which includes Android phones [10] [11].

2.5. *Read Smart can help Address the Current Efficiency of the News Website*

Interactivity issues

With the use of the voice assistant and speech recognition, content can be accessed within seconds through the users' interaction with the voice assistant. The present systems are based on the users needing to navigate through each section of the complicated web page to find their desired content.

Multitasking

Because users spend too much time on the content, it prevents them from multitasking. In contrast, ReadSmart enables multitasking while working on other tasks. It will also benefit those who have poor vision, such as those who need glasses to read (Table 1).

Table 1. Problems and Solutions provided for the current system

Problems (Existing)	Solutions (Proposed)
Complicated UI.	Simple and clean UI.
Hard to navigate directly to a category of news.	Can easily navigate to a category using Voice Assistant.
News is congested and spread all over the page	News Can be Searched According to Terms, Categories, or latest news, making it simple and less congested.
Cannot Multitask while navigating.	Can Multitask as the Voice Assistant will help read the headlines.

3. Methodology

This project employed a quantitative survey methodology so that data could be utilised to understand the users on their perception about the current websites and a user acceptance test along with a questionnaire is carried out to confirm that the Prototype is in a standard level for use.

As the project's primary requirements are set, ReadSmart will utilize the iterative model (Fig. 1) of the System Development Life Cycle (SDLC).

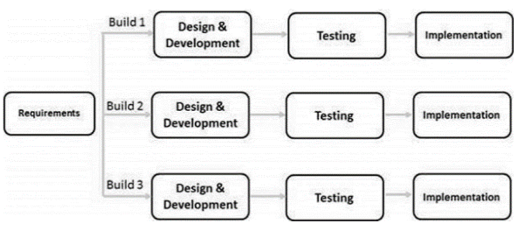


Fig. 1. Iterative model

3.1. *Implementation*

The implementation phase comes once the Design Phase is decided. The developers will need to pick a suitable programming language with adequate documentation, instructional videos, and a suitable community during this step. This will guarantee that the community will be there to support or resolve issues if the developer runs into any coding-related concerns.

At this stage of the project, the system's real construction starts. The analysis and design that were produced during the design stage will serve as the stage's guidelines. In this phase, every requirement, tactic, and design plan have been implemented and coded. The chosen design will be implemented by the developer using accepted coding and measurement standards. At this stage, the prototype will go through iterations to become a fully functional system. Depending on the project, the difficulty of the task and the length of time needed for this iteration will change.

For the Implementation Stage, the following deliverables shall be created.

- Website
- Source Code for Frontend and Backend

4. Results and Discussion

Since the development of e-reading, online reading has grown in popularity. All generations find reading articles, novels, or journals to be highly interesting. Even though people used to read before the rise of the internet, the findings indicate that most users preferred the new navigation system(prototype) over the previous one. The complexity was decreased, and the user interface was friendlier in the new proposed solution. When compared to current models on the market, it had proper functionalities and a cleaner appearance. Due to the large number of voice assistants on the market, firms ensure to offer adequate security, which itself is maintained for this project.

5. Conclusion

Finally, it can be said that the system was maintained throughout the project's developmental phases, which led to its effective development and execution of the project's main goals. The major goal was to research existing online news websites on the market before developing an AI-based news website to address issues with the current system. Studying current systems and literatures is a minor step toward overcoming a gap in the development of AI-based news websites. Both on the website and the phone, the prototype was successfully constructed and functions flawlessly. A superior outcome would also come from a more thorough analysis of the variables. The assistant can be into a wide range of additional applications in diverse types of fields. This study provides a framework for future academics to consider in the event of a comparable pandemic scenario affecting not only the newspaper sector but also other industries.

References

1. R. Mishra and P. Koundal, "Print and Online Newspapers: An Analysis of the News Content and Consumption Patterns of Readers," June 2020. [Online].
2. Preißner, S., Raasch, C., & Schweisfurth, T. (2024). When necessity is the mother of disruption: Users versus producers as sources of disruptive innovation. Journal of Product Innovation Management, 41(1), 62-85.
3. Hill, J. (2016). Newspapers in Decline. In The British Newspaper Industry: The Future of the Regional Press (pp. 69-79). London: Palgrave Macmillan UK.

4. Weber, M. S., & Mathews, N. (2024). Explicating local: An audience-based framing of local community and local news. Digital Journalism, 12(1), 28-47.
5. Tan, Y. R., Subaramaniam, K., & Kolandaisamy, R. (2023, July). Developing Interface Designs with Personality Types: Self-management Application–Luvlife. In International Conference on Human-Computer Interaction (pp. 74-89). Cham: Springer Nature Switzerland.
6. Chan, M., Lee, F. L., & Chen, H. T. (2024). Avoid or authenticate? A multilevel cross-country analysis of the roles of fake news concern and news fatigue on news avoidance and authentication. Digital Journalism, 12(3), 356-375.
7. Subaramaniam, K., Ern-Rong, J. L., & Palaniappan, S. (2021). Interface designs with personality types: an effective e-learning experience. EVERGREEN – Joint Journal of Novel Carbon Resources Sciences & Green Asia Strategy. Vol. 08, Issue 03, pp. 618-627.
8. Faruk, L. I. D., Babakerkhell, M. D., Mongkolnam, P., Chongsuphajaisiddhi, V., Funilkul, S., & Pal, D. (2024). A review of subjective scales measuring the user experience of voice assistants. IEEE Access.
9. Murad, C., Munteanu, C., Cowan, B. R., & Clark, L. (2019). Revolution or evolution? Speech interaction and HCI design guidelines. IEEE Pervasive Computing, 18(2), 33-45.
10. Rashid, A. B., & Kausik, A. K. (2024). AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. Hybrid Advances, 100277.
11. Jalil, A., Kolandaisamy, R., Subaramaniam, K., Kolandaisamy, I., & Khang, J. Q. G. (2020). Designing a mobile application to improve user's productivity on computer-based productivity software. Journal Of Advanced Research In Dynamical And Control Systems.
12. Sheresheva, M., Skakovskaya, L., Bryzgalova, E., Antonov-Ovseenko, A., & Shitikova, H. (2021). The print media convergence: overall trends and the COVID-19 pandemic impact. Journal of Risk and Financial Management, 14(8), 364.

Authors Introduction

Mr. Mohammed Mohi Uddin

He is currently an undergraduate student, pursuing his study in Bachelor of Science (Hons) in Computing from Institute of Computer Science and Digital Innovation (ICSDI) at UCSI University, Malaysia.

Dr. Ghassan Saleh ALDharhani

He is an assistant professor who is currently teaching at UCSI University. His research interests include cybersecurity, ethical hacking and machine learning. He is also the Head of Digital Innovation Department, a TTT Certified HRDF Trainer and Cyber Security Instructor,

Dr. Kay Hooi Keoy

He is an Associate Professor at UCSI Graduate Business School, UCSI University. He is holding the post of Deputy Dean and Centre for Business Informatics and Industrial Management at Graduate Business School. His extensive expertise encompasses a broad spectrum within the realms of business and management, with a specific focus on operations management, knowledge management, entrepreneurship, and sustainability..

Dr. Chit Su Mon

She is an Assistant Professor at the School of Mathematical and Computer Sciences and Programme Director of Studies at Heriot-Watt University Malaysia. She holds a PhD in Computing from Sunway University, completed in 2022. Her research interests include Human-Computer Interaction (HCI), Haptics, Assistive Technology, and Data Analytics.

Dr. Kasthuri Subaramaniam

She received her Bachelor's and Master's degrees in Computer Science from University of Malaya. She earned her Ph.D. in Informatics from Malaysia University of Science & Technology. Her research interests include human-computer interaction, human personality types, augmented reality, e-learning, mobile commerce and e-commerce.

Developing a Mobile Healthcare Application - MyHealth

Abdulrahman Salmo Alhamada
Institute of Computer Science and Digital Innovation, UCSI University, Cheras, 56000 Kuala Lumpur, Malaysia

Ghassan Saleh ALDharhani
Institute of Computer Science and Digital Innovation, UCSI University, Cheras, 56000 Kuala Lumpur, Malaysia

Kasthuri Subaramaniam
Department of Decision Science, Faculty of Business and Economics, University of Malaya, 50603 Kuala Lumpur, Malaysia

Raenu Kolandaisamy
Institute of Computer Science and Digital Innovation, UCSI University, Cheras, 56000 Kuala Lumpur, Malaysia

Email: 1002162511@ucsiuniversity.edu.my, ghassan@ucsiuniversity.edu.my, s_kasthuri@um,edu.my, raenu@ucsiuniversity.edu.my

Abstract

MyHealth targets common mobile healthcare problems like medication forgetfulness and basic health knowledge gaps. This mobile application gives medication reminders, health lessons and appointment scheduling. It employs user centric design principles in combination with modern technology for user engagement & health management. System architecture, user interface design and development process are discussed with regard to healthcare accessibility and patient compliance impact.

Keywords: Mobile healthcare, Health education, Appointment scheduling, Reminder

1. Introduction

With the speed of life today, managing healthcare effectively can be a challenge, especially for those with chronic conditions or busy lifestyles. Traditional methods of healthcare management fail to address issues like medication adherence and reliable health information access [1]. The "MyHealth" mobile application fills this need by combining medication reminders, educational content and appointment scheduling functions.

With the introduction of smartphones and mobile internet, mobile applications are an essential tool in many sectors including healthcare [2]. Mobile health (mHealth) applications can provide personalised and accessible healthcare solutions that are indispensable for modern health management. These applications can help patients interact with healthcare providers in real time & increase patient engagement [3].

Patients, caregivers and professionals can use MyHealth application to have access to important health management tools at all times [4]. MyHealth stemmed from addressing common healthcare challenges like medication non- adherence, which is estimated to contribute to major health problems and increased healthcare costs globally [5]. Also, through educational content, MyHealth aims to inform users about health and decision making.

This article describes the conceptualization, design and development phases of MyHealth application and how it can change personal healthcare management [7]. The study also looks at the existing mobile healthcare applications landscape to identify gaps and propose solutions to which MyHealth hopes to contribute. Through user-centric design principles & advanced technology MyHealth aims to improve healthcare accessibility, patient compliance and health outcomes.

2. Literature Review

2.1. *Further Elaboration on Existing Systems Analysis*

Mobile applications are well studied for improving healthcare delivery and patient outcomes [6], [12]. For example, use of mHealth solutions has been linked with better medication adherence and management of chronic diseases. Existing applications like Medisafe and WebMD give insight into user preferences and feature effectiveness.

Medisafe, a medication management app, has reminders and a visual pillbox. Similarly, WebMD combines medication tracking with health journal features to document symptoms and health progress. These applications stress that user-friendly interfaces and personalized reminders are key to adhesion promotion. Also, applications like HealthTap and WebMD provide libraries of health information so users can get solid medical advice and information [11], [13].

2.2. *Benefits of using "MyHealth" Application*

Several benefits of the MyHealth application are aimed at the existing systems:

- Comprehensive Health Management: Many existing apps handle only medication reminders - MyHealth handles medication management, health education and appointment scheduling all in one place.
- User-Centric Design: MyHealth focused on user experience and making the application simple to use. This user-centric design approach drives user engagement & compliance [8], [9].
- Health Education: MyHealth has educational content That gives users knowledge to make sound health decisions.
- Integrated Appointment Scheduling: MyHealth makes booking & managing appointments for healthcare easy.

2.3. *Weaknesses in current Existing Systems*

Table 1. Literature on Existing Systems

Feature	Medi safeapp	Doctor on demand app	WebMD app	MyHealth
Medication Reminders	Yes, customizable reminders	No	Yes	Yes, with a countdowntimer
Health Education	Yes, in-depth information on a dozen conditions	No	Yes	Yes, preventive measures and self-care resources
Appointment Scheduling	No, but you can add appointment reminder manually	Yes	Yes	Yes, easy and streamlined scheduling
User Interface (UI) & UX	Intuitive, easy to use[	Simple, user-friendly for all ages	Informative, content-heavy	Simple, user-friendlyfor all ages

There are many benefits of existing mHealth applications as shown in Table 1, however there are some limitation as stated below.

Lack of Comprehensive Features: Many existing apps either do not offer medication reminders, appointment scheduling, or health education, requiring users to use multiple apps for complete healthcare management.

User Experience Issue: an app like "WebMD" app has a content-heavy and a little overwhelming interface, which can be challenging for users to navigate and use efficiently.

2.4. *Addressing Existing System weaknesses*

All-in-One Healthcare Management: Integrates medication reminders, health education, and appointment scheduling into a single app, providing users with a comprehensive tool for managing their health.

Enhanced User Experience: Features a simple and intuitive user interface, making it easy for users of all ages to navigate and use the app effectively.

3. Methodology

Iterative development in Agile methodology was followed, allowing flexibility and continuous improvement based on user feedback [10]. The development process involved key phases such as:

Requirement Analysis: Determining the functional and non- functional requirements of an application and documenting them.

Design: Wireframes and prototypes for the user interface.

Development: Build the application in Android Studio & Firebase with focus on modularity.

Testing: conduct strict tests such as unit tests, integration tests & user acceptance testing (UAT) to ensure application reliability and performance.

4. Synthesis

4.1. *Use Case Diagram*

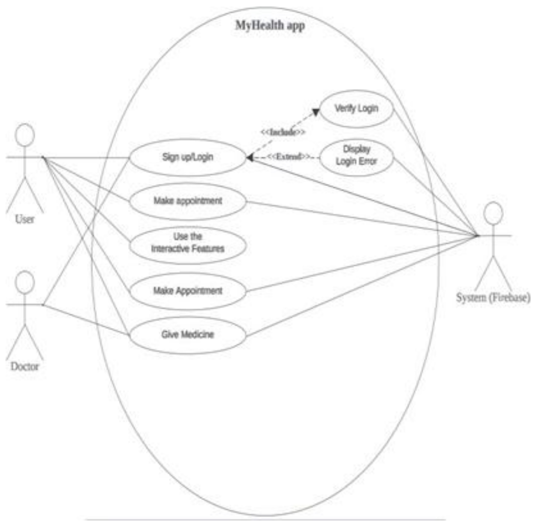


Fig. 1. Use Case Diagram

A use case diagram for "MyHealth," in Fig. 1 shows exactly how a User interacts with a Doctor and also with the System (using firebase as a backend). The User has several choices to engage in such actions as registering / signing in, scheduling appointments, utilizing interactive features, and also getting medication from a Doctor. The

Doctor may make appointments and offer medication to the User. The System is responsible for confirming the User's login and displaying any login errors. The MyHealth app diagram displays the most crucial functions as well as interactions.

4.2. *Activity Diagram*

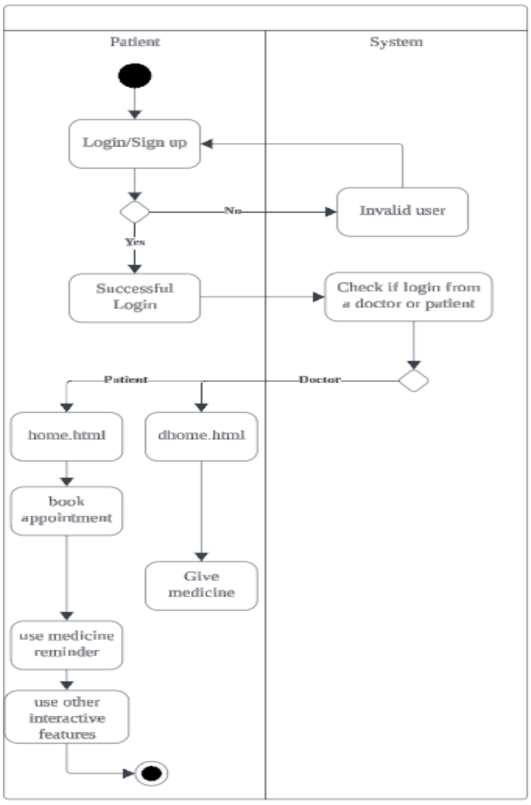


Fig. 2. Activity Diagram

This activity diagram (Fig. 2) visualize the user interactions. The screen displays the login / sign up procedure, where a patient can use the home page, appointment booking, medicine reminder along with other interactive options after logging in successfully. The user is verified as either a patient or a dcotor by the system and an invalid user is marked as a such. Doctors can give medicine to patients via the system. The flow shows the primary features of "MyHealth" app including user authentication, access to patient-specific tools, and the ability for doctors to deliver care remotely.

"MyHealth" application moves through its user flow and important characteristics within this flowchart. The process begins with a login / sign-up stage where the user's login is checked. The system then verifies if the person is a doctor or a patient in case the login is correct. The system will take the user to either the doctor page, or the patient page based upon their type. the patient can make an appointment, interact with interactive features, set a reminder for medicine, and schedule a meeting. The doctor then delivers medicine to the patient.

5. Implementation

The following shows the user interfaces for "MyHealth" app.

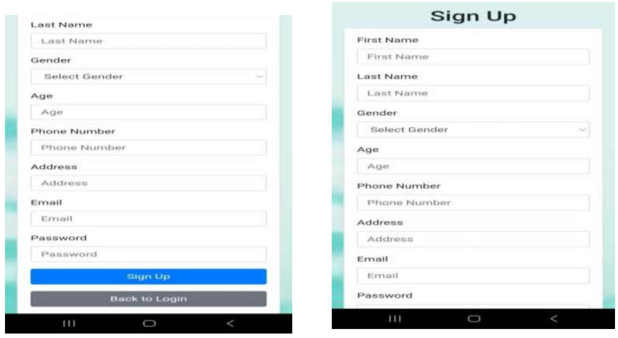


Fig. 3. Sign up

Fig. 3 shows the signup page for the users, the user will need to enter their first name, last name, gender, age, phone number, address, email, and password. And then they can go back to the login page by clicking the "back to login" button.

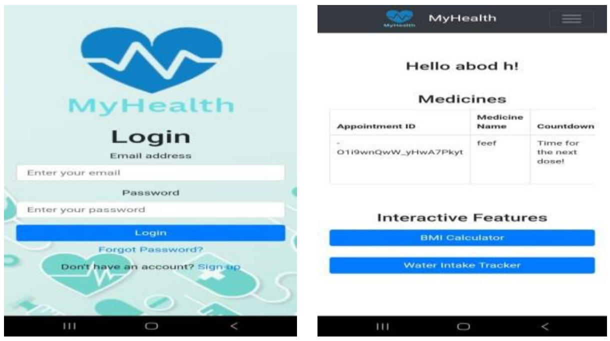


Fig. 4. Login and Home page

Fig. 4 shows the login screen for the MyHealth app is simple and clean. It displays the app logo & app name at the top. Users type in an email address along with a password to sign. The Fig. 4 also shows the home page of MyHealth app, it shows the medicine table with the appointment ID and the medicine name and the countdown until when to take the next dose of the medicine. There are also some interactive features like a BMI Calculator and water intake tracker.

6. Conclusion

MyHealth is an important step forward in mobile healthcare - manage medications, get health education & book appointments. By combining user-centered design principles and modern technology the application aims at improving healthcare accessibility and patient compliance. Future work includes expanding the application's features, integrating it with wearable devices and conducting extensive user trials to collect feedback and improve.

MyHealth project illustrates how common healthcare problems can be solved using innovative solutions and how mobile applications can improve healthcare delivery and patient outcomes.

References

1. Han, Q. (2024). Topics and Trends of Health Informatics Education Research: Scientometric Analysis. JMIR Medical Education, 10, e58165.
2. Galetsi, P., Katsaliaki, K., & Kumar, S. (2023). Exploring benefits and ethical challenges in the rise of mHealth (mobile healthcare) technology for the common good: An analysis of mobile applications for health specialists. Technovation, 121, 102598.
3. Li, C., Wang, J., Wang, S., & Zhang, Y. (2024). A review of IoT applications in healthcare. Neurocomputing, 565, 127017.
4. López Martínez, A., Gil Pérez, M., & Ruiz-Martínez, A. (2023). A comprehensive review of the state-of-the-art on security and privacy issues in healthcare. ACM Computing Surveys, 55(12), 1-38.
5. AbdulRaheem, M., Awotunde, J. B., Chakraborty, C., Adeniyi, E. A., Oladipo, I. D., & Bhoi, A. K. (2023). Security and privacy concerns in smart healthcare system. In Implementation of Smart Healthcare Systems using AI, IoT, and Blockchain (pp. 243-273). Academic Press.
6. Familoni, B. T., & Babatunde, S. O. (2024). User experience (UX) design in medical products: theoretical foundations and development best practices. Engineering Science & Technology Journal, 5(3), 1125-1148.
7. Hyzy, M., Bond, R., Mulvenna, M. D., Bai, L., Daly, R., & Leigh, S. (2023, July). Objectively assessing and comparing the user experience of two thousand digital health apps. In International Conference on Human-Computer Interaction (pp. 335-343). Cham: Springer Nature Switzerland.
8. Tan, Y. R., Subaramaniam, K., & Kolandaisamy, R. (2023). Developing Interface Designs with Personality Types: Self-management Application–Luvlife. In International Conference on Human-Computer Interaction (pp. 74-89). Cham: Springer Nature Switzerland.
9. Subaramaniam, K., Ern-Rong, J. L., & Palaniappan, S. (2021). Interface designs with personality types: an effective e-learning experience. EVERGREEN – Joint Journal of Novel Carbon Resources Sciences & Green Asia Strategy. Vol. 08, Issue 03, pp. 618-627,
10. Jalil, A., Kolandaisamy, R., Subaramaniam, K., Kolandaisamy, I., & Khang, J. Q. G. (2020). Designing a mobile application to improve user's productivity on computer-based productivity software. Journal Of Advanced Research In Dynamical And Control Systems.
11. Morawski, K., Ghazinouri, R., Krumme, A., Lauffenburger, J. C., Lu, Z., Durfee, E., ... & Choudhry, N. K. (2018). Association of a smartphone application with medication adherence and blood pressure control: the MedISAFE-BP randomized clinical trial. JAMA internal medicine, 178(6), 802-809.
12. "Doctor on Demand - apps on Google Play." https://play.google.com/store/apps/details?id=com.doctoro nd emand.android.patient&hl=en (accessed Jul. 27, 2024).
13. "WebMD: Symptom Checker - apps on Google Play." https://play.google.com/store/apps/details?id=com.webmd. an droid&hl=en (accessed Jul. 27, 2024).

Authors Introduction

Mr. Abdulrahman Salmo Alhamada

He is currently an undergraduate student, pursuing his study in Bachelor of Computer Science (Hons) Mobile Computing and Networking from Institute of Computer Science and Digital Innovation (ICSDI) at UCSI University, Malaysia.

Dr. Ghassan Saleh ALDharhani

He is an assistant professor who is currently teaching at UCSI University. His research interests include cybersecurity, ethical hacking and machine learning. He is also the Head of Digital Innovation Department, a TTT Certified HRDF Trainer and Cyber Security Instructor.

Dr. Kasthuri Subaramaniam

She received her Bachelor's and Master's degrees in Computer Science from University of Malaya. She earned her Ph.D. in Informatics from Malaysia University of Science & Technology. Her research interests include human-computer interaction, human personality types, augmented reality, e-learning, mobile commerce and e-commerce.

Dr. Raenu Kolandaisamy

He is the director at the Institute of Computer Science & Digital Innovation, UCSI University. He has a PhD and masters degrees in Computer Science. He holds a bachelors degree in computer system engineering. His research interests are in VANET, wireless networking & security, IoT and cloud computing.

Developing a Body Posture Detection for Fitness

Kai Xuan Chong
Institute of Computer Science and Digital Innovation, UCSI University, Cheras, 56000 Kuala Lumpur, Malaysia

Abdul Samad Bin Shibghatullah
College of Computing & Informatics, Universiti Tenaga Nasional, 43000 Kajang, Selangor, Malaysia

Kasthuri Subaramaniam
Department of Decision Science, Faculty of Business and Economics, University of Malaya, 50603 Kuala Lumpur, Malaysia

Chit Su Mon
School of Mathematical and Computer Sciences, Heriot-Watt University Malaysia, 62200 Putrajaya, Malaysia

Email: 1002058415@ucsiuniversity.edu.my, abdul.samad@uniten.edu.my, s_kasthuri@um,edu.my, c.mon@hw.ac.uk

Abstract

The Body Posture Detection System for Fitness is an innovative technology that aims to enhance exercise technique and movement patterns by providing real-time monitoring and feedback. It utilizes computer vision and machine learning algorithms to track and analyze body movements during fitness. The system's ability to provide immediate feedback and correction significantly improves exercise effectiveness and user safety. It also comes with the user-friendliness of the system, potentially incorporating a Graphical User Interface (GUI) for easy navigation and accessibility. To address time and budget constraints, the approach of this research will choose the Rapid Application Development (RAD) Model as the system development approach. This methodology consists of four phases which are: requirement planning, user design, rapid construction, and transition, and each is accompanied by deliverables. These efforts are aimed at enhancing the functionality and usability of the Body Posture Detection System for Fitness while addressing user needs and optimizing fitness training experiences.

Keywords: OpenCV, Tkinter, Body Posture Detection, Graphical User Interface

1. Introduction

Body Posture Detection is a system that for user can maintaining their body posture during fitness. It is an Artificial Intelligent technology and computer vision that are able to detecting and understanding the position and structure of human body parts from user input. With the Artificial Intelligent and Computer Vision techniques it owns the ability like a human to “see” and understand the visual information but it has the ability to generate an appropriate output based on it understanding [1]. This Technology are able to do a wide range of task such as image classification by assigning a label or category to an image, image localization by identifying and locating multiple objects in an image, image segmentation by partitioning an image, and key-point detection by identifying specific points or landmarks in an image [12]. So, with these technologies it greatly helps a lot in the industry of Object detection, facial recognition, hand gesture recognition, body posture detection, and so on [1], [13]. In this project, Body Posture Detection for Fitness is greatly suitable for using these technologies for detecting those joint parts.

In the worlds nowadays, Fitness become a popular exercise for people, this exercise it brings a lot of benefits such as reducing the risk of cardiovascular disease and metabolic syndrome, controlling weight, strengthening bones and muscles, and promoting mental health. Doing exercise is important but without a proper posture then it will be dangerous, because incorrect posture of exercise during fitness will lead to serious of injuries in the part of muscle and ligaments [2]. Then, it is importance to own a personal trainer during fitness, he or she can oversee your workout session and monitor your posture is proper or not, but unfortunately not everyone is able to hire a personal trainer or guidance due to many reasons such as financial issue, social distance issue, and more.

So, with this situation the Artificial Intelligent based application will become an important role in identifying the user posture during fitness and giving the instant feedback to help the user improve their body posture during fitness. Also, with the power of Artificial Intelligence and Computer Vision, the user is able to know the correct posture during fitness by the real-time guidance and help improving exercise safety and effectiveness (Fig. 1).

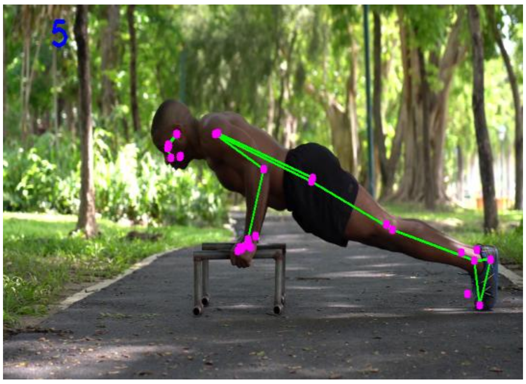


Fig. 1. Body Posture Detection

2. Literature Review

2.1. Comparison of Body Posture Detection Systems

A comparison of different methodologies in various body posture detection systems are shown in Table 1.

Table 1. Comparison of Body Posture Detection Systems

Paper Name (Publication Year)	Techniques	Advantages
Smart exercise counter using computer vision [3] - 2022	OepnCV, MediaPipe	Good Accuracy: Computer vision and Artificial Intelligence algorithms, the system can accurately detect the joint of body and count exercises, ensuring precise monitoring of fitness activities.
Body Posture Detection and Motion Tracking using AI for Medical Exercises and Recommendation System [4] - 2022	MediaPiPe Django	More Accurate compare to the others object detection model: MobileNetV2-SSD: 97.95% MediaPipe: 98.61%
A wearable-based posture recognition system with AI-assisted approach for healthcare IoT [5] - 2022	Multi-Posture Recognition (MPR) Cascade-AdaBoosting-CART (CACT)	High Accuracy and Reliability: The WMHPR system, with its combination of MPR and CACT posture recognition algorithms, achieves a high level of accuracy (98.06% True Positive Rate) and reliability (0.19% False Positive Rate). This is a significant advantage in healthcare applications, ensuring accurate recognition of human postures.
A Survey on Artificial Intelligence in Posture Recognition [6] - 2023	Sensor-Based Recognition Vision-Based Recognition RF-Based Recognition	Sensor-based recognition is cost-effective but requires sensor wear in real-time. Vision-based recognition is highly accurate but affected by lighting, background, and privacy concerns. RF-based recognition is contactless and sensitive to environmental changes but can be influenced by how the human body interacts with RF signals.

2.2. Approach of Body Posture Detection

To develop a body posture detection, recognition, or classification system, it required a Recognition Approaches. This Recognition Approaches is allowing the system to capture or collect the data of body movements. So, in these studies we have found the main approaches of recognition which is Vision-Based.

Vision-based methods for body posture recognition utilize visual information, typically derived from images or video frames, to analyze and interpret body postures. These techniques employ computer vision to extract pertinent features from the visual data, facilitating the recognition and understanding of human body poses. Here are some techniques based on vision-based methods in body posture recognition:

- Image Processing and Analysis: Vision-based methods often commence with image processing techniques to detect and extract significant features from the images. This could include identifying key body joints, angles between joints, and the overall configuration of the human body [7].
- Skeletonization Techniques: Vision-based methods may use skeletonization techniques to represent the human body as a simplified skeleton, emphasizing key joints and connections. This skeletonized representation facilitates the recognition of different body postures [8].

In conclusion, vision-based methods offer a powerful and versatile approach to body posture recognition, leveraging the rich visual information available in images or video frames for accurate and real-time analysis of human poses.

2.3. The Study of Proper Posture in Fitness

The posture you maintain while performing exercises plays a crucial role in determining the effectiveness of your workout and preventing injuries while fitness. So, there are some of the correct postures for some common exercises such as push-ups, sit-ups, and dumbbell lifting.

Push-ups are a foundational movement in strength training. However, they are often performed incorrectly. The tips are, first, set your hands at a distance slightly wider than shoulder-width apart on the ground. Second, your hands should be angled in a way that feels comfortable to you. Your feet should be set up in a way that feels right and comfortable and in balance. Third, your body should form one giant straight line from the top of your head down through your heels. Last, your butt shouldn't be sticking way up in the air or sagging [9].

Sit-ups is one of the exercises to strengthening your core. The tips are, first, lie on your back, with knees bent and feet firmly on the ground. Second, place your hands behind your head, or place each hand on the opposite shoulder, so arms cross over the front of your body. Last, slowly lift your upper body off the ground, keeping your chin tucked into your chest as you do so [10].

Dumbbell lifting is a great way to build muscle and improve strength. The tips are, first, warm up your body before you start lifting. Second, choose the right weight for you. Too much weight will only give you fewer sets before tiring out while too lightweight dumbbells won't target your muscles enough. Third, lift slowly, you want to feel the weight of the dumbbell as you lift it up and down and make your muscles work with the utmost time. Last, cool down once you're done with your workout [11].

3. Analysis

The survey has collected a lot of useful data and allows us to understand the preferences of those who took part in the survey for gathering information about the Body Posture Detection System for Fitness. Also, the most of participants are young adults aged between 18 and 25. The participants are in different races, most of them are come from Chinese, and from various university backgrounds.

Moreover, most of the participants during fitness are doing different poses, the most common usually are such as weightlifting, squatting, and push-ups. Also, the aim of the fitness goals of these participants is intend to weight loss and muscle gain.

After that, participants also noticed the negative effects of incorrect postures, such as bringing some downsides like muscle strain and joint problems. So, there is a positive attitude toward the use of a Body Posture Detection System, suggesting a potential interest in using technology to improve workout safety and effectiveness.

Also, participants have mentioned the specific features of the system should consist of real-time monitoring, accuracy, user-friendly interfaces, and educational resources. Most participants support the idea of integrating body posture detection into their fitness routines, particularly at home. The strong demand for educational resources within the system reflects the participants' dedication to learning and enhancing their fitness skills.

4. Synthesis

The figure below (Fig. 2) shown the use cases of the Body Posture Detection System for Fitness. This system used to detect the body movement and body posture of user. So, in this situation the system first will detect the user body key point through the algorithm and allow the algorithm work with the camera. After that, the computer vision is able to see people through camera, and starting to detect the body posture/movement of the user. Then it will estimate the body posture of the user by using key point, after that it will check the body posture that user perform is correct or not. Last, it will display the result.

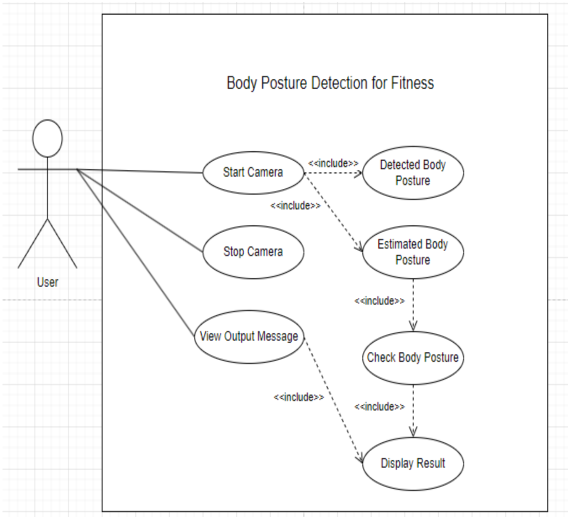


Fig. 2. Use Case Diagram of Body Posture Detection System

5. Evaluation

5.1. Proposed Method

This system is used to detect and calculate different types of exercises such as squats, push-ups, and weightlifting. Below are the main components of this program:

- App Class: This is the main class of the system which use to run the GUI and pose estimation. It initializes the application, sets up the GUI layout with buttons and labels, and initializes variables used for pose estimation and exercise counting.
- MyVideoCapture Class: This class is used to capture video the web camera. It has methods to get the current frame and release the video source when it is no longer needed.

- Update methods: The update_pose_estimation, update_squat_estimation, update pushUp_ estimation and update_lift_estimation methods in the App class are used to update the GUI with the current frame and pose estimation results from the video source. These methods are called repeatedly after a certain delay to provide real-time pose estimation.

These methods work together to create a dynamic fitness application that uses pose estimation to detect the user's posture in each frame of the video, calculates joint angles to determine the posture, and computes the number of exercises based on changes in posture. The results are displayed in the GUI, allowing the user to start and stop posture estimation and select the type of exercise.

6. Conclusion

All in all, the development of Body Posture Detection for Fitness with Python has successfully been developed. This system was developed with Python and its libraries such as cv2 library, mediapipe library, tkinter library. Users expressed interest in features like real-time monitoring, accuracy, user-friendly interfaces, and educational resources, showing a strong demand for integrating such a system into their fitness routines, especially at home.

References

1. Gollapudi, S., & Gollapudi, S. (2019). Artificial intelligence and computer vision. Learn Computer Vision Using OpenCV: With Deep Learning CNNs and RNNs, 1-29.
2. Kumar, A., Kumar, A., Maggu, S., Kapoor, B., Kaushik, A. K., & Kalonia, S. (2024). Human Posture Detection and Correction Using MediaPipe and OpenCV. In International Conference on Computing and Communication Systems for Industrial Applications (pp. 135-147). Singapore: Springer Nature Singapore.
3. M.Senthil Kumar, G.Sai Sunitha, M.Sashrutha, and R.Ramya. (2022). Smart Exercise Counter Using Computer Vision. International Journal of Creative Research Thoughts (IJCRT), Volume 10(5).
4. Patil, A., Rao, D., Utturwar, K., Shelke, T., & Sarda, E. (2022). Body posture detection and motion tracking using AI for medical exercises and recommendation system. In ITM web of conferences (Vol. 44, p. 03043). EDP Sciences.
5. Hong, Z., Hong, M., Wang, N., Ma, Y., Zhou, X., & Wang, W. (2022). A wearable-based posture recognition system with AI-assisted approach for healthcare IoT. Future Generation Computer Systems, 127, 286-296.
6. Jiang, X., Hu, Z., Wang, S., & Zhang, Y. (2023). A survey on artificial intelligence in posture recognition. Computer modeling in engineering & sciences: CMES, 137(1), 35.
7. Lin, C. B., Dong, Z., Kuan, W. K., & Huang, Y. F. (2020). A framework for fall detection based on OpenPose skeleton and LSTM/GRU models. Applied Sciences, 11(1), 329.
8. Kothari, S. (2020). Yoga pose classification using deep learning.
9. Kamb, S., & Kamb, S. (2024). How to Do a (Proper) Push-up | Nerd Fitness. Nerd Fitness.
10. Ces, A. C. C. C. (2020, September 30). How to do Sit-Ups. Verywell Fit.
11. Lemus, R. (2022, August 14). How to Lift Dumbbells Properly - Simple Fitness Hub. Simple Fitness Hub.
12. Jalil, A., Kolandaisamy, R., Subaramaniam, K., Kolandaisamy, I., & Khang, J. Q. G. (2020). Designing a mobile application to improve user's productivity on computer-based productivity software. Journal Of Advanced Research In Dynamical And Control Systems.
13. Susanto, I. C., Subaramaniam, K., & Shibghatullah, A. S. B. (2022, January). Gesturenomy: Touchless Restaurant Menu Using Hand Gesture Recognition. In Proceedings of International Conference on Artificial Life and Robotics (pp. 229-236).

Authors Introduction

Mr. Kai Xuan Chong

He is currently an undergraduate student, pursuing his study in Bachelor of Computer Science (Honours) from Institute of Computer Science and Digital Innovation (ICSDI) at UCSI University, Malaysia.

Dr. Abdul Samad Bin Shibghatullah

He received his Bachelor's degree in Accounting from Universiti Kebangsaan Malaysia, M.Sc. degree in Computer Science from Universiti Teknologi Malaysia and Ph.D. in Computer Science from Brunel University, UK. He is currently an Associate Professor at the Universiti Tenaga Nasional, Kajang, Malaysia.

Dr. Kasthuri Subaramaniam

She received her Bachelor's and Master's degrees in Computer Science from University of Malaya. She earned her Ph.D. in Informatics from Malaysia University of Science & Technology. Her research interests include human-computer interaction, human personality types, augmented reality, e-learning, mobile commerce and e-commerce.

Dr. Chit Su Mon

She is an Assistant Professor at the School of Mathematical and Computer Sciences at Heriot-Watt University Malaysia. She holds a PhD in Computing from Sunway University, completed in 2022. Her research interests include Human-Computer Interaction (HCI), Haptics and Assistive Technology.

Medical Mate: Healthcare and Medical Chat Bot

Harris Hue Chee Kin
Institute of Computer Science and Digital Innovation, UCSI University, Cheras, 56000 Kuala Lumpur, Malaysia

Javid Iqbal Thirupattur
Department of Computing and Information Systems, School of Engineering and Technology, Sunway University, 47500 Selangor, Malaysia

Kasthuri Subaramaniam
Department of Decision Science, Faculty of Business and Economics, University of Malaya, 50603 Kuala Lumpur, Malaysia

Shaik Shabana Anjum
School of Computer Science, Faculty of Innovation & Technology, Taylor's University, 47500 Selangor, Malaysia

Email: 1001955666@ucsiuniversity.edu.my, javidt@sunway.edu.my, s_kasthuri@um,edu.my, anjum.shaik@taylors.edu.my

Abstract

The proposed idea is to develop a web-based medical chat bot called "Medical Mate" that will be placed by every patient's bedside, serving as a companion, and providing necessary care and support. It aims to offer convenience and accessibility for patients while easing the workload of hospital staff. Usability tests, employing mixed methods research, were conducted to ensure the chat bot's design and functionalities meet user requirements. Medical Mate pivoted to a web-based application using HTML, CSS, JS, jQuery, and MySQL. Ultimately, the goal of Medical Mate is to be a helpful and reliable companion for patients during their hospital stay, offering care and reducing the burden on medical personnel.

Keywords: Healthcare, Chat bot, Medical mate, Graphical user interface

1. Introduction

The concept of chat bots has been around for some time, with their main purpose being to engage in conversations with users by responding to their questions based on predefined sets of intents. The vision for Medical Mate is to be a specialized chat bot designed specifically for the medical field, acting as a personal healthcare companion for patients in hospitals. Unlike traditional medical chat bots that offer basic information and customer services, Medical Mate aims to go beyond by providing constant support, companionship, and healthcare advice. It will be available 24/7, offering patients someone to interact with whenever they need it. The chat bot will engage in general conversations and small talk, acknowledging the emotional aspect of healthcare and providing social interaction for patients during their hospital stay.

Additionally, Medical Mate will leverage AI, ML, NLP, and NLU to offer healthcare guidance and serve as a communication facilitator between patients and medical staff. The inclusion of food ordering functionality aims to add convenience, while special attention is given to assisting introverted patients, providing a non-judgmental environment for them to communicate and access healthcare information comfortably. The overall goal is to enhance patient experience, promote engagement, and optimize healthcare delivery within the hospital setting [1], [8].

Existing chat bots in the medical field are limited in functionalities, often focused on basic tasks like customer service and simple health diagnostics. There is a need for improvement to create a more comprehensive system that can cater to a larger number of patients, providing better assistance and companionship during their hospital or clinic visits. The challenge lies in ensuring advanced and well-trained Natural Language Processing (NLP) to avoid misinterpretation of user questions and prevent providing false or irrelevant responses. Additionally, some chat bots lack appealing graphical user interfaces (GUI) and intuitive navigation, which may lead to disengaged users, particularly those unfamiliar with technology or older generations not exposed to such systems. Creating a user-friendly and visually appealing interface is essential for successful chat bot adoption [5].

This study aims to develop a new chat bot application system for hospitals or clinics (medical field) which will focus on being the patient's personal healthcare assistance or companions but can also be used by doctors to update patient's details. The chat bot for hospital or clinics aims to understand its patients and to provide the best responses about their current health. A chat bot can be implemented at every side of the bed to reduce human efforts so that not all nurses or doctors have to be present in the room unless of emergency. The chat bot is also available every hour so patients can access them whenever they need assistance or companions.

2. Review of Literature

The idea of chat bots originated with Alan Turing in 1950, who pondered whether machines could think like humans. With the rise of Artificial Intelligence (AI), conversational systems gained attention, and chat bots emerged as natural language user interfaces for data providers. The first-ever chat bot, "ELIZA," was created in 1966 by Joseph Weizenbaum, scanning user input for keywords to provide relevant responses. It was later improved and enhanced by Kenneth Colby, becoming "PARRY." Decades later, Richard Wallace developed the chat bot A.L.I.C.E, utilizing artificial intelligence markup language (AIML) for more advanced and improved responses to user queries [2], [3].

2.1. Types of Chat Bot

In the current technological era, there are three main types of chat bots: Menu/Button-based, Keyword Recognition-based, and Contextual Chat Bots. Menu/Button-based chat bots are basic and limited, offering pre-defined options for users to select. They are easy to use but may not provide accurate or personalized responses. Keyword Recognition-based chat bots use NLP algorithms to identify specific keywords in user input and can handle more complex questions, but they may still struggle with full message understanding and misinterpretation. Contextual chat bots are technologically advanced, utilizing AI, ML, and NLP to fully understand user input, learn from interactions, and provide personalized responses. However, they are expensive and complex to develop and maintain [4].

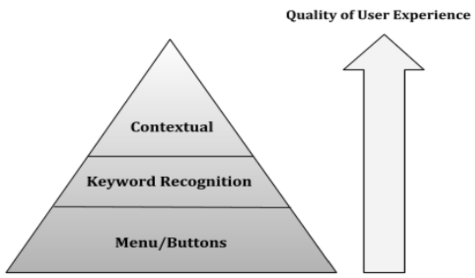


Fig. 1. Preference of chat bots

We can conclude from Fig. 1, that contextual chat bots provide a superior user experience compared to menu/button-based chat bots. Their sophistication and advanced capabilities allow them to answer complex user questions effectively. In contrast, menu/button-based chat bots are limited in dialogue and cannot handle intricate inquiries, resulting in a less enjoyable user experience [4].

2.2. Adoption of Chat Bots

The adoption of chat bots has significantly increased in various industries, transforming tasks and customer service. Businesses and organizations use chat bots as virtual personal assistants to assist customers with general inquiries, simplifying tasks and providing help 24/7. Chat bots can learn from vast amounts of data, known as big data, allowing them to understand patterns in user interactions and offer personalized responses. Advanced chat bots use machine learning to collect and analyse data, enhancing their decision-making algorithms and increasing efficiency and productivity. Their 24/7 accessibility and ability to handle multiple conversations at once make them cost-effective for businesses, reducing the need for additional customer service staff. While chat bots have advantages, they are best suited for handling simpler inquiries, and more complex questions may still require human intervention. Businesses can continue to improve customer experiences by enhancing chat bot capabilities and leveraging their ability to learn from big data. Chat bots are very common in a lot of industries these days [4], [6]. Table 1 illustrates examples of chat bots in different industries and their specific functionalities.

Table 1. The planning and control components

Chat bots	Functionalities
Online shopping	Personalized recommendation, reply to customers with basic information until owner is online, available 24/7
Education	Teach users and provide resourceful information, 24/7 availability for students to ask questions.
Customer service	Reply customers with basic information until someone is available to entertain them, available 24/7
Banking	Customer service 24/7 and can query any information regarding their bank details
Virtual assistant (phone)	Carry out simple task and provide the best possible answer, can be user's friend whenever they feel bored
healthcare	Provide basic healthcare advice, scheduling doctor's appointment, managing medications and prescriptions
Travelling (airlines)	Help manage customers flight, basic customer services

In summary, chat bots offer a convenient and efficient way for users to obtain information through voice and text recognition skills. They provide quick and helpful responses, making them cost-effective for businesses compared to hiring human customer service workers. Chat bots offer 24/7 availability, handle multiple conversations simultaneously, and use big data to provide personalized recommendations and solutions [7]. They are easily scalable to meet growing demand without sacrificing quality. Overall, chat bots simplify user experiences, improve customer satisfaction, and have the potential to transform various industries as technology continues to advance and they continue to learn and adapt.

2.3. Human Computer Interaction (HCI)

In recent years, Human-Computer Interaction (HCI) has made significant advancements, particularly in the design of Graphical User Interfaces (GUIs). HCI focuses on the interaction between humans and computers, and it plays a crucial role in the development of programs or applications that require GUIs [12], [13]. GUIs are essential for simplifying the interaction between users and various digital devices, and they must be designed with HCI principles in mind to ensure usability and user-friendliness. HCI becomes even more important when creating chat bots, as it facilitates natural and intuitive conversations between users and the system. HCI professionals use various disciplines to understand user requirements and preferences, incorporating these insights into interface design to enhance the user experience. In conclusion, a well-designed GUI is crucial for HCI, and it is considered the future of technology companies, especially in the context of chat bots and natural language interactions [2], [9].

2.4. Chat bots in the medical and healthcare industries

Chat bots offer valuable solutions to the challenges of accessing mental health services, providing instant help and support with increased efficiency. They can enhance empathy through human-like interactions, improving therapeutic effects and treatment plan compliance. Healthcare-focused chat bots improve user engagement and usability, offering convenient and personalized interactions for individuals who may find traditional healthcare methods intimidating. By providing 24-hour availability and more human-like automated messages, chat bots address accessibility issues and offer a promising avenue for efficient and effective healthcare support, potentially revolutionizing healthcare service delivery [7], [8], [10].

3. Methodology

Data gathering and prototype creation were conducted, followed by testing among chosen participants to gather feedback for improving the design and knowledge proficiency of the chat bot. Test scenarios included asking random questions to assess response time and accuracy of the chat bot. The research approach chosen for the development of the medical chat bot, Medical Mate, is mixed methods research, which combines both quantitative and qualitative research methods. Quantitative research focuses on discovering patterns and trends using numerical data, while qualitative research seeks to understand people's experiences and perspectives through verbal or narrative data [11]. The aim of the research is to obtain feedback on the chat bot's interface, conversation ability, and overall usefulness. By using both quantitative and qualitative methods, a comprehensive understanding of the chat bot's performance and user experience can be achieved. The data collection will be split into 2 different approaches, and these are focus group interview and questionnaire survey.

3.1. How does a chat bot work

A chat bot functions by processing user input and attempting to match it with the most appropriate intent from its predefined set. In the case of Medical Mate, it has a predefined set of intents stored in a Json file, which it uses to train for keyword-recognition. This allows Medical Mate to identify users' input and provide the best possible response based on the identified keywords.

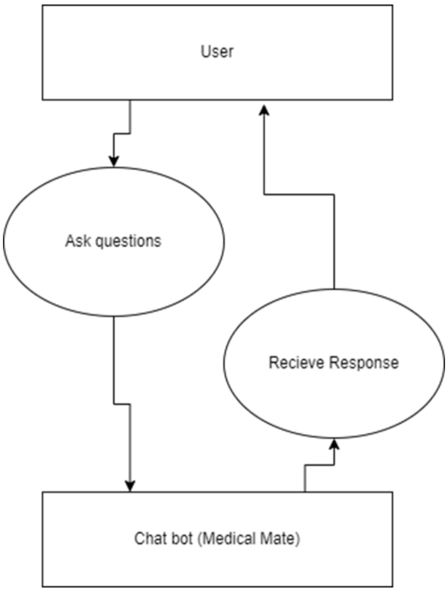


Fig. 2. Chat bot use case diagram

A simple use case diagram (Fig. 2) representing the design of the Medical Mate. This use case diagram also shows interaction between user and the Medical Mate. This diagram shows that the user input which is to ask the chat bot questions and it will process the questions and identify the keyword to provide a response to the user when it is done.

4. Discussion

Medical Mate, being a prototype, has some acknowledged limitations. These limitations include a restricted set of features and functionalities and an incompletely trained chat bot. However, the research methodologies and continuous user testing employed during its development have provided valuable insights into user experience and feedback. Despite its current limitations, Medical Mate has been instrumental in gathering valuable information about user interactions and perceptions. It is primarily a keyword-recognition based chat bot with a predefined set of intents used to provide responses to users. Additionally, the chat bot is programmed using php and MySQL instead of python, which further contributes to its limitations.

5. Conclusion

The research findings demonstrate the potential and benefits of a medical chat bot system like Medical Mate. Participants found the chat bot to be a good idea with room for improvements. The system was praised for its user-friendly design and ease of navigation. Medical Mate's aim to answer medical questions, provide companionship, and assistance was well-received by users. To further enhance its performance, integrating Python programming and comprehensive training with datasets is recommended. The ultimate goal is for Medical Mate to become a reliable and intelligent healthcare companion for patients, reducing human workload and improving the overall hospital experience. The hope is that with continuous development and improvements, Medical Mate can achieve its purpose and become an indispensable tool in the medical field.

References

1. Dahiya, M. (2017). A tool of conversation: Chatbot. International journal of computer sciences and engineering, 5(5), 158-161.
2. Følstad, A., & Brandtzæg, P. B. (2017). Chatbots and the new world of HCI. interactions, 24(4), 38-42.
3. Huang, X., & CIS, A. (2021). Chatbot: design, architecture, and applications. Univesity of Pennsylvania: School of Engineering and Applied Science, Pennsylvania, 1.
4. Nayani, A. R., Gupta, A., Selvaraj, P., Singh, R. K., & Vaidya, H. (2024). Chatbot Detection with the Help of Artificial Intelligence. International Journal of Multidisciplinary Innovation and Research Methodology, ISSN, 2960-2068.
5. Aggarwal, H., Kapur, S., Bahuguna, V., Nagrath, P., & Jain, R. (2022). Chatbot to map medical prognosis and symptoms using machine learning. In Cyber Security and Digital Forensics: Proceedings of ICCSDF 2021 (pp. 75-85). Springer Singapore.
6. Reshmi, S., & Balakrishnan, K. (2018). Empowering Chatbots with Business Intelligence by Big Data Integration. International Journal of Advanced Research in Computer Science, 9(1).
7. Gökçearslan, S., Tosun, C., & Erdemir, Z. G. (2024). Benefits, challenges, and methods of artificial intelligence (AI) chatbots in education: A systematic literature review. International Journal of Technology in Education, 7(1), 19-39.
8. Skjuve, M., & Brandtzæg, P. B. (2018). Chatbots as a new user interface for providing health information to young people. Youth and news in a digital media environment–Nordic-Baltic perspectives..
9. Traubinger, V., & Gaedke, M. (2024, June). Interaction Design Patterns of Web Chatbots. In International Conference on Web Engineering (pp. 309-317). Cham: Springer Nature Switzerland.
10. Holmes, S., Moorhead, A., Bond, R., Zheng, H., Coates, V., & McTear, M. (2019, September). Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces?. In Proceedings of the 31st European Conference on Cognitive Ergonomics (pp. 207-214).
11. Kurniadi, A., Wulandari, F., & Arinnisa, H. (2024). Usability Evaluation of Hospital Information System Applications Based on The System Usability Scale. Procedia of Engineering and Life Science, 6, 44-49.
12. Jalil, A., Kolandaisamy, R., Subaramaniam, K., Kolandaisamy, I., & Khang, J. Q. G. (2020). Designing a mobile application to improve user's productivity on computer-based productivity software. Journal Of Advanced Research In Dynamical And Control Systems.
13. Subaramaniam, K., Ern-Rong, J. L., & Palaniappan, S. (2021). Interface designs with personality types: an effective e-learning experience. EVERGREEN – Joint Journal of Novel Carbon Resources Sciences & Green Asia Strategy. Vol. 08, Issue 03, pp. 618-627.

Authors Introduction

Mr. Harris Hue Chee Kin

He is currently an undergraduate student, pursuing his study in Bachelor of Science (Hons) in Computing from Institute of Computer Science and Digital Innovation (ICSDI) at UCSI University, Malaysia.

Dr. Javid Iqbal Thirupattur

He earned his Ph.D. in Information and Communication Technology from the National Energy University Malaysia. He holds a Bachelor's degree in Computer Science and Engineering and a Master's degree in Communication Systems, both from Dr. M.G.R. Educational and Research Institute. His research interests encompass Augmented Reality, and Virtual Reality.

Dr. Kasthuri Subaramaniam

She received her Bachelor's and Master's degrees in Computer Science from University of Malaya. She earned her Ph.D. in Informatics from Malaysia University of Science & Technology. Her research interests include human-computer interaction, human personality types, augmented reality, e-learning, mobile commerce and e-commerce.

Dr. Shaik Shabana Anjum

She received a B.Eng. degree in Computer Science and Engineering and M.Eng. degree (Hons.) in Software Engineering from Anna University, Chennai, India and a Ph.D degree in Computer Science from the University of Malaya, Malaysia. Her research interests include the IoT, wireless sensor networks and radio frequency identification.

Crime Identification System for Campus Safety and the Threat of Suspicious Student Conduct

Wong Zhen Bang
Institute of Computer Science and Digital Innovation, UCSI University, Cheras, 56000 Kuala Lumpur, Malaysia

Kay Hooi Keoy
Center for Business Informatics and Industrial Management, UCSI Graduate Business School, UCSI University, Cheras, 56000 Kuala Lumpur, Malaysia

Kasthuri Subaramaniam
Department of Decision Science, Faculty of Business and Economics, University of Malaya, 50603 Kuala Lumpur, Malaysia

Sellappan Palaniappan
Help University, Bukit Damansara 50490 Kuala Lumpur, Malaysia

Oras Baker
Faculty of Computing, Ravensbourne University London, SE10 0EW, United Kingdom

Email: 1001954107@ucsiuniversity.edu.my, alankeoy@gmail.com, s_kasthuri@um,edu.my, sellappan.p@help.edu.my, O.alhassani@rave.ac.uk

Abstract

Ensuring campus safety is of paramount importance for educational institutions. With the increasing prevalence of crime and the potential threat of suspicious student conduct, there is a need for an effective crimes identification system. This paper aims to propose a comprehensive system that utilizes advanced technologies, such as video surveillance, data analytics, and behavioral monitoring, to detect and mitigate potential threats on campus. The proposed Crimes Identification System (CIS) integrates various components to enhance campus safety. These cameras capture real-time footage that is then processed using advanced video analytics algorithms. The algorithms analyze the footage for suspicious activities, unusual behavior patterns, or known criminal profiles. The system incorporates a centralized monitoring and alert system that notifies security personnel immediately upon detecting any suspicious behavior or potential threats. This enables a swift response to mitigate the risks and ensure the safety of the campus community.

Keywords: Campus safety, Suspicious student conduct, Behavioral monitoring, Threat detection

1. Introduction

In the 21st century, school crime has become a serious problem and now the number of school crimes keeps increasing. Especially among secondary school and university students. Suppose the student should learn and study in a safety environment, but due to the stress and did not have proper way to release it. Students might do illegal activities to release stress. According to [1] stated that increased stress among university students leads to several problems with their academic performance, including depression, a lack of interest in going to class, difficulty understanding the material, and suicidal thoughts because of low self-esteem. And might lead student to do illegal activities such as substance abuse, vandalism, and more [2].

Crime identification system refers to the methods and processes used by law enforcement agencies to identify and track criminal activity [9]. This can include various technologies such as biometric identification, facial recognition, and fingerprint analysis, as well as more traditional methods like eyewitness testimony and forensic analysis [3]. For now, crime identification systems have become more advanced, the AI technology combine with security camera. The accuracy of AI surveillance hugely improved compared to traditional security cameras. AI algorithms can analyze and process vast amounts of data from security cameras in real-time, making it easier to identify potential security threats and respond quickly. Also, AI surveillance have better object recognition, AI-powered cameras can detect specific objects and recognize patterns, such as vehicles or individuals, which can help identify potential security risks [4].

The research of Kakadiya proposed that using the AI (artificial intelligence) with the CCTV (Closed circuit television system) for detection of dangerous weapon and illegal items. Kakadiya trained the AI surveillance by using the CNNs (Convolutional Neural Network) algorithm and found out the result is very high accuracy. Convolutional Neural Networks (CNNs) are a type of deep learning algorithm that have several benefits which perform well in image and video recognition. CNNs have automated feature extraction which is able to automatically extract relevant features from an image or video, reducing the need for manual feature engineering. Below figure is the example result that AI surveillance using CNNs algorithm [5].

2. Literature Reviews

Campus safety is an important issue for educational institutions all around the world. With developments in technology and data analytics, crime identification systems have become important resources for improving campus security. The purpose of this literature review is to investigate existing studies on crime detection systems for campus safety and their effectiveness in reducing criminal activity [6].

Developing a prototype crime identification system is a difficult and time-consuming procedure that involves careful planning, research, and implementation [10]. These systems are critical tools for improving law enforcement's ability to detect and capture offenders. Based on the material supplied by researcher Kranthi Kumar, this article covers the essential stages required in developing a prototype crime detection system. These phases will include a variety of topics, such as technology selection, system development, testing, and implementation [7].

Another popular supervised learning method is Support Vector Machines (SVM), which is particularly successful for classification problems. SVM is frequently used to determine an ideal hyperplane in the feature space that successfully divides different groups or categories. When dealing with non-linear correlations and complex feature interactions, k-NN is extremely helpful. It has adaptable categorization boundaries and can adjust to changing crime patterns and distributions. Although k-NN is simple to build and provides relatively accurate results, it may encounter challenges when dealing with high-dimensional data or inconsistent class distributions. Choosing a suitable value for k and maximizing the distance measure are essential parts of effectively implementing k-NN [8].

Table 1. Comparison of the current Facial Emotion Recognition system [7]

Name	Version Type	Accuracy	Feature extraction Algorithm
YOLOv3	320	33.50%	Convolutional Neural Networks
FaceVideo	API v1.0	48.35%	Convolutional Neural Networks
Cognitive Services: Face	API v1.0	52.61%	N/A
YOLOv4	N/A	43.8%	Convolutional Neural Networks

The comparison results are shown in Table 1 of current Facial Emotion Recognition system.

3. Methodology

A crime identification system is a multidisciplinary effort that use technology to detect and understand criminal activity based on behavioral and visual indicators [11]. A mix of quantitative methods, historical research, and questionnaire methodology will be used to create an effective crime identification system.

Quantitative research methods can be used to collect and analyze numerical data on illegal items with different objects. This data can then be used to train machine learning algorithms that can recognize items and classify whether it whether is illegal item or not in real-time. In this project, MS COCO will be used to perform object identification. MS COCO (Microsoft Common Objects in Context) is a large-scale image dataset containing 328,000 images of everyday objects and humans. MS COCO dataset was created by Microsoft to train and evaluate many state-of-the-art object detection and segmentation models. The images were collected from the internet and are of different sizes and qualities. MS COCO is a widely used dataset in the field of object recognition and has been used for training and evaluating deep learning models for crime identification systems.

In the structure of the crime identification system, historical research methodologies are also used to understand the cultural and social variables surrounding criminal activities and how they have appeared through different behavioral and visual indicators. This technique can give important insights into the evolution of criminal patterns and law enforcement strategies by evaluating historical criminal cases and researching how crime has been investigated and documented.

4. Analysis and Discussion

An online survey was used to gather information from a sample of 100 university students for the purpose of this study. The participants were chosen at random from a broader group of college students from different universities. The purpose of the questionnaire was to collect necessary data about the research subject and the application of crime identification system technology in the educational field. The data gathered will be carefully examined in order to reach relevant conclusions and learn insightful information. It is anticipated that the sample size of 100 students will give an accurate representation of how students view and feel about the use of crime identification system technology in the educational setting.

Furthermore, the survey results revealed that 9% of respondents were unaware of the campus safety risks.

This research emphasizes the need to raise awareness and give information about campus safety measures, processes, and resources. It is critical that all members of the campus community are well-informed and prepared to deal with possible safety issues.

5. Synthesis

5.1. System Architecture Design

The system will required import some library to build up the object identification. The required essential library is IPython.display, cv2_imshow, Image, and %matplotlib inline. To make a successful detection, the system will adjust the screen to 640*480 to make sure the system go smoothly, when the resolution raised up it will cause the system crash down due to the limitation of the hardware

The "yolov4-csp.weights" neural network model was trained and tested using the MS COCO dataset. The authors used a convolutional neural network (CNN) to get an accuracy of 0.67002. This pre-trained model is intended to accurately recognize and categorize objects in pictures or video frames. Darknet is a framework for both training and running pictures or video frames through learned neural networks. It is a popular open-source deep learning framework for applications like object recognition and categorization. The Crime Identification System may use the power of the "yolov4-csp.weights" model for item recognition and classification by loading it into Darknet.

After loading the model, the next step is to define a function that identify the object based on the classification. The "darknet_helper" will help the system enable to detect the object, it will scan the object and find out the object feature to compare with the system database, then it will provide the result to inform the user the object name. Furthermore, it encodes the data from the image path by changing the picture's colour to RGB from BGR using cv2.COLOR_BGR2RGB. This extracts the RGB data from the images and places it in the encoded List array. After the image's encoding procedure is complete, the camera is turned on to capture the real image from reality and compare it to the make image.

5.2. User Interface Design and System Flow

Fig. 1. Example of Illegal Object Detected

Firstly, the user will operate the system, the system will start scanning once detected the object. When the system detects an illegal object (Fig. 1), the system will proceed to next step.

Fig. 2. Alert Sound Occur when Detect Illegal Object

Thus, the system will perform alert sound when detected an illegal object (Fig. 2).

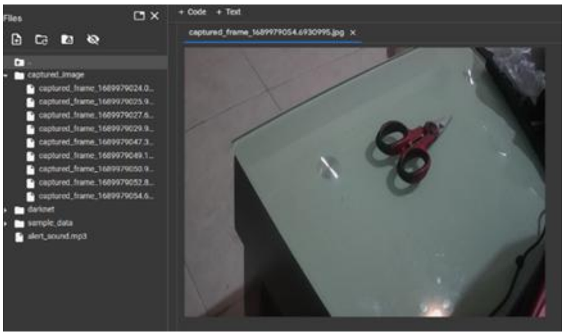


Fig. 3. Illegal Object Record

The system will immediately capture the image as evidence and store it in the database (Fig. 3).

6. Evaluation

System testing is conducted to ensure that the system works as intended.

6.1. Results and summary of the Usability Testing

After conducting usability testing on the Crime Identification System, the following results have been gathered:

- Accuracy of Object Detection
  Participants noted that the accuracy of the system varied depending on the lighting conditions, camera angle, and facial expressions of the individuals in the video clips. Some participants also reported that the system had difficulty recognizing more subtle or nuanced emotions.
- User Interface Design
  Several participants commented the system did not have GUI and make the system become not user friendly. Hence, the GUI will bring into future development.
- Respond Time
  Some participants expressed dissatisfaction and reduced efficiency because the system was slow to respond to their instructions.

7. Conclusion

In conclusion, the development of the Crime Identification System has been successfully achieved, meeting the aims and objectives outlined during the initial planning phase. This system is developed using Python and its libraries, including the cv2 library for image processing and analysis. Additionally, the development process incorporates the Rapid Application Development methodology to ensure a responsive and efficient approach.

References

1. Rana, A., Gulati, R., & Wadhwa, V. (2019). Stress among students: An emerging issue. Integrated Journal of Social Sciences, 6(2), 44-48.
2. Hingson, R., Heeren, T., Winter, M., & Wechsler, H. (2005). Magnitude of alcohol-related mortality and morbidity among US college students ages 18–24: Changes from 1998 to 2001. Annu. Rev. Public Health, 26(1), 259-279.
3. Khoei, T. T., & Singh, A. (2024). A survey of Emotional Artificial Intelligence and crimes: detection, prediction, challenges and future direction. Journal of Computational Social Science, 1-44.
4. Tsapin, D., Pitelinskiy, K., Suvorov, S., Osipov, A., Pleshakova, E., & Gataullin, S. (2024). Machine learning methods for the industrial robotic systems security. Journal of Computer Virology and Hacking Techniques, 20(3), 397-414.
5. Kakadiya, R., Lemos, R., Mangalan, S., Pillai, M., & Nikam, S. (2019, June). Ai based automatic robbery/theft detection using smart surveillance in banks. In 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 201-204). IEEE.
6. Muneer, S., Farooq, U., Athar, A., Ahsan Raza, M., Ghazal, T. M., & Sakib, S. (2024). A Critical Review of Artificial Intelligence Based Approaches in Intrusion Detection: A Comprehensive Analysis. Journal of Engineering, 2024(1), 3909173.
7. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.
8. Dupré, D., Krumhuber, E. G., Küster, D., & McKeown, G. J. (2020). A performance comparison of eight commercially available automatic classifiers for facial affect recognition. Plos one, 15(4), e0231968.
9. Kolandaisamy, R., Li, A., Subaramaniam, K., Jalil, A. B., & Kolandaisamy, I. (2022). Implementation Face Recognition Attendance System For Higher Education Institution. Specialusis Ugdymas, 1(43), 4120-4133.
10. Subaramaniam, K., Ern-Rong, J. L., & Palaniappan, S. (2021). Interface designs with personality types: an effective e-learning experience. EVERGREEN – Joint Journal of Novel Carbon Resources Sciences & Green Asia Strategy. Vol. 08, Issue 03, pp. 618-627, September 2021.
11. Kin, H. C., Subaramaniam, K & Shibghatullah A.S. (2024, February). App Alert System for Smart Phones. In App Alert System for Smart Phones. The 2024 International Conference on Artificial Life and Robotics (ICAROB2024), J:COM HorutoHall, Oita, Japan, 2024, pp. 1066–1069.

Authors Introduction

Mr. Wong Zhen Bang

He is currently an undergraduate student, pursuing his study in Bachelor of Computer Science in Data Science with Honours from Institute of Computer Science and Digital Innovation (ICSDI) at UCSI University, Malaysia.

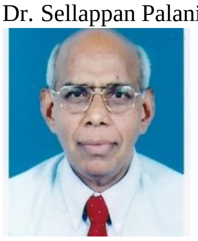
Dr. Kay Hooi Keoy

He is an Associate Professor at UCSI Graduate Business School, UCSI University. He is holding the post of Deputy Dean and Centre for Business Informatics and Industrial Management at Graduate Business School. His extensive expertise encompasses a broad spectrum within the realms of business and management.

Dr. Kasthuri Subaramaniam

She received her Bachelor's and Master's degrees in Computer Science from University of Malaya. She earned her Ph.D. in Informatics from Malaysia University of Science & Technology. Her research interests include human-computer interaction, human personality types, augmented reality, e-learning, mobile commerce and e-commerce.

Dr. Sellappan Palaniappan



He is a Professor at Help University, Malaysia. He received the Ph.D. degree in Interdiscplinary Information Science from Universiti of Pittsburg. He has more than 40 years of teaching experience, authored several IT/CS books, published more than 100 journal/ conference papers and keynote speaker in conferences His research areas include machine learning, artificial intelligence, data science and cybersecurity.

Dr. Oras Baker

He is an Associate Professor and Head of Masters in Cyber Security and Cyber Security Management at Ravensbourne University London, UK. With 25 years of distinguished experience spanning academia, research, and industry, he specialises in Artificial Intelligence, Software Engineering, Cyber Security, Data Mining, and Machine Learning.

Prototype of MixVRT Which Is a Visual Regression Testing Tool That Highlights Layout Defects in Web Pages

Naoki Aridome*, Tetsuro Katayama*, Yoshihiro Kita†,
Hisaaki Yamaba*, Kentaro Aburada*, and Naonobu Okazaki*
* *Department of Computer Science and Systems Engineering, Faculty of Engineering, University of Miyazaki*
1-1 Gakuen-kibanadai nishi, Miyazaki, 889-2192 Japan
†*Department of Information Security, Faculty of Information Systems, Siebold Campus, University of Nagasaki*
1-1-1 Manabino, Nagayo-cho, Nishi-Sonogi-gun, Nagasaki, 851-2195 Japan
E-mail: aridome@earth.cs.miyazaki-u.ac.jp, kat@cs.miyazaki-u.ac.jp, kita@sun.ac.jp,
yamaba@cs.miyazaki-u.ac.jp, aburada@cs.miyazaki-u.ac.jp, oka@cs.miyazaki-u.ac.jp

Abstract

As a method for detecting layout defects in web pages, image-based visual regression testing is proposed. However, it has the problem that it takes time to detect unintended layout differences that are not based on HTML code changes. This paper proposes a prototype of MixVRT which is a tool to detect layout defects in web pages. It is a visual regression testing tool that highlights layout defects in web pages. From evaluation experiments, the time find to layout defects can be reduced.

Keywords: MixVRT, web page, layout defects, visual regression testing, HTML code

1. Introduction

In recent years, web pages need to be updated more frequently from the perspective of user experience (UX) and search engine optimization (SEO) [1]. Here, in changing to update a web page, there is a problem that intended layout differences that are based on HTML code changes and unintended layout differences that are not based on HTML code changes are mixed, and it takes time to distinguish between the two. In this paper, differences in layout unintended by the developer are referred to as layout defects.

To solve the problem, there is visual regression testing [2]. In the visual regression testing for web pages, the web pages before and after changes are placed side by side and differences in layout are highlighted to make it easier to spot differences in layout. Many existing studies of visual regression testing compare only images before and after changes and highlight all layout differences, including differences in the developer's intended layout [3][4].

However, finding layout defects is time-consuming because it requires comparing the differences between all the displayed layouts and the HTML codes to determine whether the differences are in line with the developer's intent. Therefore, to reduce the time to find layout defects in web pages, this paper develops a prototype of MixVRT which is a visual regression testing tool that highlights layout defects in web pages.

2. Functions of MixVRT

In this chapter, we describe the functions of MixVRT. It detects differences by comparing images and detects changes due to HTML codes changes on the web pages before and after changes and compares them to detect layout defects. In addition, it highlights the following three areas by surrounding them with red or green borders.

- Difference Areas
  Difference Areas is the areas removed from the web page before the change and the areas added to the web page after the change by comparing the images of the web pages before and after changes.
- Modified Areas
  Modified Areas is the areas of screen elements affected by changes in either the body element or the style element, or both, of the HTML code. These changes are detected by comparing the HTML codes of the web pages before and after changes. In this paper, Modified Areas is the areas of the intended layout differences.

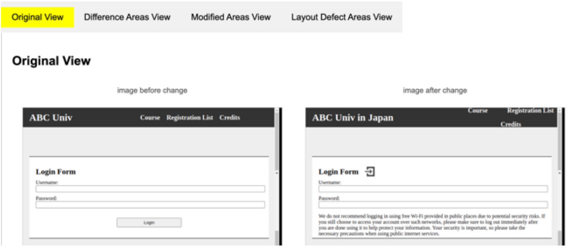


Fig. 1. Appearance of Original View

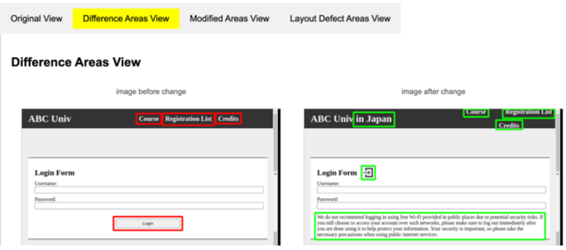


Fig. 2. Appearance of Difference Areas View

- Layout Defect Areas
  Layout Defect Areas is the areas of layout differences excluding the Modified Areas from the Difference Areas. In this paper, Layout Defect Areas is the areas of the unintended layout differences. In addition, Layout Defect Areas shows areas where it is possible that layout defects occur.

3. Appearance and Structure of MixVRT

In this chapter, we describe the appearance and structure of MixVRT. It takes as input URLs of the web page before and after changes. Then, for each of the four views shown below that make up the appearance of MixVRT, it generates a PNG format image corresponding to each view and outputs the image to a web page running on the local server for display.

- Original View
  The before and after images of the web page are displayed side by side on the left and right. The appearance of Original View is shown in Fig. 1.
- Difference Areas View
  The before and after images of the web page are displayed side by side on the left and right, with the Difference Areas highlighted with red or green borders. The appearance of Difference Area View is shown in Fig. 2.
- Modified Areas View
  The before and after images of the web page are displayed side by side on the left and right, with the Modified Areas highlighted with red or green borders. The appearance of Modified Area View is shown in Fig. 3.

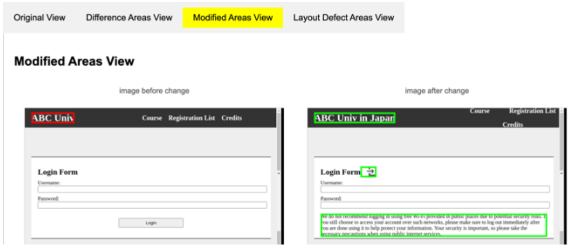


Fig. 3. Appearance of Modified Areas View

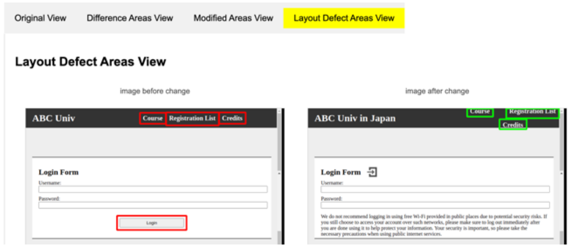


Fig. 4. Appearance of Layout Defect Areas View

- Layout Defect Areas View
  The before and after images of the web page are displayed side by side on the left and right, with the Layout Defect Areas highlighted with red or green borders. The appearance of Layout Defect Areas View is shown in Fig. 4.

The structure of MixVRT is shown in Fig. 5. MixVRT consists of the following five processing sections.

- Data Manager
  Data Manager is responsible for communicating data between each processing section, exchanging data with other processing sections, and maintaining data. The data in this paper are the images and HTML codes of the web pages before and after processing.
- Data Getter
  Data Getter acquires the images of the web page from the URLs of the web page received as input using Selenium WebDriver [5]. It also acquires the HTML codes of the web page using Requests [6]. These processes are performed for both the URL of the web page before and after changes to obtain the before and after changes images of the web page. The acquired data is output to the Data Manager.

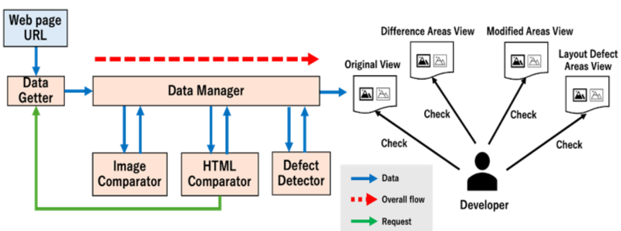


Fig. 5. Structure of MixVRT

- Image Comparator
  Image Comparator compares the before and after changes images of the web page received from the Data Manager in pixel units and generates the before and after changes images of the web page, highlighting the Difference Areas. The generated images are output to the Data Manager.
- HTML Comparator
  HTML Comparator compares the before and after changes HTML codes of the web page received from the Data Manager line by line using difflib [7] and generates a difference code. Then, based on the difference code, it adds HTML codes that encloses the changed HTML elements with red or green borders to both the before and after changes HTML codes. The added before and after HTML codes are then rendered as web pages, and by capturing images of the rendered web pages, the before and after changes images of the web page, highlighting the Modified Areas, are obtained. The obtained images are output to the Data Manager.
- Defect Detector
  Defect Detector detects Layout Defect Areas by comparing the images highlighting the Difference Areas with the images highlighting the Modified Areas and generates the before and after changes images of the web page, highlighting the Layout Defect Areas. The generated images are output to the Data Manager. The algorithm for detecting Layout Defect Areas compares the red or green borders of the Difference Areas and Modified Areas, respectively, and if the overlapping areas of the borders is less than 50%, it is determined to be Layout Defect Areas.

4. Evaluation of MixVRT

In this chapter, to evaluate the usefulness of MixVRT, we experiment to compare the time to find Layout Defect Areas and the detection accuracy of Layout Defect Areas between using MixVRT and conventional methods of image-based visual regression testing. The subjects are four students of University of Miyazaki.

In preparation for the experiment, we first prepare two web pages. Let these be Web page α and Web page β before the change, respectively. Web page α consists of 261 lines of code and 17 CSS classes, while Web page β consists of 472 lines of code and 28 CSS classes. Next, we change each of these web pages, each containing three Layout Defect Areas with layout defects. After the changes, Web page α consists of 275 lines of code and 20 CSS classes, and Web page β consists of 489 lines of code and 31 CSS classes. Let these be Web page α and Web page β after the change, respectively. In addition, we prepare an experimental file for the subjects to record the Layout Defect Areas when they found it.

4.1 *Experimental Methods*

We experiment using Web page α before and after changes, and Web page β before and after changes, under the following two cases:

- Case A: Using MixVRT, we measure the time to find Layout Defect Areas. It is the time from when a subject views the Layout Defect Areas View to when all Layout Defect Areas are found.
- Case B: Using conventional methods, we measure the time to find Layout Defect Areas. It is the time from when a subject views the Difference Areas View to when all Layout Defect Areas are found.

Here, the subjects can view the HTML codes before and after changes of the Web page for the experiment.

4.2 *Evaluation of time to find Layout Defect Areas*

Table 1 shows the time to find Layout Defect Areas in the two cases. From Table 1, it reduced an average of 12 m 59.5 s (about 91.8%) of time in Web page α and reduced an average of 33 m 41.5 s (about 97.9%) of time in Web page β. Here, we considered the reason why the reduction rate is higher for Web page β than for Web page α. This is because Web page β has more lines of code and CSS classes than Web page α (see Section 4), which increases the time to check the HTML codes. As a result, using the conventional methods takes significantly longer to find Layout Defect Areas compared to MixVRT. From the above, MixVRT can reduce the time to verify whether layout defects are in the areas where layout changes have occurred by checking the HTML codes.

Table 1. The time to find Layout Defect Areas

	Web page *α*		**Web page *β***	
Subjects	**Case A**	**Case B**	**Case A**	**Case B**
1	-	18m 41s	52s	-
2	-	9m 38s	45s	-
3	1m 36s	-	-	32m 43s
4	44s	-	-	36m 7s
Average	1m 10s	14m 9.5s	48.5s	34m 25s
Diff	**12m 59.5s**		**33m 41.5s**	

4.3 *Evaluation of detection accuracy of Layout Defect Areas*

Table 2 shows the number of Layout Defect Areas over-detected and under-detected in Case B. In evaluation experiments, the subjects sometimes over-detect or under-detect Layout Defect Areas using the conventional method. In contrast, by using MixVRT, the subjects were able to detect Layout Defect Areas without over- or under-detection. In summary, MixVRT is more useful for detecting layout defects than the conventional methods of image-based visual regression testing.

Table 2. The number of Layout Defect Areas over-detected and under-detected in Case B

Web page	Layout Defect Areas	Subjects	Over-detected	Under-detected
α	3	1	1	1
		2	0	1
β	3	3	0	0
		4	0	1

5. Conclusion

In this paper, we have developed a prototype of MixVRT which is a visual regression testing tool that highlights layout defects in web pages, to reduce the time to find layout defects in web pages.

We have experimented with two Web page. As a result, MixVRT can reduce the time by 91.8% and 97.9%. We also have confirmed that MixVRT detects all Layout Defect Areas without over- or under-detection. On the other hand, in the conventional methods, Layout Defect Areas have been over-detected or under-detected. Consequently, MixVRT is useful in reducing the time to find layout defects compared to the conventional method.

The following are future issues in this paper.

- Support for multiple browsers
- Support for different window sizes before and after web page changes

References

1. CAES Office of Information Technology: "Why is web content so important?", https://oit.caes.uga.edu/why-is-web-content-so-important/. (Accessed 2024-12-10)
2. Visual regression testing: awesome-regression-testing, https://github.com/mojoaxel/awesome-regression-testing. (Accessed 2024-12-10)
3. Haruto Tanno, Yu Adachi, Yu Yoshimura, et al.: "Region-based Detection of Essential Differences in Image-based Visual Regression Testing", Journal of Information Processing, Vol.28, pp.268-278, 2020.
4. Akihiro Tsukakoshi: "Empirical Evaluation of a Visual Regression Testing Difference Detection Method Using Hierarchical Structures of GUI Elements" (in Japanese), Software Quality Profession, Vol.5, pp.207-214, 2020.
5. Selenium WebDriver: "WebDriver", Selenium Documentation, https://www.selenium.dev/ja/documentation/webdriver/. (Accessed 2024-12-10)
6. PyPI: "Requests", https://pypi.org/project/requests/. (Accessed 2024-12-10)
7. Python Software Foundation: "difflib − Helper for computing deltas", Python 3.13.1 Documentation, https://docs.python.org/ja/3/library/difflib.html. (Accessed 2024-12-10)

Authors Introduction

Tetsuro Katayama

Tetsuro Katayama received a Ph.D. degree in engineering from Kyushu University, Fukuoka, Japan, in 1996. From 1996 to 2000, he has been a Research Associate at the Graduate School of Information Science, Nara Institute of Science and Technology, Japan. Since 2000 he has been an Associate Professor at the Faculty of Engineering, Miyazaki University, Japan. He is currently a Professor with the Faculty of Engineering, University of Miyazaki, Japan. His research interests include software testing and quality. He is a member of the IPSJ, IEICE, and JSSST.

Yoshihiro Kita

Yoshihiro Kita received a Ph.D. degree in systems engineering from the University of Miyazaki, Japan, in 2011. He is currently an Associate Professor with the Faculty of Information Systems, University of Nagasaki, Japan. His research interests include software testing and biometrics authentication.

Hisaaki Yamaba

Hisaaki Yamaba received the B.S. and M.S. degrees in chemical engineering from the Tokyo Institute of Technology, Japan, in 1988 and 1990, respectively, and the Ph D. degree in systems engineering from the University of Miyazaki, Japan in 2011. He is currently an Assistant Professor with the Faculty of Engineering, University of Miyazaki, Japan. His research interests include network security and user authentication. He is a member of SICE and SCEJ.

Kentaro Aburada

Kentaro Aburada received the B.S., M.S, and Ph.D. degrees in computer science and system engineering from the University of Miyazaki, Japan, in 2003, 2005, and 2009, respectively. He is currently an Associate Professor with the Faculty of Engineering, University of Miyazaki, Japan. His research interests include computer networks and security. He is a member of IPSJ and IEICE.

Naonobu Okazaki

Naonobu Okazaki received his B.S, M.S., and Ph.D. degrees in electrical and communication engineering from Tohoku University, Japan, in 1986, 1988 and 1992, respectively. He joined the Information Technology Research and Development Center, Mitsubishi Electric Corporation in 1991. He is currently a Professor with the Faculty of Engineering, University of Miyazaki since 2002. His research interests include mobile network and network security. He is a member of IPSJ, IEICE and IEEE.

Proposal of a Method for Automatic Fill-in Fields Detection and for Labels Assignment to Generate Electronic Forms

Yuya Kimura*, Tetsuro Katayama*, Yoshihiro Kita†, Hisaaki Yamaba*, Kentaro Aburada*, and Naonobu Okazaki*

* *Department of Computer Science and Systems Engineering, Faculty of Engineering, University of Miyazaki*
1-1 Gakuen-kibanadai nishi, Miyazaki, 889-2192 Japan

†*Department of Information Security, Faculty of Information Systems, Siebold Campus, University of Nagasaki*
1-1-1 Manabino, Nagayo-cho, Nishi-Sonogi-gun, Nagasaki, 851-2195 Japan

E-mail: kimura@earth.cs.miyazaki-u.ac.jp, kat@cs.miyazaki-u.ac.jp, kita@sun.ac.jp, yamaba@cs.miyazaki-u.ac.jp, aburada@cs.miyazaki-u.ac.jp, oka@cs.miyazaki-u.ac.jp

Abstract

The digitalization of forms is being promoted. One of the effective ways to manage contents filled in fields is using electronic forms. Several tools have been developed to generate them automatically. However, when you use a paper form, the layout of the original form may change, and it takes time to generate electronic one because it is necessary to place fill-in fields on an electronic form by dragging with a mouse. This paper proposes a method for automatic fill-in fields detection and labels assignment to reduce time required to place fill-in fields without changing the layout. The proposed method can reduce the time to place fill-in fields.

Keywords: image processing, electronic form, fill-in fields, area coordinates obtainment, label assignment.

1. Introduction

In April 2019, Electronic Books Maintenance Act was amended [1], and the digitalization of them is being promoted. On the other hand, about half of Japanese companies still use paper form for their management system [2]. Forms can be digitized by taking pictures with a scanner or camera. This make it easy, but it is required to see contents filled in fields on manual. Using electronic forms is one of the ways to manage them efficiently. Some tools have been developed to generate electronic forms automatically. However, when using these tools, it has the following two problems.

- The layout of original forms may change.
- It takes time to place fill-in fields because it requires placing fill-in fields on an electronic form by dragging with a mouse.

To solve them, this paper proposes a method for automatic fill-in fields detection and for labels assignment to generate electronic forms. There is a previous study to detect rectangular fill-in fields to recognize the category of form documents and to store them [3]. Based on it, our proposed method is performed image-based fill-in fields detection.

2. Proposed method

In this chapter, we present the structure and the behavior of the proposed method. Fig. 1 shows the structure of the proposed method. It consists of four parts: Area coordinates obtainment part, Texts information obtainment part, Labels assignment part, and Files output part. It takes as input a form image, and outputs two form images highlighted fill-in fields and a JSON file including fill-in fields coordinates and labels.

An image of a digitized document and an image of an electronic document are received as input. Electronic documents refer to documents created as digital information, such as Word or text files. Digital documents refer to documents that have been scanned from paper documents and saved in PDF or image format. The image of the digitized document is assumed to be an image taken with a smartphone. In this paper, the fill-in fields obtained coordinates with the proposed method are called area coordinates, and ones indicated rectangle are called rectangular areas, and ones indicated underline are called underlined areas.

2.1. Area coordinates obtainment part

Area coordinates obtainment part obtains area coordinates for rectangular areas and underlined areas. The coordinates of rectangular area acquire four xy coordinates as rectangular area coordinates for each vertex. The coordinates of underlined area acquire two xy coordinates as underlined area coordinates for both endpoints. To obtain area coordinates, image processing is performed using OpenCV. Furthermore, we use DeblurGANv2 [4] to remove blur in the image. DeblurGANv2 is a tool that applies generative adversarial network (GAN) to deblurring. The following is the flow of Area coordinates obtainment part behavior.

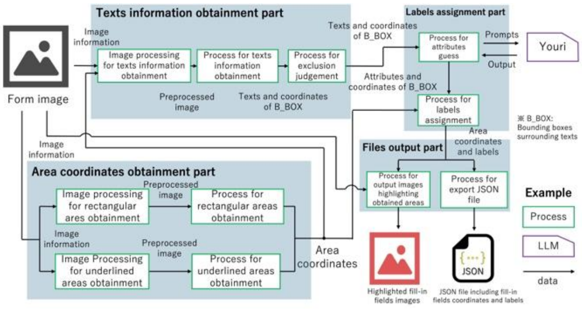


Fig. 1. The structure of the proposed method

(i) Process image to obtain rectangular area coordinates with imread, cvtColor, GaussianBlur, threshold, getStructuringElement, and dilate function of OpenCV.
(ii) Deblur the form image with DeblurGANv2.
(iii) Detect rectangle and obtain rectangular area coordinates with findContour function of OpenCV.
(iv) Process image to obtain underlined area coordinates with imread, cvtColor, and threshold function of OpenCV.
(v) Detect underline and underlined area coordinates with createFastLineDetector function of OpenCV Contrib, which is one of OpenCV's extra modules.

2.2. Texts information obtainment part

Texts information obtainment part obtains texts and the coordinates of the bounding box surrounding them as texts information using optical character recognition (OCR). Using OCR engine is Tesseract [5] with Japanese trained data and dictionary. In recognizing texts, the accuracy of texts recognition can be improved by binarizing the image. In this process, some pixels of fill-in fields affect the binarization for character recognition optimization. To solve this problem, the most frequent RGB values among all pixels in the image are obtained as the background color's RGB values. The following is the flow of Texts information obtainment part behavior.

(i) Obtain the area coordinates from Area coordinates obtainment part.
(ii) Load the image with imread function of OpenCV.
(iii) Deblur the loaded image with DeblurGANv2.
(iv) Obtain the most frequent RGB values with histogram function of Numpy.
(v) Fill the area with the background color referenced from the obtained area coordinates.
(vi) Generate an image after process (v).
(vii) Obtain texts information with OCR from the output image of process (vi) with Tesseract.

2.3. Labels assignment part

Labels assignment part assigns labels to obtained area coordinates. By assigning labels, it is possible to add information required for validation. In this study, one is selected from three data types: "date", "number", or "string". It uses Youri [6], Japanese large language model (LLM) to predict data types as attributes of texts to be filled in from the recognized texts. The data type is linked as the label by referencing the coordinates of the bounding box surrounding the texts. The following is the flow of Labels assignment part.

(i) Input prompt including obtained texts to guess the data type of them into Youri.
(ii) Obtain the answers from Youri as attributes.
(iii) Initialize labels for all area coordinates as string.
(iv) Assign labels of rectangular areas to attributes of the bounding box if the x, y coordinates of the bounding box center are greater than ones of the rectangular region center.
(v) Assign labels of underlined areas to attributes of the bounding box if the x, y coordinates of the bounding box center are greater than ones of the rectangular region center.

2.4. Files output part

Files output part outputs two form images and one JSON file. One of the form images highlights rectangular areas, and the other highlights underlined areas. The JSON file includes area coordinates and labels of them. The following is the JSON file composition.

- A unique ID for each area coordinate.
- The label assigned to each area coordinate.
- The object indicating each area coordinate.

The following is the flow of File output part behavior.

(i) Obtain area coordinates and labels of them from Labels assignment part.
(ii) Generate two form images A and B with copy function of Python.
(iii) Generate three random integer values and determine a random RGB color from each color space with randint function of Numpy.
(iv) For form image A, draw rectangles colored the RGB color generated in process (iii) by referencing the rectangular area coordinates in the form image with drawContours function of OpenCV.
(v) For form image B, draw green lines by referencing the underlined area coordinates in the form image with line function of OpenCV.
(vi) Draw the values of id key and label key at the top left of each area coordinate in the form images with putText function of OpenCV.
(vii) Save form image A as highlighting rectangular area image and form image B as highlighting underlined area image with imwrite function of OpenCV.

3. Application Example

In this chapter, we confirm that the proposed method works correctly by using an implemented prototype. A part of an example of the output of JSON file is shown in List. 1, and a part of an example of a highlighted rectangular areas image is shown in Fig. 2.

From List. 1, The label is “string” for the rectangular area with id 4, and the label is “number” for the one with id 5. From Fig. 2, The fourth rectangular area is a fill-in field for "品名", and the fifth rectangular area is a fill-in field for "数量". Here, “品名” is the Japanese word for the item name, and “数量” is the Japanese word for the quantity. For each label in each rectangular area is “string” and “number”. Therefore, we have confirmed they are the correct labels for “品名” and “数量”. And, we have confirmed their area coordinates are correct.

4. Discussion

In this chapter, to evaluate the usefulness of the proposed method, we experiment with six students of University of Miyazaki. For the experiment, we used a GUI tool to generate electronic forms.

There are three measuring times: execution time, placement time, and total time. The following is the meaning of each measurement time.

- Execution Time: the time taken by the program from the start to the end of its execution.
- Placement Time: the time to place all fill-in fields on the electronic form with the GUI tool.
- Total Time: sum of execution time and placement time.

The steps of the experiment regarding evaluation placing fill-in fields with the GUI tool on an electronic form are shown below.

(i) The experimenter prepares two form images: a form image of an electronic document as Form image A, and a form image of a digitized document as Form image B.
(ii) The experimenter instructs the participants to place the fill-in fields in the electronic form with the GUI tool and starts measuring the placement time.
(iii) The participants place the fill-in fields as instructed

Fig. 3 shows the two form images used in the experiment. There are 86 fill-in fields to place in Form image A, and 54 ones to place in Form image B.

List. 1. A part of an example of output JSON file

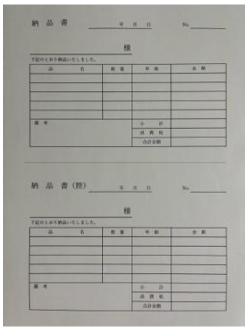

```
{
  "id": 4,
  "label": "string",
  "coords": {
    "top_left": {
      "x": 275,
      "y": 817
    },
    "bottom_left": {
      "x": 275,
      "y": 903
    },
    "bottom_right": {
      "x": 1008,
      "y": 903
    },
    "top_right": {
      "x": 1008,
      "y": 817
    }
  }
},
{
  "id": 5,
  "label": "number",
  "coords": {
    "top_left": {
      "x": 1016,
      "y": 817
    },
    "bottom_left": {
      "x": 1016,
      "y": 903
    },
    "bottom_right": {
      "x": 1308,
      "y": 903
    },
    "top_right": {
      "x": 1308,
      "y": 817
    }
  }
},
```


Fig. 2. A part of an example of a highlighted rectangular areas image

4.1. Evaluation on accuracy of area coordinates

Table 1 shows the precision rate and recall rate about area coordinates. We consider the reason why the recall rate of A is lower than that of B is because there are six filled-in fields to be filled in that is not indicated by either the rectangle or the underline, and it cannot detect them.

a. Form image A



b. Form image B

Fig. 3. Form images used in the experiment

Table 1. The precision rate and recall rate about area coordinates obtainment (percent)

Precision Rate		Recall Rate	
Form image A	**Form image B**	**Form image A**	**Form image B**
91.49	81.82	93.48	100.00

4.2. Evaluation on accuracy of assigned labels

Table 2 shows the precision rate and recall rate about assigned labels. The area coordinates which have correct labels is correct too about the coordinates. We considered the reason why two rates in Form image A are less than that of Form image B is that a few texts are not recognized, so an attribute of other texts are assigned as the label incorrectly.

4.3. Evaluation on time to place fill-in fields with a GUI tool

There are two cases below for participants to place fill-in fields in the electronic form.

- Case A: using only the GUI tool to Form image A, and using it with proposed method to Form image B.
- Case B: using only the GUI tool to Form image B, and using it with proposed method to Form image A.

Table 3 shows the average of each measurement time for each form image. From Table 3, it reduced 3 minutes 54.5 seconds (about 31.4 %) of Form image A, and reduced 2 minutes 16.4 seconds (about 27.0 %) of Form image B. It shows that proposed method can reduce the time to place fill-in fields in an electronic form.

Table 2. The precision rate and recall rate about assigned labels (percent)

Precision Rate		Recall Rate	
Form image A	**Form image B**	**Form image A**	**Form image B**
72.22	73.13	81.25	90.74

Table 3. The measurement times it took the participants to place fill-in fields in electronic form

Average Measurement time	Form image A		Form image B	
	Case A	**Case B**	**Case A**	**Case B**
Execution	-	2:54.3	2:28.4	-
Placement	12:26.0	5:37.2	3:40.6	8:25.4
Total	12:26.0	**8:31.5**	**6:09.0**	8:25.4

5. Conclusion

This paper has reduced the time to place fill-in fields on an electronic form to generate it without changing the layout. We experimented with two form images. As a result, the proposed method can reduce the time by 31.4% and 27.0 %. Therefore, the proposed method has solved the two problems.

The future works are as follows.

- Improving recognition accuracy of fill-in fields and texts when using a colored form.
- Judging between filled-in fields themselves and fields indicating the contents to be filled in.
- Dealing with various shooting environment.

References

1. National tax agency, Special website for the Electronic Bookkeeping System (in Japanese) (Accessed 2024-12-13)
2. Keyman's Net, Status of digitalization of forms (2022) / Part 1 (in Japanese) (Accessed 2024-12-13)
3. Kuo-Chin Fan, Yuan-Kai Wang and Mei-Lin Chang, "Form document identification using line structure based features," Proceedings of Sixth International Conference on Document Analysis and Recognition, Seattle, WA, USA, 2001, pp. 704-708
4. O. Kupyn, T. Martyniuk, J. Wu and Z. Wang, "DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 8877-8886
5. GitHub, Tesseract, *tesseract*, https://github.com/tesseract-ocr/tesseract (Accessed 2024-12-13)
6. rinna, Rinna releases "Youri 7B", a Japanese language pre-learning model for Llama 2 (in Japanese) (Accessed 2024-12-13)

Authors Introduction

Yuya Kimura

Yuya Kimura received the Bachelor's degree in engineering (computer science and systems engineering) from the University of Miyazaki, Japan in 2024. He is currently a Master's student in Graduate School of Engineering at the University of Miyazaki, Japan. His research interests image processing.

Tetsuro Katayama

Tetsuro Katayama received a Ph.D. degree in engineering from Kyushu University, Fukuoka, Japan, in 1996. From 1996 to 2000, he has been a Research Associate at the Graduate School of Information Science, Nara Institute of Science and Technology, Japan. Since 2000 he has been an Associate Professor at the Faculty of Engineering, Miyazaki University, Japan. He is currently a Professor with the Faculty of Engineering, University of Miyazaki, Japan. His research interests include software testing and quality. He is a member of the IPSJ, IEICE, and JSSST.

Yoshihiro Kita

Yoshihiro Kita received a Ph.D. degree in systems engineering from the University of Miyazaki, Japan, in 2011. He is currently an Associate Professor with the Faculty of Information Systems, University of Nagasaki, Japan. His research interests include software testing and biometrics authentication.

Hisaaki Yamaba

Hisaaki Yamaba received the B.S. and M.S. degrees in chemical engineering from the Tokyo Institute of Technology, Japan, in 1988 and 1990, respectively, and the Ph D. degree in systems engineering from the University of Miyazaki, Japan in 2011. He is currently an Assistant Professor with the Faculty of Engineering, University of Miyazaki, Japan. His research interests include network security and user authentication. He is a member of SICE and SCEJ.

Kentaro Aburada

Kentaro Aburada received the B.S., M.S, and Ph.D. degrees in computer science and system engineering from the University of Miyazaki, Japan, in 2003, 2005, and 2009, respectively. He is currently an Associate Professor with the Faculty of Engineering, University of Miyazaki, Japan. His research interests include computer networks and security. He is a member of IPSJ and IEICE.

Naonobu Okazaki

Naonobu Okazaki received his B.S, M.S., and Ph.D. degrees in electrical and communication engineering from Tohoku University, Japan, in 1986, 1988 and 1992, respectively. He joined the Information Technology Research and Development Center, Mitsubishi Electric Corporation in 1991. He is currently a Professor with the Faculty of Engineering, University of Miyazaki since 2002. His research interests include mobile network and network security. He is a member of IPSJ, IEICE and IEEE.

A Study on Methodology of Measurement for the Physical Burden on Preschool Children

Sachiko Kido
Interdisciplinary graduate school of Agriculture and Engineering, University of Miyazaki, 1-1, Gakuen Kibanadai-Nishi, Miyazaki, 889-2192, Japan

Hiroki Tamura
Faculty of Engineering, University of Miyazaki, 1-1, Gakuen Kibanadai-Nishi, Miyazaki, 889-2192, Japan

E-mail: z321t01@student.miyazaki-u.ac.jp, htamura@cc.miyazaki-u.ac.jp

Abstract

Measuring the burden placed on the body of a preschool child is difficult and has rarely been measured using motion capture. In this paper, the AnyBody Modelling System was used to verify whether it is possible to calculate the burden on the preschool child's body using motion capture. The lumbar burden value for preschool children was calculated, defining the burden on the lumbar region as the burden on the body as a whole. Few measurements have been made on young children with the AnyBody Modelling System. Therefore, the validity of the pre-school child figure was verified by comparing the lumbar burden with that of adult male. In addition, by setting the adult male data based on the preschool child's height and weight and comparing the calculated values with the preschool child's actual values, the possibility of simulating children's body burden in various movements using adult body models in the future was examined.

Keywords: Preschool child, Physical activity, Lumbar burden,

1. Introduction

In contemporary Japan, where the decline in children's physical fitness [1] is an issue, it is very meaningful to analyze the physical activity of children, especially pre-school children, by various methods.

Methods for measuring physical activity include continuous heart rate recording, electromyography, pedometers, calorie counters and accelerometers [2],[3],[4]. However, analysis with motion capture data is less common in pre-school children. Possible reasons for this include the inability to accurately read body movements due to their small size, and the fact that children often move differently from adults and cannot respond to these movements.

In this paper, we have examined whether it is possible to calculate the burden on the body of pre-school children using motion capture, using the AnyBody Modelling System [5], a musculoskeletal mechanics analysis software that is widely used worldwide. The lumbar burden values for pre-school children were calculated based on the definition that ‘the burden on the lumbar region is the burden on the body’. Few measurements have been made on pre-school children using the ‘AnyBody Modelling System’. Therefore, the lumbar burden values of adult males were compared to those of preschool children to verify the validity of the preschool children's values. In addition, the adult male data was set to the height and weight of a girl, and the calculated values were compared and examined with those of a pre-school child to verify whether it would be possible to simulate the burden on a child's body in various movements with an adult's body in the future.

2. Methodology

The subjects were one 5-year-old girl and one adult male. Their respective heights and weights are shown in Table 1. To verify whether it is possible to calculate the burden on the infant's body using motion capture, videos of ‘balancing’ and ‘squatting’ (Fig. 1) were taken and the burden on the lower back was calculated using the AnyBody Modelling System. To verify whether the values were valid, videos of the two movements were also taken of adult males and compared with the calculated values. In addition, the adult male data was analyzed with the height and weight of a 5-year-old girl set and compared and examined with the actual values for a 5-year-old girl to verify whether simulation of the movements by an adult would be possible in the future.

Table 1. Adult male squatting

	Preschool Child	Adult Male
Height (cm)	100	180
Weight (kg)	17	78

Fig. 1. Adult male squatting

The AnyBody Modeling System's analysis procedure

After the video was recorded, the video was converted to mp4, the part to be analyzed was cut out for about 10 seconds, and motion capture data (hereinafter referred to as BVH) was created. The lumbar burden was calculated as the burden on the most painful lumbar vertebrae L4L5.

3. Results and Discussion

3.1. *Relevance of the definition*

For the purposes of this paper, we define "the burden on the hips is the burden on the body. Fig. 2 shows a 5-year-old girl performing a series of balancing movements from behind. The color of the muscles used is changed to indicate the areas of high burden. The color of the lower back in addition to the color of one leg, which is weight-bearing, indicates that the burden is placed mainly on the lower back. Fig. 3 shows a series of crouching movements viewed from behind. The fact that the two movements are mainly burdened on the lower back shows that they meet the definition of "the burden on the lower back is the burden on the body".

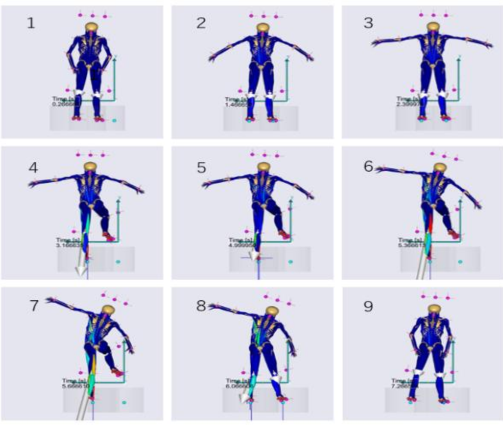


Fig. 2. Sequence of balance movements from behind

Fig. 3. Sequence of squat movements from behind

3.2. *Comparison of Values between a 5-year-old girl and an adult male*

The burden on the body, based on the standing posture, is approximately 100 [N] for a weight of 10 kg. Therefore, a 5-year-old girl weighs 17 kg, so the burden on her body in the standing posture is approximately 170 [N]. An adult male weighs 78 kg, so the standard is approximately 780 N. The burden increases when the body is carrying a load in its hands or when the body axis is tilted.

The average and maximum values of the burden value of the lower back in the balancing movement were 257.866[N] and 425.870[N], respectively, for a 5-year-old girl (Fig. 4). A 5-year-old girl's maximum value posture is posture 8 in Fig. 2, in which the body is tilted due to loss of balance. The mean and maximum lumbar strain values for the balancing movements of adult males were 1012.574 [N] and 1939.441 [N], respectively (Fig. 5). The posture with the maximum value for the adult male was the same as that of the 5-year-old girl, in which he was off-balance, and his body was tilted.

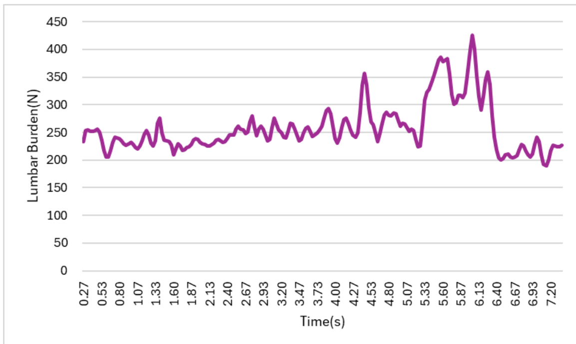


Fig. 4. Graph of lumbar burden values for Balance in a preschool child

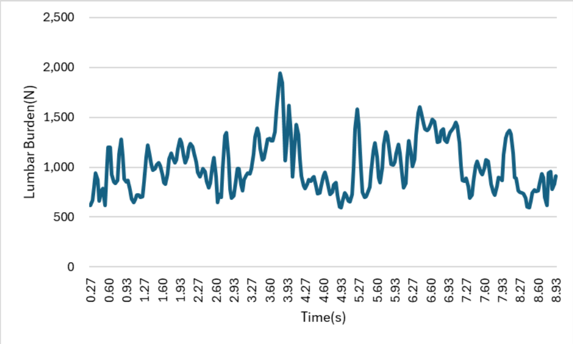


Fig. 5. Graph of lumbar burden values for Balance in an adult male

The mean and maximum values of the squatting burden were 310.173[N] and 556.171[N], respectively, for the 5-year-old girl (Fig. 6). The mean and maximum lumbar strain values for the squatting motion for adult males were 1101.497[N] and 2456.474[N], respectively (Fig. 7). The posture with the maximum value is the most squatting posture.

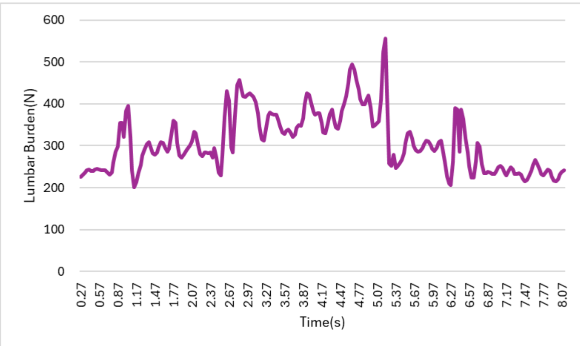


Fig. 6. Graph of lumbar burden values for Squat in a preschool child

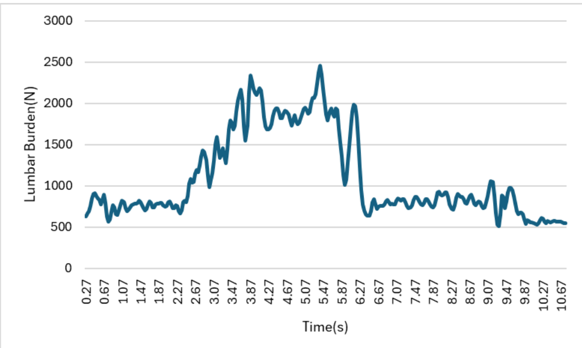


Fig. 7. Graph of lumbar burden values for Squat in an adult male

The difference between the criterion for 5-year-old girls and the mean value of the balance movement was about 88 [N], and the difference between the criterion and the mean value of the squatting movement was about 140 [N]. The difference between the mean values of balance and squatting movements was about 232[N] and 321[N], respectively, from the standard for adult males. Considering the body leaning and squatting movements, we believe that both values are reasonable. In addition, when comparing the average values between the 5-year-old girl and the adult male, the difference in balance movement was about 755[N] and the difference in squatting movement was about 791[N], which we consider to be a reasonable difference since the difference is thought to be caused by differences in height and weight.

3.3. *Analysis of adult male data in the setting of a 5-year-old girl*

Using the adult male BVH and analyzing in the setting of a 5-year-old girl's height and weight, the mean and maximum values for the balance movement were 290.559 [N] and 418.476 [N], respectively. (Fig. 8) The mean value for squatting movements was 342.096 [N] and the maximum value was 560.767 [N]. (Fig. 9) Compared to the figures for 5-year-old girls, the difference in mean values for balance movements was about 33 [N] and for squatting movements about 32 [N], a small difference for both balance and squatting movements, Simulation of children's movements with motion capture data is considered to be possible for adults.

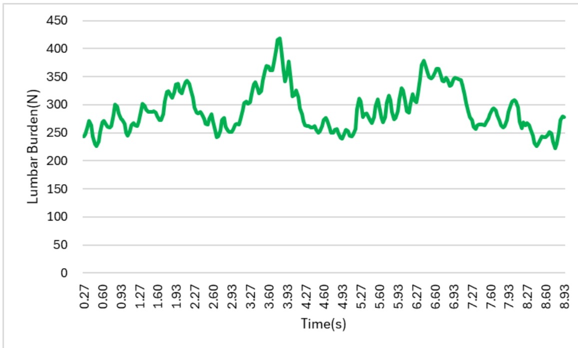


Fig. 8. Graph of lumbar burden values for Balance in an adult male

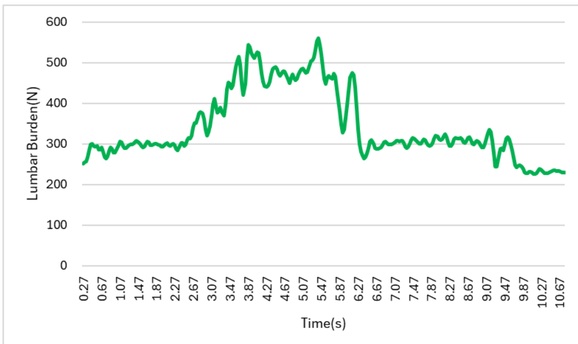


Fig. 9. Graph of lumbar burden values for Squat in an adult male

4. Conclusion

The figures calculated in this paper are shown in Table 2. Reasonable results were obtained in all comparisons of the values of the two movements with the 5-year-old girl's standards, with the adult male's standards and with the values of the two movements, and with the values of the 5-year-old girl and the adult male. In addition, values analyzed using the adult male BVH with the height and weight of a five-year-old girl were calculated to be almost identical to the actual values of a five-year-old girl. This suggests that it is possible to simulate children's movements in adults with an awareness of the amount of physical activity in children's movements. In the future, many movements should be analyzed.

Table 2. Figures calculated in this paper

		Adult Male	Preschool Child	Adult for Child's Setting
Balance	Average[N]	1012.574	257.866	290.559
	Max[N]	1939.441	425.870	418.476
Squat	Average[N]	1101.497	310.173	342.096
	Max[N]	2456.474	556.171	560.767

References

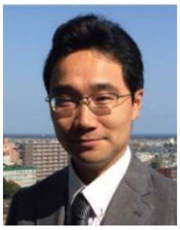
1. Ministry of Education, Culture, Sports, Science and Technology https://www.mext.go.jp/sports/b_menu/toukei/chousa04/tairyoku/kekka/k_detail/1421920_00011.htm (Last viewed on 14 December 2024)
1. A. Sugimoto, "The Measurement of Physical Activity : A Review of Recent Progress", The Japanese Journal of Rehabilitation Medicine, vol. 37, 2000, pp. 53-61.
2. S. Tanaka, "A Study of the Relationship between Young Children's Motor Abilities and Physical Activities : Focusing Five Year Old Children's Physical Activity Countsina Day, The Hoikugaku Kenkyu vol. 47, no.2, 2009, pp. 8–16.
3. R. Tsuda, M. Watanabe and K. Suzuki, "Daily physical activities of public nursery school children during winter : a case study of eight public schools in Ishikawa Prefecture", the Japan Journal of Human Growth and Development Research, vol. 90, 2021, pp. 57-64.
4. M. Damsgaard, J. Rasmussen, S. T. Christensen, E. Surma and M. D. Zee, "Analysis of musculoskeletal systems in the AnyBody Modeling System", Simulation Modelling Practice and Theory, vol. 14, 2006, pp. 1100-1111.

Authors Introduction

Ms. Sachiko Kido

She received her M.S. from Nakamura Gakuen University, Graduate School of Human Development in 2015. Currently enrolled in the doctoral program at the Graduate School of Agriculture and Engineering, University of Miyazaki. Assistant Professor, Department of Early Childhood Education, Higashikyushu Junior College, 2010. Lecturer, Faculty of Education, Miyazaki International University, since 2019. Her main research interest is the development of motor skills in young children.

Prof. Hiroki Tamura



He received the B.E. and M.E. degree from Miyazaki University in 1998 and 2000, respectively. From 2000 to 2001, he was an Engineer in Asahi Kasei Corporation, Japan. In 2001, he joined Toyama University, Toyama, Japan, where he was a Technical Official in the Department of Intellectual Information Systems. In 2006, he joined Miyazaki University, Miyazaki, Japan, where he was an Assistant Professor in the Department of Electrical and Electronic Engineering. Since 2015, he is currently a Professor in the Department of Environmental Robotics. His main research interests are Neural Networks and Optimization Problems. In recent years, he has had interest in Biomedical Signal Processing using Soft Computing.

Evaluation of Ankle Joint Movements in Frontal Plane for a Normal Coordinated Gait

Praveen Nuwantha Gunaratne
Interdisciplinary Graduate School of Agriculture and Engineering, University of Miyazaki, 1-1, Gakuen Kibanadai-Nishi, Miyazaki, 889-2192, Japan

Hiroki Tamura
Faculty of Engineering, University of Miyazaki, 1-1, Gakuen Kibanadai-Nishi, Miyazaki, 889-2192, Japan
Email: ti20060@student.miyazaki-u.ac.jp, htamura@cc.miyazaki-u.ac.jp

Abstract

As per the records, around 15% of the global population is experiencing some form of disabilities in lower extremity resulting in loss of accessibility to their basic routine movements. The ankle joint complex plays an important role as a weight bearing articulation in the lower extremity and is a key contributor to the power behind human locomotion. While sagittal plane ankle movements are crucial for gait, several studies have proven that inversion-eversion, the frontal plane movements oversee the pressure distribution at the ankle joint to ensure a well-coordinated gait. This paper presents an evaluation of prediction of such ankle joint movements using Electromyogram (EMG), Inertial Measurement Unit (IMU) and Force-Sensitive Resistor (FSR) measurements, which can later be adapted for use in anthropometric active ankle orthosis designs to assist dynamic ankle movements during normal gait in real-time.

Keywords: Ankle-joint movements, Frontal plane mechanics, EMG, Active ankle orthoses

1. Introduction

The World Health Organization (WHO) estimates an increase in the global population of individuals aged 65 and above, rising from 524 million in 2010 to 1.5 billion by 2050, reflecting a growth from 8% to 16% of the world's population. Alongside age-related degenerations, a larger proportion of the population is anticipated to experience mobility-related impairments caused by chronic diseases [1],[2]. Walking, as the most fundamental human activity of daily life, serves as a critical indicator of individual health, and gait impairments can significantly compromise independence and quality of life. Consequently, understanding and addressing the biomechanical factors contributing to normal and abnormal gait patterns is crucial for improving rehabilitation strategies and outcomes [3].

The ankle joint complex, a critical weight-bearing articulation that supports nearly 100% of the ground force load, plays a key role in locomotion by contributing approximately 45% of the total power required for walking and running [4],[5]. Its primary movements are plantarflexion-dorsiflexion, occurring in the sagittal plane; inversion-eversion, occurring in the frontal plane and abduction-adduction, occurring in the transverse plane. While the ankle is often simplified as a hinge joint facilitating sagittal plane movements for gait analysis, its function is far more intricate involving simultaneous, complex actions across all three planes such as supination and pronation during locomotion [6]. Notably, movements in frontal plane, inversion and eversion play a crucial role in maintaining balance and stability through uniform distribution of ground reaction forces during gait. These movements adjust the foot contact area with the ground, allowing for efficient pressure absorption and propulsion while minimizing the strain at lower extremities [7],[8].

Given the complexity of ankle biomechanics, evaluating frontal plane movements during normal coordinated gait is particularly important for understanding how these motions contribute to stability and efficiency. Research has shown that frontal plane mechanics play a critical role in dynamic postural control and load redistribution, which are essential for normal gait [9],[10]. The intricate coordination of frontal plane movements with other joint motions ensures proper alignment and minimizes compensatory forces that could lead to injuries or further gait dysfunction. By evaluating the frontal plane ankle joint movements in healthy individuals, researchers can establish baseline parameters for normal gait, which could then be used to develop targeted rehabilitation strategies, particularly through the use of active ankle joint orthosis devices. These devices provide support and assistance on such compromised movement and further it can be customized to individual needs. Therefore, a detailed evaluation of these movements is crucial for optimizing rehabilitation efforts and enhancing the effectiveness of orthotic interventions, especially in the context of a rapidly aging population.

2. Considerations for Evaluation

2.1. Anatomy of the ankle joint complex

The ankle joint complex is comprised of the talocrural, subtalar, and transverse-tarsal joints, each contributing to the intricate biomechanics of foot movement. The talocrural joint, functioning primarily as a hinge, facilitates plantarflexion-dorsiflexion movements of the foot [11]. However, its oblique axis of rotation suggests a more complex motion than that of a simple hinge joint.

The axis of rotation of the talocrural joint for the respective sagittal plane movements occurs around the line shown in Fig.1(a). The subtalar joint, with its geometry, is primarily responsible for inversion-eversion of the foot. The axis of rotation of the subtalar joint is also oblique, as illustrated in Fig.1(b). This oblique axis allows the subtalar joint to produce multiple motions during plantarflexion-dorsiflexion, resulting in pronation and supination [12]. The transverse-tarsal joint is considered as part of the functionally coupled unit with the subtalar joint due to the nature of shared axis of motion, further contributing to inversion-eversion of the foot [13].

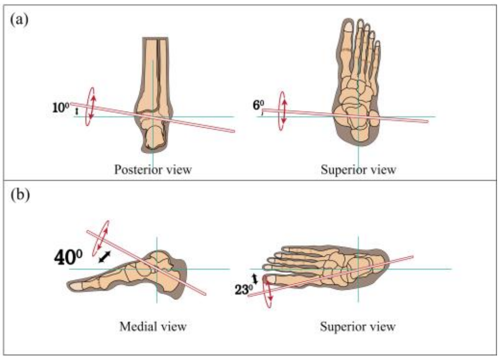


Fig.1. Oblique axes of rotation in ankle joint complex; a) rotational axis of talocrural joint, b) rotational axis of subtalar joint

The motion of the foot and ankle is largely driven by twelve extrinsic muscles, organized into four compartments. The anterior compartment includes four muscles responsible for dorsiflexion and inversion-eversion of the foot, while the lateral compartment contains two muscles that produce plantarflexion and eversion. The posterior compartment consists of three muscles contributing only to plantarflexion, and the deep posterior compartment contains three muscles that facilitate both plantarflexion and inversion of the foot. Understanding the coordinated actions of the ankle joint complex and the related muscles is crucial for the data acquisition processes and evaluations of the dynamic ankle movements [13],[14].

2.2. Gait cycle

During walking, each leg follows a repetitive sequence of steps known as the gait cycle, which begins with the heel strike (HS) of one leg and ends with the subsequent HS of the same leg, marking the start of the next cycle. A gait cycle is broadly divided into two phases: the stance phase and the swing phase. The stance phase, comprising approximately 60% of the cycle, considers when the foot is in contact with the ground, beginning at HS and ending at toe-off (TO). Later, the swing phase starts when the foot is off the ground and in motion (see Fig.2(a)) [15].

The HS begins with the ankle in a slightly plantarflexed position, occurring within the sagittal plane, and transitions to a flat-foot (FF) state. Following this, the ankle moves from plantarflexion to dorsiflexion, facilitating the forward progression of the body during the stance phase. As the heel lifts off the ground, the ankle transitions back into plantarflexion, continuing to its maximum point at TO, where the power generation is achieved to propel the body into the next phase of the gait cycle. During the swing phase, the ankle dorsiflexes to ensure better ground clearance and then repositions to a slightly plantarflexed state, preparing for the next HS (see Fig.2(b)). While these sagittal plane movements dominate the gait cycle, simultaneous movements occur in the frontal plane.

Inversion of the foot is observed up to the HS, enhancing initial ground contact and transitioning to eversion during mid-stance (MST), which facilitates a stable push-off into the swing phase (see Fig.2(c)). Studies have documented an inversion-eversion, within a range of approximately 15 degrees, highlighting the importance of frontal plane dynamics in achieving a coordinated gait [16].

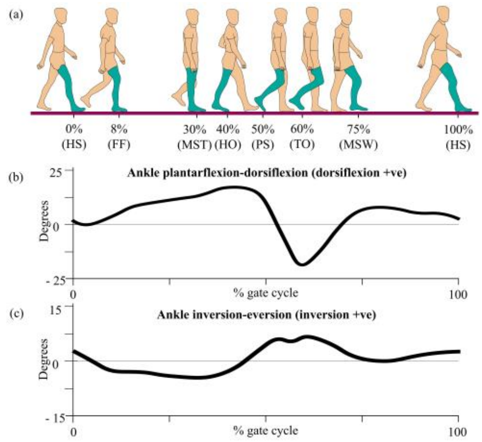


Fig.2. An overview of the gait cycle; a) different states of gait cycle (HO – heel off, PS – pre swing, MSW – Mid swing)[17], b) movements within the sagittal plane across the gait cycle, c) movements within the frontal plane across the gait cycle[11].

3. Methodology
3.1. Subjects

This study reports experiments conducted on four healthy male subjects (average age: 23 years) who provided informed consent to participate and self-reported no history of biomechanical or neuromuscular disorders.

3.2. Experimental setup

The experimental setup included two sensor systems for data acquisition: four wireless surface EMG sensors (Delsys Trigno Avanti), integrated with IMUs, and a four-channel force-sensitive resistor (FSR) sensor

(Delsys Trigno 4-Ch FSR Adapter), enabling precise collection of muscle activity, motion, and foot pressure data during the trials. The sEMG sensors were strategically placed on the muscle bellies of the tibialis anterior, fibularis longus, soleus and gastrocnemius muscles of the right-foot (see Fig.3), as these muscles are known to significantly contribute to ankle joint functions according to previous research [18]. Data from the sEMG sensors were recorded using EMGworks software, developed by Delsys Inc. at a sampling rate of 1926 samples/sec (sa/s) with a bandwidth of 20 – 450 Hz, while acceleration data from the integrated IMUs were captured at a sampling rate of 148 sa/s with a bandwidth of 24 – 470 Hz. The FSR adapter, attached to a sandal, was used to measure the pressure distribution of foot contact with the ground. The FSR membranes were affixed to the inner sole of the sandal, modified to each subject's individual foot configuration to ensure accurate data acquisition. The Fig.4 illustrates the layout of FSR membranes used for different subjects. Data from the FSR sensors were recorded at sampling rates of: 1926 sa/s for channel 1 and 148 sa/s for channels 2 – 4, with a bandwidth of 50 Hz.

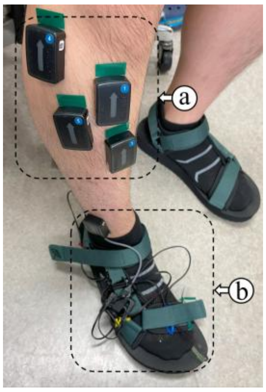


Fig.3. Sensor arrangement for data acquisition; a) wireless sEMG sensor assembly, b) wireless FSR sensor attachment

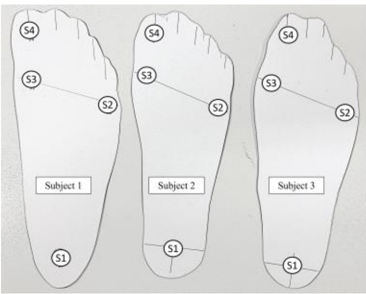


Fig.4. Placement layout of FSR membranes for different subjects based on their unique foot configurations (S1- membrane connected to channel 1, S2 – connected to channel 2 and so on)

3.3. Experimental procedure

The subjects were instructed to walk at their normal gait pattern over a fixed distance of 12 meters to collect the data. This test was repeated for five sessions per subject, while maintaining consistent intervals between each session to ensure reliable data acquisition.

4. Analysis and Results

The collected data were subjected to a series of steps, including pre-processing, filtering, rectification, and normalization, using a Python program. These steps were performed separately based on each sensor's sampling rate and bandwidth to ensure accurate data handling. Following the processing, the datasets were input into an evaluation model for further analysis. Fig.5 illustrates a sample FSR output acquired from Subject 1. Fig.6 subsequently shows the average EMG waveform along with the calculated EMG envelope, derived to assess the average activation of the EMG signals. These figures provide a detailed representation of the average EMG calculations and the EMG envelopes, facilitating a deeper understanding of muscle activation patterns for further evaluation.

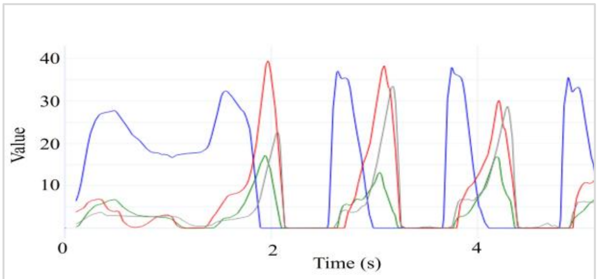


Fig.5. A segment of the FSR sensor output acquired from subject 1 (Blue - pressure sensor 1, Green - sensor 2, Red - sensor 3, Gray - sensor 4)

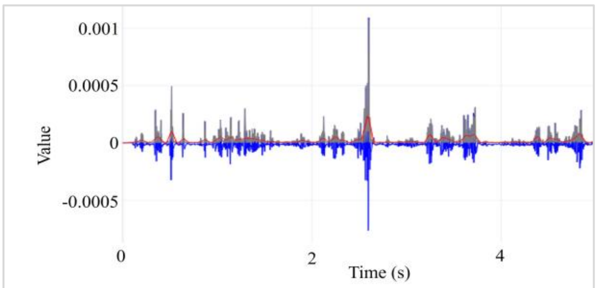


Fig.6. A segment of the EMG envelope corresponds to the original data represented (Blue – original waveform, Gray – avg. rectified waveform, Red – EMG envelope).

5. Discussion

This paper emphasizes the significance of evaluating frontal plane ankle joint movements by critically reviewing the key anatomical and gait related factors. The study further discusses the processes and methodologies that are crucial for assessing dynamic movements within the ankle joint complex, considering its detailed structure. The data obtained in this study demonstrate that a well-defined evaluation model can provide reliable estimations for inversion – eversion movements during normal gait. This approach has the potential to be helpful in optimizing the control of active ankle orthoses, facilitating gait rehabilitation, and offering continuous real-time assistance for individuals with gait impairment, which will be the focus of future studies.

References

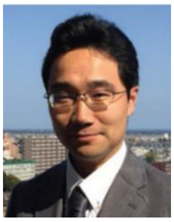
1. M. Grimmer, R. Riener, C. J. Walsh, and A. Seyfarth, "Mobility related physical and functional losses due to aging and disease - a motivation for lower limb exoskeletons," *Journal of NeuroEngineering and Rehabilitation*, vol. 16, no. 1, Jan. 2019, doi: https://doi.org/10.1186/s12984-018-0458-8.
2. "Global Health and Aging," World Health Organization, Oct. 2011. Available: https://www.nia.nih.gov/sites/default/files/2017-06/global_health_aging.pdf
3. O. Rasmussen, "Stability of the Ankle Joint," *Acta Orthopaedica Scandinavica*, vol. 56, no. sup211, pp. 1–75, Jan. 1985, doi: https://doi.org/10.3109/17453678509154152.
4. J. H. Calhoun, F. Li, B. R. Ledbetter, and S. F. Viegas, "A Comprehensive Study of Pressure Distribution in the Ankle Joint with Inversion and Eversion," *Foot & Ankle International*, vol. 15, no. 3, pp. 125–133, Mar. 1994, doi: https://doi.org/10.1177/107110079401500307.
5. C. M. Thalman, T. Hertzell, and H. Lee, "Toward A Soft Robotic Ankle-Foot Orthosis (SR-AFO) Exosuit for Human Locomotion: Preliminary Results in Late Stance Plantarflexion Assistance," *3rd IEEE International Conference on Soft Robotics (RoboSoft)*, May 01, 2020. https://ieeexplore.ieee.org/abstract/document/9116050
6. P. Procter and J. P. Paul, "Ankle joint biomechanics," *Journal of Biomechanics*, vol. 15, no. 9, pp. 627–634, 1982, doi: https://doi.org/10.1016/0021-9290(82)90017-3.
7. C. Moriguchi, T. Sato, and H. Gil Coury, "Ankle movements during normal gait evaluated by flexible electrogoniometer," *Revista Brasileira de Fisioterapia*, vol. 11, no. 3, Jun. 2007, doi: https://doi.org/10.1590/s1413-35552007000300006.
8. M. A. Lafortune, P. R. Cavanagh, H. J. Sommer, and A. Kalenak, "Foot inversion-eversion and knee kinematics during walking," *Journal of Orthopaedic Research*, vol. 12, no. 3, pp. 412–420, May 1994, doi: https://doi.org/10.1002/jor.1100120314.
9. A. Leardini, J. J. O'Connor, and S. Giannini, "Biomechanics of the natural, arthritic, and replaced human ankle joint," *Journal of Foot and Ankle Research*, vol. 7, no. 1, Feb. 2014, doi: https://doi.org/10.1186/1757-1146-7-8.
10. J. R. Close, "Some Applications of the Functional Anatomy of the Ankle Joint," *The Journal of Bone & Joint Surgery*, vol. 38, no. 4, pp. 761–781, Jul. 1956, doi: https://doi.org/10.2106/00004623-195638040-00005.
11. C. L. Brockett and G. J. Chapman, "Biomechanics of the ankle," *Orthopaedics and Trauma*, vol. 30, no. 3, pp. 232–238, Jun. 2016, doi: https://doi.org/10.1016/j.mporth.2016.04.015.
12. S. K. SARRAFIAN, "Biomechanics of the Subtalar Joint Complex," *Clinical Orthopaedics and Related Research*, vol. &NA;, no. 290, p. 17???26, May 1993, doi: https://doi.org/10.1097/00003086-199305000-00003.
13. P. Procter and J. P. Paul, "Ankle joint biomechanics," *Journal of Biomechanics*, vol. 15, no. 9, pp. 627–634, 1982, doi: https://doi.org/10.1016/0021-9290(82)90017-3.
14. R. Drake, A. Wayne Vogl, and Adam, *Gray's Anatomy for Students*. Elsevier Health Sciences, 2009.
15. Arnaldo Leal-Junior and Anselmo Frizera-Neto, "Gait analysis: overview, trends, and challenges," *Elsevier eBooks*, pp. 53–64, Jan. 2022, doi: https://doi.org/10.1016/b978-0-32-385952-3.00011-1.
16. B. M. Nigg, V. Fisher, and J. L. Ronsky, "Gait characteristics as a function of age and gender," *Gait & Posture*, vol. 2, no. 4, pp. 213–220, Dec. 1994, doi: https://doi.org/10.1016/0966-6362(94)90106-6.
17. Tengxu Xia, "Real-time Gait Event Tracker (RGET): An Innovative Approach to Digital Human Modeling through Integration of Inertial Measurement Unit and Dynamic Feature Extraction," *Journal of Electrical Systems*, vol. 20, no. 2, pp. 598–615, Apr. 2024, doi: https://doi.org/10.52783/jes.1214.
18. H. Dallali, L. Knop, L. Castelino, E. Ficanha, and M. Rastgaar, "Estimating the multivariable human ankle impedance in dorsi-plantarflexion and inversion-eversion directions using EMG signals and artificial neural networks," *International Journal of Intelligent Robotics and Applications*, vol. 1, no. 1, pp. 19–31, Jan. 2017, doi: https://doi.org/10.1007/s41315-016-0004-4.

Authors Introduction

Mr. Praveen Nuwantha Gunaratne

He received his Bachelor's degree in Engineering from the Faculty of Engineering, University of Moratuwa, Sri Lanka, in 2018. In 2022, he completed his Master's degree at the University of Miyazaki, Japan. He is currently pursuing his Doctoral studies at the University of Miyazaki, Japan.

Prof. Hiroki Tamura



He received the B.E. and M.E. degree from Miyazaki University in 1998 and 2000, respectively. From 2000 to 2001, he was an Engineer in Asahi Kasei Corporation, Japan. In 2001, he joined Toyama University, Toyama, Japan, where he was a Technical Official in the Department of Intellectual Information Systems. In 2006, he joined Miyazaki University, Miyazaki, Japan, where he was an Assistant Professor in the Department of Electrical and Electronic Engineering. Since 2015, he is currently a Professor in the Department of Environmental Robotics. His main research interests are Neural Networks and Optimization Problems. In recent years, he has had an interest in Biomedical Signal Processing using Soft Computing.

Development of a Real-Time Multi-Person 3D Keypoint Detection System Using Stereoscopic Cameras and RTMPose

Taufik Hidayat Soesilo
Graduate School of Engineering, University of Miyazaki, 1-1 Gakuen Kibanadai-Nishi, Miyazaki-shi, Miyazaki, 889-2155, Japan

Praveen Nuwantha Gunaratne
Interdisciplinary Graduate School of Agriculture and Engineering, University of Miyazaki, 1-1, Gakuen Kibanadai-Nishi, Miyazaki, 889-2192, Japan

Hiroki Tamura
Faculty of Engineering, University of Miyazaki,1-1 Gakuen Kibanadai-Nishi, Miyazaki-shi, Miyazaki, 889-2155, Japan
E-mail: ti22068@student.miyazaki-u.ac.jp, ti20060@student.miyazaki-u.ac.jp, htamura@cc.miyazaki-u.ac.jp

Abstract

The feature offers a way of tracking the movement patterns of many people at once, and is critical in occupational health, sports performance, and team-based work settings. The selected traditional biomechanical analysis systems, in turn, are largely considered in detecting single person movement, which weakens their applicability to movement analysis that involves interacting with multiple people. In this paper we consider a real time multi-person detection and analysis system using stereoscopic cameras and RTMPose, a novel high real-time pose estimation framework. RTMPose offers real time analysis of 2D key points for the individuals and this data is later augmented with depth data coming from stereoscopic imaging to give 3D skeletal data. The benefit of employing RTMPose is that the system is able to perform accurate and fast multiple persons tracking despite present occlusion scenarios. Consequently, the system overcomes the drawbacks of prior methods, including reliance on wearable devices and unsuitability for out-of-door environments, by employing stereoscopic cameras and RTMPose with low-latency and high-accurate inference. Experimental results demonstrate the system's ability to provide detailed real-time analysis of posture and movement for multiple individuals in diverse scenarios. This research highlights the potential of RTMPose-powered systems to advance multi-person biomechanical analysis for applications ranging from workplace monitoring to sports performance assessment.

Keywords: RTMpose, real-time biomechanical analysis, computer vision, multi-person 3D keypoint detection

1. Introduction

The requirement for the decomposition of multiple individuals at once in real-time has emerged in some fields including occupational health, sports, and ergonomics. Extended conventional frameworks of biomechanical assessment have generally been developed to examine one person at a time, thus being restricted in multi-subject environments. For instance, tasks of analysis in a team environment including construction, healthcare setting or production line work involve capacity to discern group motions for safety, efficiency and order to enhance performance. That is why there is a need to create new systems that would be able to detect and analyze multiple people with increasing complexity of backgrounds [1],[2].

In particular, the progresses in computer vision and stereoscopic imaging technology enable the multi-person detection in real time. While many older approaches only use singe-camera equipment or body-worn sensors, stereoscopic systems are based on two-camera configurations for depth acquisition. This enables the realist rendition of 3D keypoints for the multiple individuals, based on the crowded or overlapping states. These capabilities can be highly employed in cases where all the body movement of the subjects has to be studied at the same time as in sports, ergonomic studies and group work areas [3],[4].

Advanced techniques including deep learning based pose estimation in multi-person detection are currently employed. Some popular benchmarks like OpenPose, AlphaPose, and RTMPose have demonstrated satisfactory accuracy of body-structured 2D human pose estimation of multiple persons [5],[6],[7]. When correlated with depth information gathered from stereoscopic cameras these methods allow for determining three-dimensional corresponding key points for each of participants. This integration solves some of the issues that include occlusion, lighting variation, and objects sharing the same space which tends to complicate the model but produces efficient and accurate tracking applicable in real-life scenarios [8],[9].

Building on these advancements, this study focuses on developing a system for multi-person detection and analysis using stereoscopic cameras. The proposed solution offers detailed 3D key-point data for multiple individuals in real-time, addressing the limitations of single-person systems. By employing state-of-the-art multi-person detection algorithms and robust depth-sensing technology, the system is designed to operate effectively across diverse conditions, making it suitable for a wide range of applications, from workplace monitoring to athletic performance analysis [10].

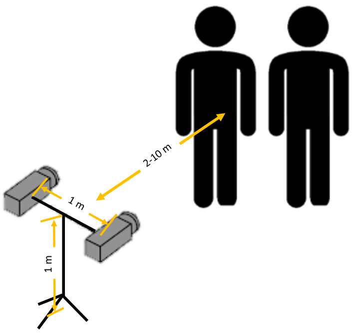


Fig. 1. Camera setup

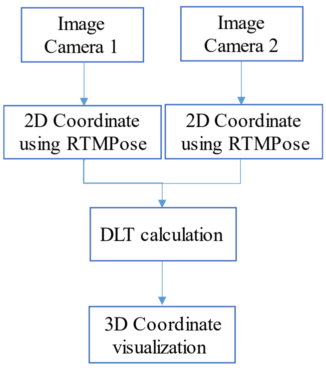


Fig. 2. System Flowchart

2. System Design and Methodology

2.1. System setup

The proposed system also utilizes a stereoscopic camera combination to acquire simultaneous videos that are helpful in the determination of multiple individuals. The setup consists of having two cameras placed in parallel with a defined baseline distance of 1 meter to keep a high level of accuracy of the depth estimation, as depicted in Fig. 1. The cameras are adjusted so as to define the intrinsic and extrinsic parameters that provide the foundations for enriching 3D data.

Similarly, the stereoscopic camera setup is positioned to obtain a rich view and capture all regions of the detected scene to identify different postures and orientations of the individuals. The videos captured are then processed in real time in order to obtain 2D key points for each person by using RTMPose. It maintains a high degree of visibility and small-object discernibility because of the proper positioning of the devices.

Key Hardware Details:

- **Camera Model:** Elgato Facecam Full HD
- **Resolution:** 1920x1080 pixels
- **Frame Rate:** 60 fps
- **Baseline Distance:** 1 meter
- **Calibration Tools:** OpenCV stereo calibration (to compute intrinsic/extrinsic parameters and distortion coefficients)

2.2. *Pose Estimation Framework*

The identification of multiple people and their skeletons is done using RTMPose, an efficient real-time pose estimation tool. RTMPose takes the synchronized video feeds from the stereoscopic cameras, and identifies the 2D key points of all people in each frame. The RTMPose model uses a deep learning model for real time object detection with high speed and accuracy the model for RTMPose is ResNet or MobileNet. This framework extracts 17 specific points (according to the COCO dataset) for every individual: shoulders, elbows, knees, ankles as well as other parts. These 2D key points are moreover linked with unique IDs corresponding to each person to afford stable correspondence from frame to frame. The system flowchart can be seen in the Fig 2

Fig. 3. Multi-Person and detected keypoints in 2D for applications of skeleton overlays.

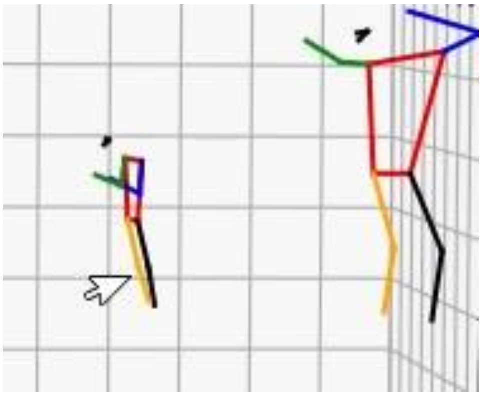


Fig. 4. The 3D coordinates for the ribs of both individuals were reconstructed.

2.3. *3D Keypoint Reconstruction*

The 3D key points for each subject are reconstructed using the 2D key point data obtained from RTMPose and the depth obtained from the stereoscopic cameras. This process involves extracting the 2D key point coordinates from the left and right camera; projected into 3D space.

The reconstruction is performed using the Direct Linear Transformation (DLT) technique that estimates the coordinates of all the important points with respect to the coordinate frame of the camera [11]. Extrinsic and intrinsic calibration data is used to compute the coefficients for lens distortion and thereby facilitates the calculation of depth.

3. Results and Discussion

3.1. Results

The proposed system can display real time multi-person detection, and 3D keypoints reconstruction with stereoscopic camera and the RTMPose. As it can be seen from the outcomes that the developed system successfully recognizes several people concurrently literally in a real conditions as well and yields quite high accuracy rates in terms of the Keypoints tracking in addition to the real life possibilities to generate the 3D bone models.

Key Observations:

1. 2D Keypoint Detection:
   - The system successfully detected key points for all bodies like some primary joints of shoulder, elbow, knee, ankles of more than one person. Congruent outlines drawn around each of the individuals established sharp and clear segmentation and identification of subjects as shown in Fig 3.
2. 3D Keypoint Reconstruction:
   - With data obtained from the stereoscopic camera setup the system constructed 3D keypoints for each person. These are well oriented with the detected individuals in the 2D images, in order to have accurate spatial correspondences in the 3D environment as the skeletal models can be seen in the Fig 4.
3. Performance Metrics:
   - Accuracy: The computed 3D keypoints of the reconstructed ego-vehicle appeared to match the position of the individuals regardless of their execution of different movements and partially obscured body parts.
   - Processing Speed: The system was able to obtain real time performance with average frame rate of the order of 15 fps.
4. Multi-Person Capability:
   - Several people were effectively detected and tracked within the field of view, in the 2D view demonstrated by the bounding boxes and in the 3D view through the skeleton overlays.

3.2. Discussion

The results validate the effectiveness of the proposed system in handling real-time multi-person detection and 3D keypoint reconstruction. Several key aspects are

Challenges and Limitations:

1. Occlusion Handling: In most cases, the system provided accurate detection of keypoints and the estimation of 3D poses, but full occlusions of the body parts led to small errors. Scanner input occlusion handling could be the subject of future work, with temporal data or limit algorithms used.
2. Complex Environments: The current testing environment was relatively controlled. Performance with signals in more difficult conditions such as in a noisy environment with many subjects and with greatly fluctuating light conditions requires assessment.
3. Depth Accuracy: Disparity based depth estimation technique may contain noises at some time period when the camera is moving fast or during large change in lighting. However, this problem could be reduced by improved calibration on the camera setup or using better filtering methods.

Potential Applications:

1. Workplace Monitoring: They include real-time tracking of multiple workers to help monitor safety and even workers' ergonomics.
2. Healthcare: To maintain correct movements for a number of patients at once, group rehabilitation sessions are used.

4. Conclusion

The proposed system is capable of real-time multi-person detection and 3D keypoint reconstruction, while using stereoscopic cameras and RTMPose. It shows high precision, low delay, and multiple person tracking ability, hence, it can be used for workplace monitoring, sports performance analysis, and group physical therapy.

The performance of the system is excellent, however, it is essential to consider the following factors: Occasion where one of the cameras is occluded by the other, noisy depth estimation, and the existence of other objects within a scene. The future work will therefore aim at enhancing these aspects as well as incorporating other features such as action recognition to the system. This system demonstrates better advance of multi-person biomechanical analysis and will have an enhanced foundation for related purposes.

References

1. Z. Cao, G. Hidalgo, T. Simon, et al., "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172–186, 2019. doi:10.1109/TPAMI.2019.2929257.
2. H.-S. Fang, J. Li, H. Tang, C. Xu, H. Zhu, and Y. Xiu, "AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 6, pp. 7157–7173, Jun. 2023. doi:10.1109/TPAMI.2022.3222784.
3. S. Dill, A. Ahmadi, M. Grimmer, D. Haufe, M. Rohr, Y. Zhao, M. Sharbafi, and C. H. Antink, "Accuracy Evaluation of 3D Pose Reconstruction Algorithms Through Stereo Camera Information Fusion for Physical Exercises with MediaPipe Pose," Sensors, vol. 24, no. 23, p. 7772, 2024. doi:10.3390/s24237772.
4. F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann, "MediaPipe Hands: On-device Real-time Hand Tracking," arXiv, Jun. 2020. doi:10.48550/arXiv.2006.10214.
5. J.-K. Tsai, C.-C. Hsu, W.-Y. Wang, and S.-K. Huang, "Deep Learning-Based Real-Time Multiple-Person Action Recognition System," Sensors, vol. 20, no. 17, p. 4758, Aug. 2020. doi:10.3390/s20174758.
6. A. Mathis, P. Mamidanna, K. M. Cury, et al., "DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning," Nature Neuroscience, vol. 21, pp. 1281–1289, Sep. 2018. doi:10.1038/s41593-018-0209-y.
7. T. Jiang, P. Lu, L. Zhang, N. Ma, R. Han, C. Lyu, Y. Li, and K. Chen, "RTMPose: Real-Time Multi-Person Pose Estimation based on MMPose," arXiv, Jul. 2023. doi:10.48550/arXiv.2303.07399.
8. T. Golda, T. Kalb, A. Schumann, and J. Beyerer, "Human Pose Estimation for Real-World Crowded Scenarios," in 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Taipei, Taiwan, Sep. 2019, pp. 1–8. doi:10.1109/AVSS.2019.8909823.
9. N. Sarafianos, B. Boteanu, B. Ionescu, and I. A. Kakadiaris, "3D Human Pose Estimation: A Review of the Literature and Analysis of Covariates," Computer Vision and Image Understanding, vol. 152, pp. 1–20, Nov. 2016. doi:10.1016/j.cviu.2016.09.002.
10. D. Rodriguez-Criado, P. Bachiller-Burgos, G. Vogiatzis, et al., "Multi-person 3D pose estimation from unlabeled data," Machine Vision and Applications, vol. 35, no. 46, Apr. 2024. doi:10.1007/s00138-024-01530-6.
11. T. Batpurev, "Real time 3D body pose estimation using MediaPipe," GitHub. [Online]. Available: https://github.com/TemugeB/bodypose3d. [Accessed: Nov. 14, 2022].

Authors Introduction

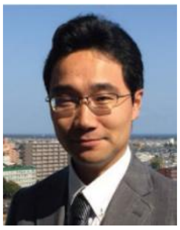
Mr. Taufik Hidayat Soesilo

He received his Bachelor's degree in Engineering in 2021 from the Faculty of Engineering, University of Brawijaya in Indonesia. He is currently a doctoral student in University of Miyazaki, Japan

Mr. Praveen Nuwantha Gunaratne

He received his bachelor's degree in engineering in 2018 from the Faculty of Engineering, University of Moratuwa, Sri Lanka. He is currently a doctoral student in University of Miyazaki, Japan.

Prof. Hiroki Tamura



He received the B.E. and M.E. degree from Miyazaki University in 1998 and 2000, respectively. From 2000 to 2001, he was an Engineer in Asahi Kasei Corporation, Japan. In 2001, he joined Toyama University, Toyama, Japan, where he was a Technical Official in the Department of Intellectual Information Systems. In 2006, he joined Miyazaki University, Miyazaki, Japan, where he was an Assistant Professor in the Department of Electrical and Electronic Engineering. Since 2015, he is currently a Professor in the Department of Environmental Robotics. His main research interests are Neural Networks and Optimization Problems. In recent years, he has had interest in Biomedical Signal Processing using Soft Computing.

Fine-registered Object LiDAR-inertial Odometry for a Solid-state LiDAR system

Hanyeol Lee
Department of Aerospace Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu
Seoul, 08826 Republic of Korea

Chan Gook Park
Department of Aerospace Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu
Seoul, 08826 Republic of Korea
E-mail: han2110@snu.ac.kr, chanpark@snu.ac.kr

Abstract

We propose the LiDAR-inertial odometry with object measurements for the solid-state LiDAR system with narrow field-of-view. Although the geometric feature has been used for the precise localization with LiDAR, the measurement vanishing can lead to the localization failure in the limited field-of-view. To address this problem, objects that are sufficiently present in a man-made environment can be used as localization measurements, but accurate registration is needed to formulate them due to partial observations. The point clouds in the object are registered and the processed measurements are coupled with the geometric measurements in the estimator. The effectiveness of the object measurements is verified through a virtual environment simulator, and the proposed algorithm shows superior localization performance compared to the case of geometric measurement alone.

Keywords: Localization, LiDAR-inertial odometry, semantic perception, point cloud registration

1. Introduction

LiDAR has superior performance in sensing the spatial information from the environment with the accurate distance measurements from ray-casting and is used in many applications such as object perception, tracking, and localization based on map-matching or odometry. Especially in the field of localization, a framework for pose estimation with whole point cloud data has been developed, which is represented by iterative closest point (ICP) and its variants [1]. On the other hand, in the last decade, the use of 3D LiDAR has led to the demand for lightweight computation and mapping, and has been used as a necessary block in the task of simultaneous localization and mapping (SLAM). To meet the above two conditions, an effective and efficient approach to extract and match feature points based on geometric information among the entire point cloud was proposed, called LiDAR odometry and mapping (LOAM) [2]. It provided a breakthrough in LiDAR localization in terms of a real-time operation and accuracy. Since then, LOAM-based LiDAR localization studies have been proposed, and in particular, tight coupling with IMU has been studied for robust application in real-world problems [3]. This has led to significant performance improvements such as mitigating the distortion problem of point clouds due to motion, robust association between features, and accurate initialization of estimators from inertial navigation system (INS).

Although LiDAR can be used as a useful sensor for robotic systems, the inherently complex structure of spinning LiDAR leads to high prices. Therefore, solid-state LiDAR with simple structures are being developed for commercialization. However, solid-state LiDAR has a narrow field of view (FoV), which introduces measurement vanishing into currently existing geometric LiDAR localization algorithms. This does not provide sufficient constraints on the estimated state variables and leads to pose drift and divergence. To address this issue, we propose an approach that additionally uses object measurements in addition to geometric measurements. Object measurements can be effectively obtained and utilized for LiDAR systems that are typically operated in man-made environments.

Recently, object perception has been the main agenda in robotics and computer vision with rapid performance improvement. Autonomous systems require high-level perception as an essential building block for manipulation, exploration, and autonomous driving. In LiDAR systems, data-driven object detectors provide an object bounding box (Bbox) [4] that can be utilized to formulate object measurements. However, LiDAR measurements have a measurement characteristic called L-shape, and there are inevitably unobserved parts of the object. These parts are implicitly estimated based on a prior model within the trained neural network, which cannot provide accurate information about each observed object. To tackle this problem, we aim to perform precise registration by utilizing the point cloud inside the obtained object, and the obtained relative pose is used as a measurement for the localization algorithm. In addition, the obtained object measurements also provide semantic information, which can play an important role in autonomous systems.

2. LiDAR localization with object feature

Geometric feature-based LiDAR-inertial odometry (LIO) starts by using both planes and edges as measurements, but plane features are dominant with robustness to data association in the current. In our algorithm, we also construct a plane-based LIO that combines geometric and statistical information to extract the plane and perform data association.

The obtained plane measurements are processed in an Extended Kalman filter with sliding window to take the advantage of the fixed-lag smoother. The state to be estimated has the form of an error state variable as

$$\delta x = \left[\delta\theta^T\ \delta p^T\ \delta v^T\ \delta b_a^T\ \delta b_g^T\ \delta\theta_1^T\ \delta p_1^T, \dots, \delta\theta_N^T, \delta p_N^T\ \right]^T. \quad (1)$$

Each element is, in order, an error representation of the attitude, position, velocity, accelerometer bias, gyro bias, and attitude and position at the current point in time. N is the size of the sliding window. The system model for the state variables is as follows.

$$\dot{\delta x} = F\delta x + \omega. \quad (2)$$

F is the Jacobian of the state transition model and ω is the process noise. In our system, this is done by the IMU. Accordingly, the equation for the discretized covariance propagation is expressed as

$$P_k = \Phi P_{k-1}\Phi^T + Q. \quad (3)$$

P_k and P_{k-1} are the covariances at time k and $k-1$, Φ is F discretized through integration, and Q is the covariance of discretized ω. The measure model is then as follows.

$$z_k = Hx_k + v. \quad (4)$$

z_k is the measurement obtained via lidar, which in our system includes geometric and object measurements. H is the Jacobian of the measurement model and v is the measurement noise. The Kalman gain is

$$K = PH^T(HPH^T + R)^{-1}. \quad (5)$$

The measurement update equation about state is shown below.

$$\delta x = K\big(z - h(\hat{x})\big). \quad (6)$$

$h(\hat{x})$ is the nonlinear mapping function from state to measurement. The covariance is updated as follows.

$$P^+ = (I - KH)P^-(I - KH)^{\mathrm{T}} + KRK^T \quad (7)$$

P^+ and P^- mean posteriori covariance and priori covariance, respectively. The object measurement used in Eq. (4) is obtained by ICP and is the solution of the optimization problem in (8).

$$(\mathrm{R}, \mathrm{t})^* = \arg\min_{(R,t)} \Sigma\|p_i - (Rq_i + t)\|. \quad (8)$$

p_i and q_i are the i-th pair of points in point sets that are associated by distance.

3. Results

To validate the localization performance of the proposed algorithm, the sensor data is corrected in CARLA, a virtual environment simulator based on the Unreal engine. The virtual environment used in the simulation is shown in Figure 1. In the configured virtual environment, a solid-state LiDAR with a FoV of 45 degrees and an IMU are attached to the ego-vehicle, and an example of the point cloud data obtained is shown in Figure 2. Considering the MEMS IMU performance, the accelerometer error and gyroscope error are introduced into the raw data, respectively. The sensor specifications are listed in Table 1. For object measurements, semantic

Figure 1. The virtual environment in simulator.

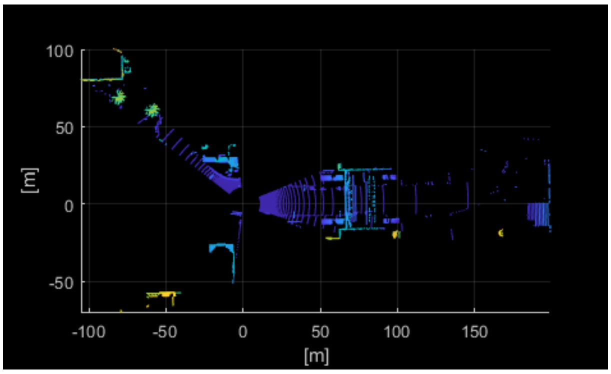


Figure 2. The example of a point cloud data

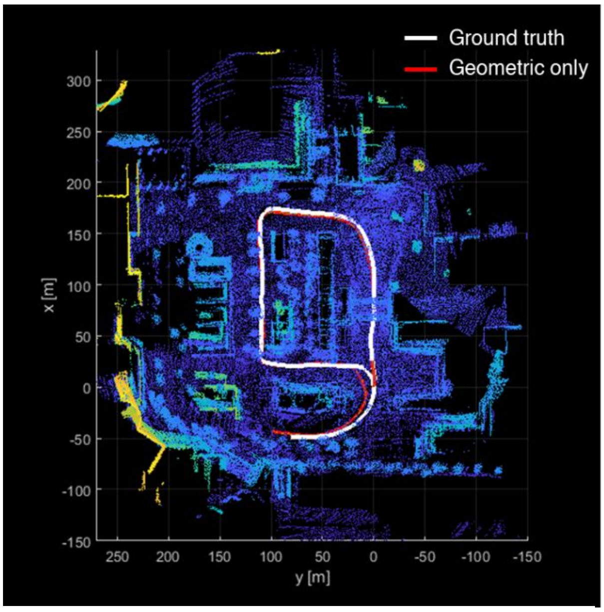


Figure 3. The trajectory of geometric only LIO

labels provided by CARLA [5] are used, and a static vehicle is formulated as a measurement.

The sliding window size of the estimator is set to 10, and the maximum iteration number of ICP is set to 20. To use stable object measurements, convergence is determined by the final loss value of ICP, and outliers are removed by Mahalanobis distance with the covariance of the measurements obtained by the estimator. For comparison, LIO using only geometric measures is used.

Table 1. IMU specification.

	Accelerometer	Gyroscope
Noise density	$0.14\text{mg}/\sqrt{Hz}$	$0.0035°/s/\sqrt{Hz}$
Turn-on bias	0.005g	$0.2°/s$

The vehicle moves the virtual map under these conditions, and the estimated trajectories from the proposed method and the compared method are shown in Figure 3 and Figure 4, respectively. The proposed result is closer to the true trajectory compared to the comparison using only geometric measurements. This indicates that when geometric measurements are lost due

Table 2. Position RMSE results

Position	Geometric only	Proposed
RMSE [m]	7.2702	**5.9435**

to the narrow FoV of solid-state LiDAR in certain segments, object measurements can lead to a robust navigation estimate. Finally, the quantitative error of the two results is presented in Table 2 using the root-mean-square-error (RMSE) metric, and it can be seen that the estimation error is also superior using the proposed object measurements.

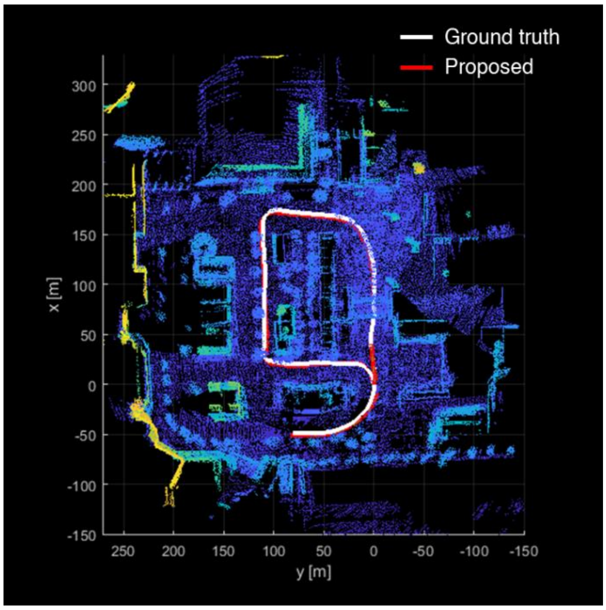


Figure 4. The trajectory of proposed LIO

4. Conclusion

We have proposed the LIO algorithm with object measurement to address the geometric measurement vanishing due to the narrow FoV of a solid-state LiDAR. Observed objects are fine-registered by ICP to formulate relative pose measurements, which are coupled with geometric measurements to update the measurements of the EKF. The proposed algorithm has been validated using a sensor dataset obtained in a virtual urban environment, and an improved localization error is presented compared to the case using geometric measurements alone. The proposed algorithm can be developed into a precise localization algorithm by mapping with object state and loop closure.

Acknowledgment

This research was supported by the National Research Foundation of Korea funded by the Ministry of Science and ICT, the Republic of Korea, under Grant NRF-2022R1A2C2012166.

References

1. P. Besl and N. McKay, “A method for registration of 3-D shapes”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, pp. 239-256, April, 1992.
2. J. Zhang and S. Singh, “LOAM: Lidar odometry and mapping in real-time”, Proceedings of 9th Robotics: Science and systems, pp. 1-9, 2014.
3. W. Xu, Y. Cai, D. He, J. Lin and F. Zhang, “FAST-LIO2: Fast Direct LiDAR-Inertial Odometry”, IEEE Transactions on Robotics, Vol. 38, No. 4, pp. 2053-2073, August, 2022.

4. A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "PointPillars: Fast encoders for object detection from point clouds", Proceedings of 32nd IEEE Conference on Computer Vision and Pattern Recognition, pp. 12689-12697.
5. A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator", Proceedings of 1st IEEE Conference on robot learning, pp. 1-16, 2017.

Authors Introduction

Mr. Hanyeol Lee

He received the B. S. degree in Mechanical and Aerospace Engineering in 2021 from Seoul National University, Seoul, South Korea, where he is currently working toward the Ph.D. degree in Aerospace Engineering. His research interests include semantic perception and SLAM for mobile robots and autonomous driving.

Dr. Chan Gook Park

He received the B.S., M.S., and Ph.D. degrees in control and instrumentation engineering from Seoul National University, Seoul, South Korea, in 1985, 1987, and 1993, respectively. In 1998, he was as a Postdoctoral Fellow with Prof. J. L. Speyer about peak seeking control for formation flight with the University of California at Los Angeles, Los Angeles, CA, USA. From 1994 to 2003, he was an Associate Professor with Kwangwoon University, Seoul. In 2003, he joined the Faculty with the School of Mechanical and Aerospace Engineering, Seoul National University, where he is currently a Professor. In 2009, he was a Visiting Scholar with the Department of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA. His research interests include advanced filtering techniques, high-precision inertial navigation system (INS), visual-inertial odometry, INS/GNSS/IBN integration, and smartphone-based/foot-mounted pedestrian dead reckoning systems. Dr. Park was the Chair of IEEE AES Korea Chapter until 2009.

A Fusion Method for Estimating the Walking Direction of Smartwatch Users

Jae Hong Lee
Department of Aerospace Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu
Seoul, 08826 Republic of Korea

Chan Gook Park
Department of Aerospace Engineering/ASRI, Seoul National University, 1 Gwanak-ro, Gwanak-gu
Seoul, 08826 Republic of Korea
E-mail: honglj@snu.ac.kr, chanpark@snu.ac.kr

Abstract

Accurately estimating the walking direction of smartwatch users is critical for applications such as exercise trajectory analysis. This study introduces a novel approach that fuses estimation direction from inertial sensors and GPS. Inertial sensors provide stable estimates as they are unaffected by environmental conditions, but their accuracy can be impacted by sensor performance and user motions, such as hand swinging. GPS, in contrast, offers higher accuracy than inertial sensors under favorable signal conditions. To leverage the strengths of both sensors, the proposed method employs an information-weighted consensus filter, integrating direction estimates and error covariances. Experimental results demonstrate that the fusion approach reduces estimation errors compared to individual sensors.

Keywords: smartwatch, inertial sensor, walking direction, pedestrian inertial navigation

1. Introduction

Pedestrian localization systems using wearable devices are a field of active research. While the Global Navigation Satellite System (GNSS) is highly effective for localizing pedestrians outdoors, it is hindered by environmental limitations [1]. In urban environments, GNSS signals often experience multipath effects or signal attenuation, leading to reduced accuracy. To address this, map-matching methods have been proposed, but these require highly detailed maps and face challenges in predicting a pedestrian's walking trajectory.

Inertial sensors are employed to mitigate GNSS's environmental limitations. These sensors are already integrated into wearable devices, eliminating the need for additional infrastructure. Pedestrian Dead Reckoning (PDR) is a method that leverages inertial sensors to continuously estimate a pedestrian's position by calculating step length and walking direction [2]. However, when devices such as smartwatches are used, aligning the device's orientation with the pedestrian's walking direction becomes challenging.

Approaches utilizing Attitude and Heading Reference Systems (AHRS) to estimate walking direction often assume alignment between the device and the pedestrian's walking direction, which limits their general applicability. Other studies employ Principal Component Analysis (PCA) methods to extract dominant components from acceleration patterns during walking [3]. These methods provide useful results when the pedestrian swings their arm while walking. However, due to the limitations of inertial sensors, estimating the absolute walking direction remains challenging. Additionally, since these methods rely on the statistical characteristics of acceleration patterns, their accuracy may decrease depending on a pedestrian's walking style or arm-swinging habits.

In this paper, a walking direction estimation algorithm that integrates inertial sensors and GNSS is proposed. For inertial sensor-based walking direction estimation, a previously developed PCA-based algorithm is utilized [4]. For GNSS-based direction estimation, Recursive Least Squares (RLS) is applied to estimate the pedestrian's walking direction from positional data. The directions estimated from the two sensors are fused using the Information-Weighted Consensus Filter (ICF), which considers the error covariance of each sensor's estimation.

The structure of this paper is as follows. Section II introduces walking direction estimation methods and details the proposed algorithm. Section III presents experimental results demonstrating the performance of the proposed algorithm. Finally, Section IV provides conclusions and discusses future work.

2. Fusion method for walking direction

The proposed method estimates a user's walking direction and error covariance by integrating inertial sensors and GNSS data. First, a previously developed inertial sensor-based walking direction estimation method is employed, which extracts principal components from the acceleration distribution over several steps to estimate the user's direction. The core equation for the PCA-based method is shown below:

$$\hat{\theta}_{PCA} = \arg\max_{\theta} \left(\sum_{i=1}^{M} \left(\boldsymbol{\beta} \cdot \boldsymbol{\alpha}_i^n \right)^2 \right) \quad (1)$$

Here, $\hat{\theta}$ represents the walking direction, $\boldsymbol{\alpha}_i^n$ denotes the accumulated acceleration vector in the navigation frame over several steps, $\boldsymbol{\beta}$ represents a specific parameter (denoted as $\boldsymbol{\beta} = [\cos\theta \ \sin\theta]^{\mathrm{T}}$), and M is the number of steps. The estimation accuracy is influenced by the number of steps.

Since PCA does not inherently provide an index for estimation error, an error index for the proposed method is introduced. This index uses the eigenvalues of the two principal components derived during PCA to represent the degree of error. The error index can be expressed as the ratio of the two eigenvalues:

$$P_{PCA} = \tilde{\theta}^2, \ \tilde{\theta} = \tan^{-1}\left(\frac{\lambda_2}{\lambda_1} \right) \quad (2)$$

where λ_1 and λ_2 are the eigenvalues of the first and second principal components, respectively. A large λ_1 and a small λ_2 indicate a sharp distribution with minimal estimation error. Conversely, a large λ_2 suggests a broader acceleration distribution, making it challenging to estimate a clear direction. Figure 1 illustrates the concept of the error index.

For GNSS-based walking direction estimation, positional data are utilized, and the Recursive Least Squares (RLS) algorithm is applied, which provides computational efficiency based on the recursive formulation [5]. The estimated walking direction is expressed as follows:

$$\hat{x}_i = \hat{x}_{i-1} + K_i (y_i - \varphi_i \hat{x}_{i-1}), \quad (3)$$

$$K_i = \frac{P_{i-1}\varphi_i^T}{1 + \varphi_i P_{i-1} \varphi_i^T} \quad (4)$$

$$P_i = \frac{12}{\Delta t^2 N (N^2 - 1)} \quad (5)$$

$$K_i = \frac{P_{i-1}\varphi_i^T}{1 + \varphi_i P_{i-1} \varphi_i^T} \quad (6)$$

where y represents the GNSS positional measurements, $x = [v^T, p^T]^{\mathrm{T}}$ denotes the state variables, and $\varphi = [(i-1)\Delta t, 1]^T$ is the regressor. To calculate the error covariance, we leverage the relationship between the

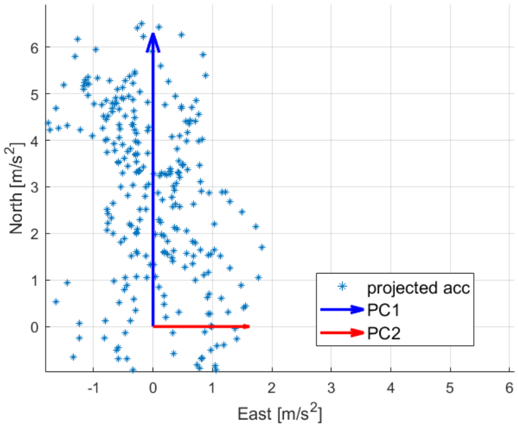


Fig. 1. Estimated walking direction error of PCA-based method.

Horizontal Dilution of Precision (HDOP) of the GNSS and the estimated velocity.

$$P_{GNSS} = P_i \cdot \frac{\sigma_{HDOP}^2}{2 \| \hat{v} \|^2} \quad (7)$$

GNSS-based direction estimation exhibits high accuracy during prolonged linear movements due to the clear tendency in the walking direction. However, during turns, the regression model becomes inconsistent, requiring calculations based only on short-term measurements.

Using the two methods described above, the walking direction and error covariance are calculated. The proposed method combines these using the Information-Weighted Consensus Filter (ICF) [6]. ICF is a widely used technique in sensor fusion, particularly effective when the sensors have varying reliability. It adjusts the weights of erroneous data to minimize their impact on the fusion results. The final direction calculated by the proposed method is expressed as:

$$\hat{\theta}_{fusion}^k = \hat{\theta}_{fusion}^{k-1} + \varepsilon W' \sum_{i \in \{PCA, GNSS\}} (\hat{\theta}_i - \hat{\theta}_{fusion}^{k-1}) \quad (8)$$

$$W_{fusion}^k = W_{fusion}^{k-1} + \varepsilon \sum_{i \in \{PCA, GNSS\}} (W_i - W_{fusion}^{k-1}) \quad (9)$$

$$W' = \frac{W_{fusion}^0}{W_{PCA} + W_{GNSS}} \quad (10)$$

where $W = P^{-1}$ represents the information matrix. This approach allows us to obtain a combined movement direction estimation by considering the contributions and error covariances of the individual sensors.

The ICF combines the two estimated directions through an iterative process, as outlined in equations (8) and (9). This process refines the fusion result by iteratively adjusting the weights of the estimated directions based on their respective error covariances. The iterative nature of ICF ensures that the final estimation accounts for the reliability of each input while minimizing the influence of erroneous data. This robust fusion mechanism is crucial for achieving high accuracy in walking direction estimation under varying conditions.

3. Experiment

To evaluate the performance of the proposed method, outdoor experiments were conducted. The participant wore a Samsung Galaxy Watch 5 on their left wrist, and data from the watch's inertial sensors and GNSS measurements were utilized to validate the estimation performance. For ground truth walking direction, the MTi-680G sensor from Xsens was employed. This sensor provides highly accurate positional information using RTK, which was differentiated to calculate the walking direction.

The experiments were conducted in an urban environment within an apartment complex, where the GNSS signal quality was poor, with HDOP values ranging from 6 to 12 meters. The participant walked for approximately 15 minutes, following a trajectory shaped like the number 9.

Figure 2 illustrates the estimated walking directions over the entire experiment. The black line represents the reference direction, the blue line corresponds to the PCA-based estimation, the red line represents the GNSS-based estimation, and the green line shows the result of the proposed fusion method. At around the 400-second mark, the participant waited at a crosswalk, resulting in no change in the estimated walking directions. Overall, it was observed that the GNSS-based walking direction exhibited large errors, while the PCA-based and fusion-based results demonstrated smaller errors.

Table 1. RMSE of estimated walking direction.

	PCA	GNSS	Fusion
RMSE [deg]	11.32	20.42	**10.98**

Table 1 presents the RMSE values for walking direction estimation using each method. The proposed method achieved lower errors compared to the individual sensor-based methods. Given the challenging GNSS conditions of the experimental site, the GNSS-based method showed significant errors, whereas the PCA-based method had smaller errors. Consequently, the fusion results of the proposed method were more influenced by the PCA-based estimation.

Figure 3 shows a zoomed-in view of the estimation errors during a straight-line walking segment. In this scenario, the errors of the PCA-based and GNSS-based results were similar, but the proposed method exhibited relatively smaller errors. This highlights the utility of the GNSS-based method in straight-line situations and provides justification for combining the two sensor-based methods in the proposed approach.

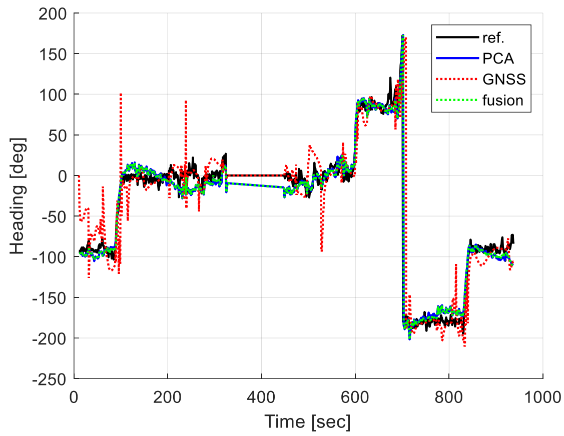


Fig. 2. Estimated walking direction results.

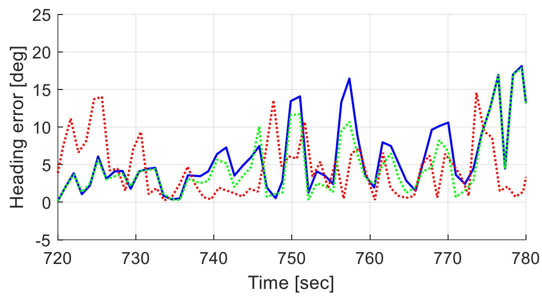


Fig. 3. Estimated walking direction results (straight trajectory).

4. Conclusion

In this study, a walking direction estimation algorithm was proposed, integrating inertial sensors and GNSS data using the ICF. The algorithm successfully addressed the limitations of individual sensor-based methods by leveraging the strengths of both PCA-based inertial sensor and GNSS measurements. Experimental results demonstrated that the proposed method achieved lower errors compared to standalone sensor approaches, particularly in urban environments with challenging GNSS conditions. Furthermore, the analysis highlighted the robustness of the fusion method in varying scenarios, such as straight-line and complex trajectories. These findings suggest that the proposed approach is effective for improving walking direction estimation in wearable devices, paving the way for more reliable pedestrian navigation systems.

Acknowledgment

This research was supported by the National Research Foundation of Korea funded by the Ministry of Science and ICT, the Republic of Korea, under Grant NRF-2022R1A2C2012166.

References

1. R. Sun, L. Fu, Q. Cheng, K.-W. Chiang, and W. Chen, ''Resilient pseudorange error prediction and correction for GNSS positioning in urban areas,'' IEEE Internet Things J., vol. 10, no. 11, pp. 9979–9988, Jun. 2023.
2. R. Harle, ''A survey of indoor inertial positioning systems for pedestrians,'' IEEE Commun. Surveys Tuts., vol. 15, no. 3, pp. 1281–1293, 3rd Quart., 2013.
3. S. A. Hoseinitabatabaei, A. Gluhak, R. Tafazolli, and W. Headley, ''Design, realization, and evaluation of uDirect—An approach for pervasive observation of user facing direction on mobile phones,'' IEEE Trans. Mobile Comput., vol. 13, no. 9, pp. 1981–1994, Sep. 2014.
4. J. W. Park, J. H. Lee, J. Park, and C. G. Park, "Smartwatch-Based Kinematic Walking Direction Estimation Using Paired Principal Component Analysis." IEEE Access, vol. 12, pp.27756-27767, 2024.
5. S. A. U. Islam and D. S. Bernstein. "Recursive least squares for real-time implementation [lecture notes]." IEEE Control Systems Magazine, vol. 39, no.3, pp.82-85, 2019.
6. A. T. Kamal, J. A. Farrell, and A. K. Roy-Chowdhury. "Information weighted consensus filters and their application in distributed camera networks." IEEE Transactions on Automatic Control, vol. 58, no.12, pp.3112-3125, 2013.

Authors Introduction

Mr. Jae Hong Lee

He received the B.S. degree from the School of Mechanical and Electrical Control Engineering, Handong Global University, in 2017, and the M.S. degree from the Department of Mechanical and Aerospace Engineering, Seoul National University, Seoul, South Korea, in 2019, where he is currently pursuing the Ph.D. degree. His research interests include pedestrian dead reckoning and inertial navigation systems.

Dr. Chan Gook Park

He received the B.S., M.S., and Ph.D. degrees in control and instrumentation engineering from Seoul National University, Seoul, South Korea, in 1985, 1987, and 1993, respectively. In 1998, he was as a Postdoctoral Fellow with Prof. J. L. Speyer about peak seeking control for formation flight with the University of California at Los Angeles, Los Angeles, CA, USA. From 1994 to 2003, he was an Associate Professor with Kwangwoon University, Seoul. In 2003, he joined the Faculty with the School of Mechanical and Aerospace Engineering, Seoul National University, where he is currently a Professor. In 2009, he was a Visiting Scholar with the Department of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA. His research interests include advanced filtering techniques, high-precision inertial navigation system (INS), visual-inertial odometry, INS/GNSS/IBN integration, and smartphone-based/foot-mounted pedestrian dead reckoning systems. Dr. Park was the Chair of IEEE AES Korea Chapter until 2009.

Multi-Frame Track-Before-Detect with Adaptive Number of Frame as Noise Level

Jehwa Lee
Department of Aerospace Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu
Seoul, 08826 Republic of Korea

Jae Hong Lee
Department of Aerospace Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu
Seoul, 08826 Republic of Korea

Chan Gook Park
Department of Aerospace Engineering/ASRI, Seoul National University, 1 Gwanak-ro, Gwanak-gu
Seoul, 08826 Republic of Korea
E-mail: blackhawk@snu.ac.kr, honglj@snu.ac.kr, chanpark@snu.ac.kr

Abstract

Multi-frame Track-Before-Detect (MF-TBD) is a batch processing method used to enhance detection and tracking performance in low SNR environments. Unlike traditional filtering techniques, MF-TBD does not apply thresholding and instead uses all observed data to reduce the risk of target loss. By integrating observations across multiple frames, it leverages space-time correlations to improve detection robustness. However, as the number of frames increases, the computational cost grows exponentially due to the need to correlate data over a larger dataset, leading to inefficiencies. Especially in high SNR conditions, where fewer frames are sufficient for accurate detection. To address this, we propose an Adaptive MF-TBD framework that dynamically adjusts the number of frames based on SNR levels.

Keywords: multi-frame track-before-detect, target tracking, adaptive number of frame, maneuvering target

1. Introduction

Target tracking is a critical process that estimates the state of a target based on continuous measurements, playing a key role in both military and commercial applications. Conventional tracking algorithms apply a threshold to each measurement, using filtered measurements for tracking [1]. However, these methods face significant challenges when the target's SNR (Signal-to-Noise Ratio) is low. In such cases, actual target-originating measurements may be filtered out by the threshold, leading to track loss [2]. To address this issue, Track-Before-Detect (TBD) was introduced, enabling tracking of targets with low SNR. However, TBD faces limitations, such as susceptibility to false alarms when clutter density is high, which reduces its tracking reliability. To improve upon these limitations, more advanced approaches like Multi-Frame Track-Before-Detect (MF-TBD) were developed [3].

MF-TBD improves tracking performance by integrating target energy across multiple consecutive frames, leveraging the spatiotemporal correlation of the target. This makes it particularly effective for weak signal targets. However, the primary drawback of MF-TBD is its computational complexity and high memory requirement. Each frame used for computation contains a large number of measurements, and processing several frames simultaneously significantly increases computational demand. This is because MF generates trajectory candidates for the target using a fixed number of consecutive frames. As the number of frames increases, the number of potential trajectory combinations grows exponentially, leading to a significant rise in computational load [4].

To overcome these limitations, this paper proposes a novel adaptive MF-TBD method that adjusts the number of frames based on signal conditions. Using the Mean Absolute Error (MAE) metric to quantify signal qualities [5], the framework minimizes computational overhead in high-SNR conditions (low MAE) by reducing the number of frames while maintaining performance. Conversely, in low-SNR conditions (high MAE), it retains a larger number of frames to ensure robust tracking performance. This adaptive strategy ensures optimal performance across diverse SNR conditions, achieving significant computational efficiency gains compared to conventional fixed-frame methods while maintaining tracking accuracy.

The adaptive MF-TBD framework and the use of the MAE metric for dynamic frame adjustment are detailed in Section II. Section III compares the performance of the proposed method with conventional TBD and fixed-frame MF-TBD in terms of accuracy and efficiency. Section IV concludes with a summary and discussion of its applicability to various tracking scenarios.

2. MF-TBD with adaptive number of frame

The target model assumes a point target performing constant velocity (CV) motion within a 2D surveillance region. The continuous target state $x_k = [x_k, \dot{x_k}, y_k, \dot{y_k}]$ evolves according to the following Markov process

$$\mathrm{x_k} = \mathrm{Fx_{k-1}} + \mathrm{w_k} \quad (1)$$

where F is the state transition matrix defined as

$$\mathrm{F} = I_{s/2} \otimes \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \quad (2)$$

T is the sampling interval, and $\mathrm{w_k}$ is the process noise.

The measurement model uses a pixelized approach commonly employed in TBD techniques to describe low-resolution sensors. The surveillance region is assumed to be divided into a grid of $N_x \times N_y$, with $\triangle_x$ and $\triangle_y$ representing the cell sizes in the x- and y-directions, respectively. The measurement at time k is expressed as

$$z_K(i,j) = \begin{cases} A_k + n_k^{H_1}, iftargetisin(i,j) \\ n_k^{H_0}, \ \text{otherwise} \end{cases} \quad (3)$$

where A_k represents the signal amplitude from the target, and $n_k^{H_1}$, and $n_k^{H_0}$ are samples from independent and identically distributed (i.i.d.) zero-mean Gaussian noise processes with variances ${\sigma_k^{H_1}}^2$ and ${\sigma_k^{H_0}}^2$, respectively.

Following structure of conventional MF-TBD framework is based on [6]. Merit function $I_k(x_k|Z_{1:K})$ evaluates the likelihood of the current state x_k being target-related while incorporating the transition probabilities from the previous state x_{k-1}. This function plays a critical role in identifying the most probable trajectory of the target over K consecutive frames. Here, K represents the number of consecutive frames used to estimate the trajectory, with larger K values enabling the algorithm to better leverage spatiotemporal correlations, even though at the cost of increased computational complexity. The merit function is computed as follows

$$I_k(x_k|Z_{1:K}) = \max_{\tau(\mathrm{x_k})} I_{k-1}(x_{k-1}|Z_{1:K-1}) + \lambda_k(x_k) \quad (4)$$

$$\tau(x_k) = \{|x_{k-1}\ y_{k-1}|^T \in \mathbb{R}^2: \\ |x_k - x_{k-1}| \leq B_1, |y_k - y_{k-1}| \leq B_2\} \quad (5)$$

The set of possible previous states, $\tau(x_k)$, represents all x_{k-1} states that can transition to x_k while satisfying the boundary constraints. Specifically, it includes states where the displacement in the x- and y-directions lies within the respective boundaries B_1 and B_2. These boundaries defined using the Kinematics Boundary Constraint (KBC) method [7], represent the maximum allowable transitions in each direction, ensuring that only realistic and feasible state transitions are considered based on the target's dynamics. $\lambda_k(x_k)$ denotes the likelihood of the current state x_k.

The function $\psi(x_k)$ identifies the most likely previous state x_{k-1} that maximizes the merit function I_{k-1} while meeting the constraints defined by $\tau(x_k)$. It is given by

$$\psi(\mathrm{x_k}) = \underset{\tau(\mathrm{x_k})}{\mathrm{argmax}}\, I_{k-1}(x_{k-1}|Z_{1:K-1}) \quad (6)$$

The state sequence history χ_i represents the recovered trajectory estimate of the target from initial state up to time k.

$$\chi_{1:k} = \begin{cases} \psi_{\mathrm{i}(\chi_{i+1}(\mathrm{x_k}))}, & i = 1, \dots, k-1 \\ x_i. & i = k \end{cases} \quad (7)$$

As the number of frames K increases, the merit function for the true target tends to improve due to the accumulation of spatiotemporal information across frames. However, this comes with a significant drawback: the set of candidate states $\tau(x_k)$, representing all possible transitions, grows exponentially with K, leading to a rapid increase in computational complexity. Furthermore, in high Signal-to-Noise Ratio (SNR) conditions, maintaining a high frame count is unnecessary, as the target signal is already strong and distinguishable.

So, the proposed method adjusts the number of frames K based on the SNR. Specifically, when the SNR is high, a smaller value of K is selected to minimize computational complexity, whereas a larger K is chosen under low SNR conditions to ensure robust tracking performance. To quantify the SNR level and make this adjustment, the Mean Absolute Error (MAE) is employed as a metric, calculated as:

$$\mathrm{MAE} = \frac{\sum|\mathrm{x}-\hat{\mathrm{x}}|}{n} \quad (8)$$

where x represents the observed values, $\hat{x}$ the estimated values, and n the total number of measurements. This SNR-dependent frame adjustment ensures both computational efficiency and tracking accuracy across varying signal environments.

3. Simulations

In this simulations, two metrics were used to evaluate performance: $P_{d-track}$, which measures the probability of the algorithm returning a valid track, and computation time. $P_{d-track}$ is defined as the case where the cell difference between the true target and the estimated target is within designed threshold ε. The experiments were conducted under the following conditions. The target maneuver constant velocity (CV) with 100m/s. The observation area was structured as a 100×100 grid, with each cell having a size of $\triangle_x$=$\triangle_y$= 15 m. The tolerance to $\epsilon =$ 4, and the maximum number of frames $K_{max} =$

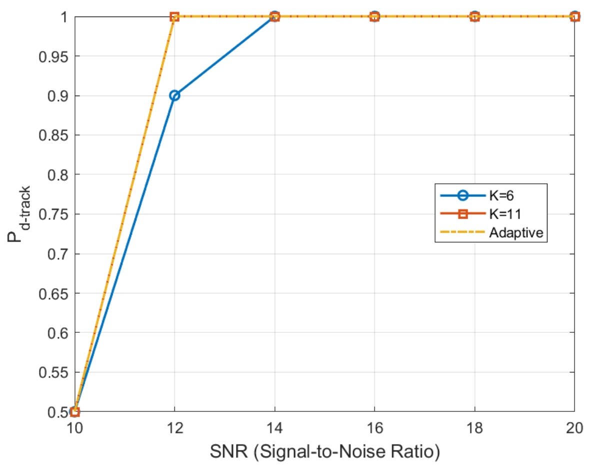


Fig. 1. $P_{d-track}$ per SNR.

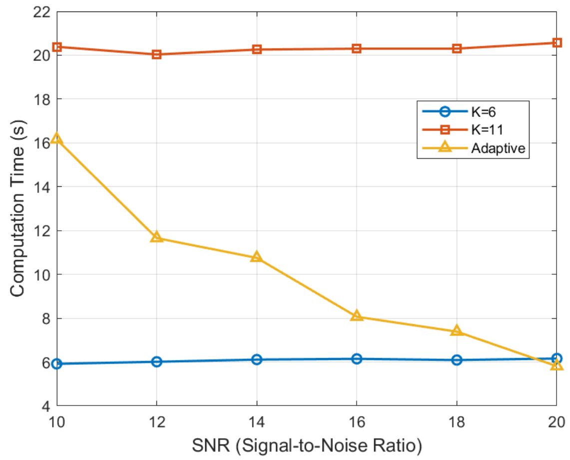


Fig. 2. Computation time per SNR.

11. The adaptive frame number K was adjusted within the range of 6 to 11.

Figure 1 shows that the $P_{d-track}$ values for each method (K=6, K=11, adaptive) as SNR level. The results show that the K=11 method outperforms the K=6 method in terms of tracking performance. Furthermore, the adaptive method achieves identical performance to the K=11 method, demonstrating its ability to maintain high tracking accuracy.

Figure 2 presents the computation time for each method as SNR level. When K=11, the large number of τ values that need to be processed results in higher computation time. Conversely, the computation time is significantly lower for K=6, as fewer τ values are required. The adaptive method exhibits selects a higher K value at low SNR, leading to increased computation time. However, as the SNR improves and the signal quality becomes better, the adaptive method reduces the K value, which progressively decreases the computation time. Proposed method demonstrated similar $P_{d-track}$ to the K=11 method, while significantly reducing computation time. Although it required more computation time than the K=5 method, it achieved better tracking performance. Overall, the results show that the proposed method strikes a favorable balance between maintaining high tracking performance and reducing computation time.

4. Conclusion

The proposed algorithm evaluates the signal quality using the MAE metric to appropriately adjust the number of frames, thereby reducing unnecessary computational costs. Compared to conventional MF-TBD method, it maintains similar tracking performance while significantly reducing computation time. This proposed algorithm will contribute to efficiently detecting and tracking maneuvering targets using the MF-TBD approach under various SNR conditions.

Acknowledgment

This research was supported by the National Research Foundation of Korea funded by the Ministry of Science and ICT, the Republic of Korea, under Grant NRF-2022R1A2C2012166.

References

1. Y. Bar-Shalom, “Tracking and data association”, J. Acoustical Soc. Amer., vol. 87, no. 2, pp. 918-919, 1990.
2. S. J. Davey, “SNR limits on Kalman filter detect-then-track”, IEEE Signal Process. Lett., vol. 20, no. 8, pp. 767-770, Aug, 2013.
3. Y. Barniv, “Dynamic programming solution for detecting dim moving targets”, IEEE Trans. Aerosp. Electron. Syst., vol. 29, no.1, pp. 44-56, Jan. 1993.
4. J. Wang, W. Yi, M. R. Morelande, and L. Kong, “A computationally efficient dynamic programming based track-before-detect”, In Proc. 2025 18th Int. Conf. Inf. Fusion, pp. 1558-1565, 2015.
5. T. Chai and R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)?”, Atmos. Chem. Phys., vol. 14, no. 3, pp. 1833-1845, Feb. 2014.
6. W. Yi, Z. Fang, W. Li, R. Hoseinnezhad, and L. Kong, “Multi-frame track-before-detect algorithm for maneuvering target tracking”, IEEE Trans. Veh. Technol., vol. 69, no. 4, pp. 4104-4118, Apr. 2020.
7. E. Grossi, M. Lops, and L. Venturino, “A novel dynamic programming algorithm for track-before-detect in radar systems”, IEEE Trans. Signal Process., vol. 61, no. 2, pp. 392-397, Jan. 2013.

Authors Introduction

Mr. Jehwa Lee

He received the B. S. degree in Weapon System Engineering in 2015 from Naval Academy, Changwon, South Korea. He is currently working toward the M.S. degree in Aerospace Engineering in Seoul National University, Seoul, South Korea. His research interests include target tracking for maneuvering target in low SNR situation.

Mr. Jae Hong Lee

He received the B.S. degree from the School of Mechanical and Electrical Control Engineering, Handong Global University, in 2017, and the M.S. degree from the Department of Mechanical and Aerospace Engineering, Seoul National University, Seoul, South Korea, in 2019, where he is currently pursuing the Ph.D. degree. His research interests include pedestrian dead reckoning and inertial navigation systems.

Dr. Chan Gook Park

He received the B.S., M.S., and Ph.D. degrees in control and instrumentation engineering from Seoul National University, Seoul, South Korea, in 1985, 1987, and 1993, respectively. In 1998, he was as a Postdoctoral Fellow with Prof. J. L. Speyer about peak seeking control for formation flight with the University of California at Los Angeles, Los Angeles, CA, USA. From 1994 to 2003, he was an Associate Professor with Kwangwoon University, Seoul. In 2003, he joined the Faculty with the School of Mechanical and Aerospace Engineering, Seoul National University, where he is currently a Professor. In 2009, he was a Visiting Scholar with the Department of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA. His research interests include advanced filtering techniques, high-precision inertial navigation system (INS), visual-inertial odometry, INS/GNSS/IBN integration, and smartphone-based/foot-mounted pedestrian dead reckoning systems. Dr. Park was the Chair of IEEE AES Korea Chapter until 2009.

Simplification of Rip Current Detection by Image Averaging Based on the Number of Wave Breaks

Ota Hamasuna[1], Leona Kimura[1], Satoshi Ikeda[1], Kaoru Ohe[1], Kenji Aoki[1], Amane Takei[1], Akihiro Kudo[2], Makoto Sakamoto[1*]
([1]Graduate School of Engineering, University of Miyazaki, Japan),
([2]National Institute of Technology, Tomakomai College, Japan)
[*]Corresponding Author
Email: saka2000@cc.miyazaki-u.ac.jp

Abstract

According to a National Police Agency report, there were 1,392 water accidents in 2023, with 368 victims (dead or missing) in the sea, mainly due to rip currents. Detecting rip currents is crucial, and past studies have used image averaging, often relying on fixed-point cameras or lengthy videos, making it difficult for individuals to apply. This study proposes using smartphone videos, with durations adjusted by the number of wave breaks, to enable easier rip current detection. To test this, smartphone footage was recorded at Hitotsuba Surf Point in Miyazaki Prefecture for analysis.

Keywords: Rip current, image averaging, image processing, wave breaking, SSIM

1. Introduction

According to a National Police Agency report [1], there were 1,392 water accidents in 2023, with 743 fatalities or missing persons. Furthermore, out of those 743 individuals, 368 were involved in water accidents at sea, with rip currents identified as a primary cause. In fact, in the author's hometown of Miyazaki Prefecture, an incident occurred on February 8, 2024, where a high school student was swept approximately 10 meters offshore at the Kaeda River mouth and drowned. Therefore, it is considered necessary to develop a system for detecting and visualizing rip currents. In addition, detecting and visualizing rip currents can provide benefits, such as helping anglers like the author find good casting points.

And as a method for detecting rip currents, "image averaging" is well-known (Fig. 1). This is a method for visualizing rip currents by converting video footage of the coastline into a sequence of images and averaging them into a single image. In previous studies, this method has been commonly used. However, most of them involve the use of fixed-point cameras or analyses that take a long time, which come with various constraints. Therefore, the purpose of this study is to propose a method that reduces the analysis time while also changing the video source from fixed-point cameras to handheld smartphone videos, enabling individuals to easily detect rip currents in various locations.

2. Rip Currents to Detect

Most previous studies have required long analysis times for image averaging. However, in the study by Shimada et al., "Investigation of Suitable Analysis Period of Time for Image Averaging to Detect Rip Current" [2], it was found that sudden rip currents can be detected in as little as 1 minute. Therefore, this study will focus on detecting these suddenly occurring rip currents.

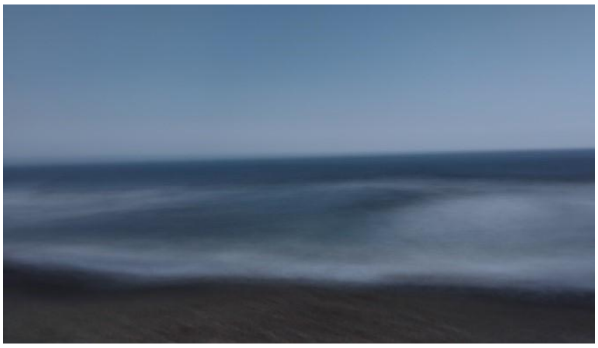
3. Proposed Method

We know that detection in 1 minute is possible through the work of Shimada et al. However, since the velocity of rip currents can reach 2 meters per second and we are assuming the use of hand-held videos from smart phones rather than fixed-point cameras, a shorter analysis time is required. Therefore, we propose a method to shorten the analysis time by using the number of wave breakings as the basis for the video used for image averaging, instead of simple time.

This proposal is based on the principle of detecting rip currents through image averaging. The reason why rip currents can be detected using image averaging lies in the characteristic that rip currents are less likely to cause wave breaking in the areas where they occur. When wave breaking does not occur, whitecaps are absent, and as a result, the pixel values in those areas become lower compared to the wave-breaking zones when image averaging is applied, creating vertical dark streaks (Fig. 2). Therefore, the number of wave breakings significantly impacts the results of image averaging. Based on this, we hypothesized that by setting the

Fig.1 Image averaging

analysis time according to the number of wave breakings, faster wave speeds could allow for shorter analysis times.



a. With rip current

b. Without rip current

Fig.2 Differences in the results of image averaging

4. Experiment

4.1. *Observation Sites and Dates*

Observations were conducted at the Hitotsuba Surf Point in Miyazaki Prefecture. This location is a high-risk area for water accidents, with frequent occurrences of rip currents (Fig. 3).

Observations were conducted a total of 108 times over four days, November 1st, 3rd, 5th, and 9th, 2024, regardless of the presence or absence of rip currents.

Fig.3 Hitotsuba Surf Point

4.2. *Experimental Environment*

The experimental environment is as shown in Table 1.

Table 1. Experimental environment

Smartphone	Redmi Note 10T
Handheld video	width : 1280px height : 720px fps : 30
Programming language	Python
Library	OpenCV
Program	convert.py : Video-to-Image Conversion ave.py : Image Averaging
Laptop	LAPTOP-LRONCBM2 OS : Windows 10

4.3. *Experimental Method and Evaluation Method*

A one-minute video was recorded at various points in the Hitotsuba Surf Point a total of 108 times. Afterward, each video was divided based on the number of wave breakings, and image averaging was performed on each segment.

The evaluation method utilized SSIM (Structural Similarity Index Measure) (Eq. (1) and (2)). The SSIM guidelines are established by JIMA (Japan Internet Media Association) [3]. The guidelines are shown in Table 2.

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (1)$$

$$\underset{\text{Luminance}}{\frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}}_{\text{Mean}} \times \underset{\text{Contrast}}{\frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}}_{\text{Standard Deviation}} \times \underset{\text{Structure}}{\frac{\sigma_{xy} + C_2/2}{\sigma_x\sigma_y + C_2/2}}_{\text{Covariance}} \quad (2)$$

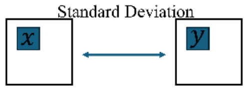


Table 2. Guidelines for SSIM

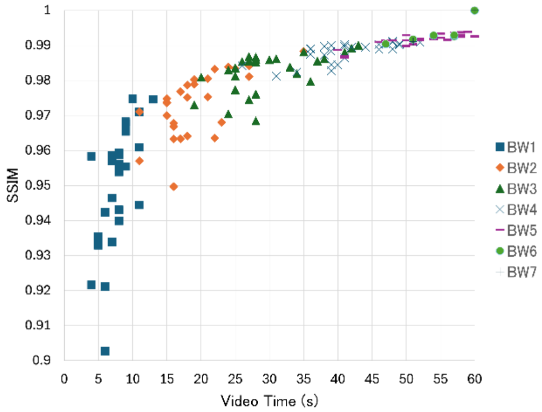
SSIM	Subjective Evaluation
0.98 ~	The original image and the compressed image are indistinguishable.
0.90 ~ 0.98	The degradation is noticeable when zoomed in.
~ 0.90	The degradation is clearly noticeable.

In this study, we aim to conclude with the number of wave breakings that reached a value close to 0.98.

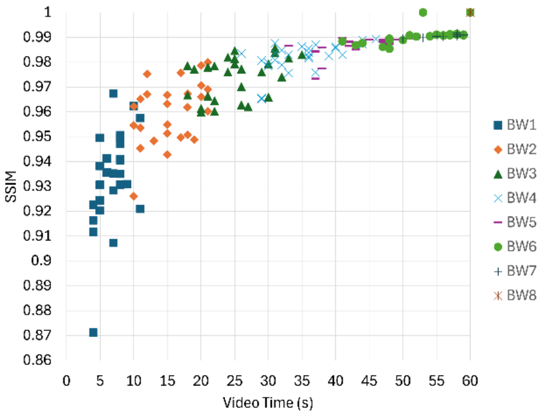
5. Results and Discussion

5.1. *SSIM Results for 108 Observations*

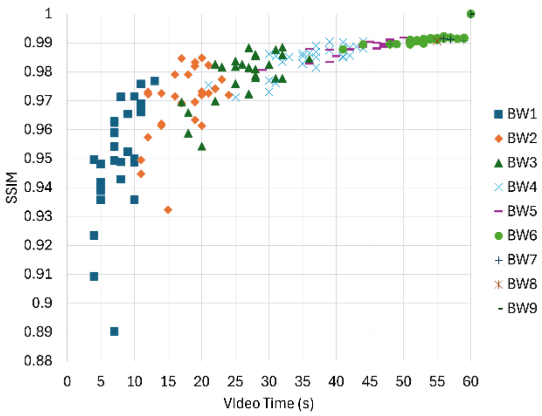
The results of 108 observations are shown in Fig. 4.



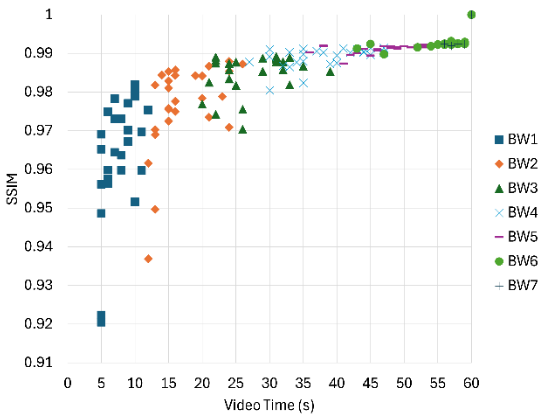
a. 2024/11/01



b. 2024/11/03



c. 2024/11/05



d. 2024/11/09

Fig.4 SSIM results

Fig. 4 a-d , it can be seen that the SSIM approaches 0.98 when the wave breaking count is 4. The average video duration for 4 wave breakings on November 1, 2024, was 41.2 seconds. Similarly, on November 3, 2024, the average video duration was 34.9 seconds, on November 5, 2024, it was 35.9 seconds, and on November 9, 2024, it was 38.1 seconds. And the total average duration was 37.5 seconds. This means a reduction of about 22 seconds compared to the conventional method.

5.2. *Results for Rip Currents Only*

The results based only on images showing rip currents among the 108 observations are shown in Fig. 5. There were 31 images that showed rip currents.

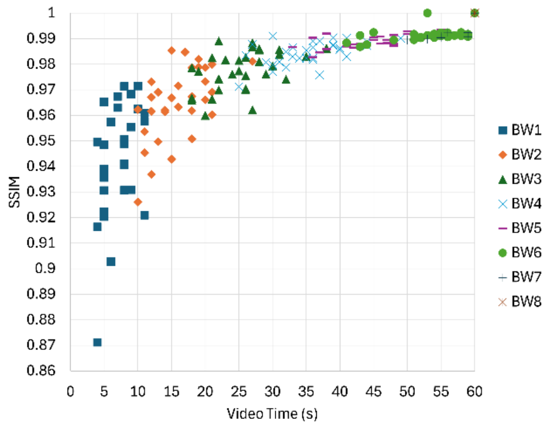


Fig.5 Rip currents only

The results for rip currents only, as shown in Fig. 5, also indicate that the SSIM approaches 0.98 when the wave breaking count is 4. And the average video duration for a wave breaking count of 4 was 35.4 seconds. This means a reduction of about 24 seconds compared to the conventional method.

6. Conclusion

We proposed a rip current detection method using image averaging with videos based on the number of wave breakings. As a result of 108 observations and experiments conducted at Hitotsuba Surf Point, it was found that the results obtained with 4 wave breakings were nearly equivalent to those obtained using the conventional 1-minute method. When considering the video duration, this method allows for a time reduction of approximately 22 to 24 seconds.

Appendix

The author is an angler. In this study, the points where rip currents were detected were recorded, and fishing was conducted around those points. Although the conditions were different, the catch was 2.4 times higher compared to last year. In particular, I was able to catch fish-eaters such as flatfish (Fig. 6) and Lateolabrax latus (Fig. 7). Rip currents tend to gather bait in their vicinity, making it an ideal casting point. Detecting rip currents can provide such benefits as well.


Fig.6 Flatfish


Fig.7 Lateolabrax latus

References

1. National Police Agency, “令和５年における水難の概況等”[online](accessed:2024/12/22).
2. R. Shimada, T. Ishikawa and T. Komine, “Investigation of Suitable Analysis Period of Time for Image Averaging to Detect Rip Current”, Journal of Japan Society of Civil Engineers Ser B2, vol. 76, no.2 pp.I_1339-I_1344, 2020.
3. JIMA, “電子化文書の画像圧縮ガイドライン”[online] (accessed:2024/12/22)

Authors Introduction

Mr. Ota Hamasuna

He is a master student at Department of Computer Science and System Engineering, University of Miyazaki. His current research interests are rip currents, augmented reality, computer graphics, and so on.

Ms. Leona Kimura

She is a master student at Department of Computer Science and System Engineering, University of Miyazaki. His current research interests are image processing, machine learning, and so on.

Dr. Satoshi Ikeda

He received PhD degree from Hiroshima University. He is an associate professor in the Faculty of Engineering, University of Miyazaki. His research interest includes graph theory, probabilistic algorithm, fractal geometry and measure theory.

Prof. Kaoru Ohe

She received her Ph.D. degrees from University of Miyazaki, Japan, in 2014. Currently she is an Associate Professor of the Center for Science and Engineering Education, Faculty of Engineering. Her research is separation engineering especially adsorption hazardous heavy metals and oxyanions

Dr. Kenji Aoki

He received Ph.D. of Engineering from Kagoshima University in 2010. He is currently working in Information Technology Center at University of Miyazaki as Associate Professor, since 2010. His research interests include bio-informatics, evolutionary computation, information system and Intelligent systems. He is a member of IPSJ and JSET.

Dr. Amane Takei

He is working as a professor for Department of Electrical and systems Engineering, University of Miyazaki. His research interest includes high performance computing for computational electromagnetism, iterative methods for the solution of sparse linear systems, domain decomposition methods for large-scale problems. Prof. Takei is a member of IEEE, an expert advisor of IEIC, a delegate of Kyushu branch of IEEJ, a director of JSST.

Prof. Akihiro Kudo

He received Ph.D. degree from Nagaoka University of Technology He is a professor in the Department of engineering for innovation , National Institute of Technology, Tomakomai college. He is a member of Acoustical society of Japan, Information and Communication Engineers (IEICE).

Prof. Makoto Sakamoto

He is presently a professor in the Faculty of Engineering, University of Miyazaki. His first interests lay in hydrodynamics and time series analysis, especially the directional wave spectrum. He is a theoretical computer scientist, and his current main research interests are automata theory, languages and computation. He is also interested in digital geometry, digital image processing, computer vision, computer graphics, etc.

Automated Classification of High-Grade Dried Shiitake Mushrooms Using Machine Learning

Leona Kimura[1]
Email: tg24020@student.miyazaki-u.ac.jp

Ota Hamasuna[1], Kaoru Ohe[1], Satoshi Ikeda[1], Kenji Aoki[1],
Amane Takei[1], Akihiro Kudo[2], Kazuhide Sugimoto[3], Makoto Sakamoto[1*]
([1]Graduate School of Engineering, University of Miyazaki, Japan),
([2]National Institute of Technology, Tomakomai Collage, Japan),
([3]SUGIMOTO Co., Ltd., Japan)
**Corresponding Author*
Email: saka2000@cc.miyazaki-u.ac.jp

Abstract

This study aims to automate shiitake mushroom sorting using an anomaly detection system with Autoencoders (AE) trained on acceptable product data. Initial experiments using CNN approaches highlighted challenges in achieving high accuracy for acceptable product classification, necessitating improvement. The AE-based approach showed progress in detecting defective products via data cleansing, augmentation, and training optimization. However, misclassification of acceptable products with features like darker areas or complex textures remains an issue. This presentation outlines current findings and strategies, including data expansion and model improvements, to address these challenges.

Keywords: Autoencoders (AE), Convolutional Neural Networks (CNN), Model Optimization, Shiitake Mushrooms

1. Introduction

Sorting tasks of shiitake mushrooms have long relied on the experience of skilled workers. However, due to aging and labor shortages, there is a strong demand to improve efficiency and automate the sorting tasks traditionally performed by humans. Specifically, the visual evaluation of shiitake mushrooms, based on attributes such as color and shape, has been challenging to automate due to the ambiguous and subjective nature of sorting criteria.

This study, conducted in collaboration with Sugimoto Shoten in Takachiho Town, Miyazaki Prefecture, aims to develop an automatic discrimination system using machine learning. The study focused on addressing the following two challenges:

Data Scarcity and Bias: How to utilize limited data, especially given the difficulty of collecting defective product data.

Improving Identification Accuracy: Designing algorithms capable of handling the variability and complexity inherent to natural objects.

2. Environment

In this study, the following environment was used:

Hardware:
CPU: Intel Core i7-11800H
GPU: NVIDIA RTX3050Ti
RAM: 16GB

Software:
Anaconda(Python/Jupyter Notebook)
Google Colaboratory

Library:
OpenCV: For data preprocessing (background removal, noise reduction)
PyTorch: For model construction and training
Scikit-learn: For statistical analysis and evaluation

In this study, we divided the time-consuming training tasks between Anaconda for the training phase and Google Colaboratory for execution. This approach distributed the high-load processing tasks, ensuring stable execution. Additionally, given the extensive use of images, we constructed an efficient data pipeline for image processing and machine learning, automating data augmentation and cleansing processes.

3. Research Methods

3.1 Data Collection and Preparation

In this study, shiitake mushroom samples were collected from Sugimoto Shoten as follows:
Good Quality Data: 3,564 images
Defective Data: 864 images
To address the data scarcity, data augmentation techniques such as rotation, scaling, and noise addition were applied, resulting in more than doubling the total data volume.

3.2 Data creasing

Using OpenCV, we removed the background[1] and applied a process to emphasize the shape of the shiitake mushrooms. During the process, there were instances where the background was not properly removed and excessively adapted to the background. For complex background data, manual corrections were made to minimize the impact of noise (Fig.1, Fig.2).

Fig.1 Before image

Fig.2 After Data Cleansing (Background Removal)

3.3 Model Design and Training Process

Autoencoder (AE):
The AE[2] is a method that learns only good quality data and detects anomalies based on reconstruction errors. Specifically, the model was constructed through the following steps:

1. Use good quality images as input data.
2. Calculate the mean and variance of reconstruction errors to set the identification threshold.
3. If the reconstruction error exceeds the threshold, the item is determined to be defective.

Convolutional Neural Network (CNN):
For the CNN, we utilized labeled data with annotations and conducted supervised learning. The following elements were incorporated:

1. Applied weighting to each layer of data to reduce bias.
2. Added preprocessing steps such as edge detection to emphasize features of defective items (e.g., color irregularities, shape anomalies).

3.4 Improvement Process

To mitigate misclassification due to data bias and insufficient training, the following improvements were implemented (Fig.3):

1. Increased the number of training iterations from 30 to 100, and then to 500, to analyze the effects of overfitting.
2. Focused data augmentation on misclassified data to enhance the model's generalization performance.
3. Utilized Google Collaboratory to enable high-precision computational processing.

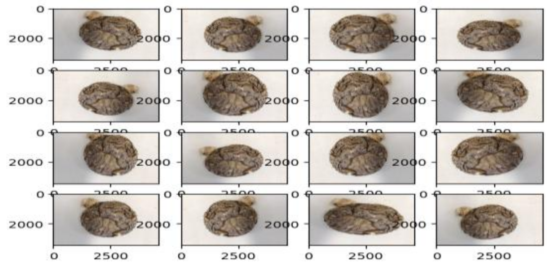


Fig3 Examples of Data

4 Results

This time, we were able to conduct up to four rounds. We reviewed up to three rounds, and in the fourth round, we decided to check the output results of overfitting.
Below are the details of the results and improvements.

First Trial:
Accuracy: Good Quality 20%, Defective 10%
Issues: Misclassifications frequently occurred due to the mixture of front and back images (Fig. 4, Fig. 5). Additionally, there were many backgrounds present (Fig. 4), leading to the hypothesis that the model excessively adapted to the background.


Fig.4 front image


Fig.5 back image

Improvement: Limited to front images and manually removed the background (Fig. 6, Fig. 7).


Fig.6 Images with a lot of background that couldn't be removed by OpenCV


Fig.7 Manually removed images

Second Trial:
Accuracy: Good Quality 14.68%, Defective 80.59%
Issues: The number of training iterations was limited to around 30, resulting in poor generation similar to the "Reconstructed" images.
Consequently, the reconstruction error values could not be determined, leading to a low classification rate (Fig.8).

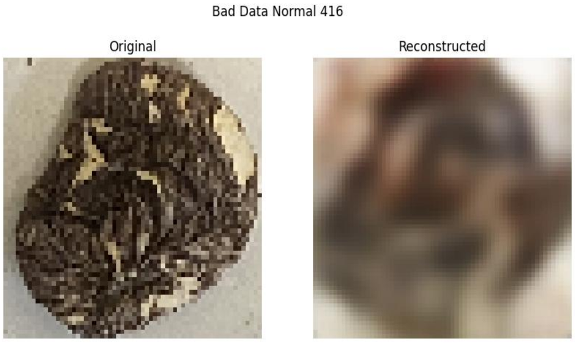


Fig.8 Misclassification of detective items

Improvement: Increase the Number of Training Iterations (Fig.9).

Third Trial:
Accuracy: Good Quality 24.36%, Defective 81.24%
Issues: Dirt on the platform caused noise, leading to misclassifications.
Improvement: Focused on augmenting data that includes dirt and corrected the training bias.

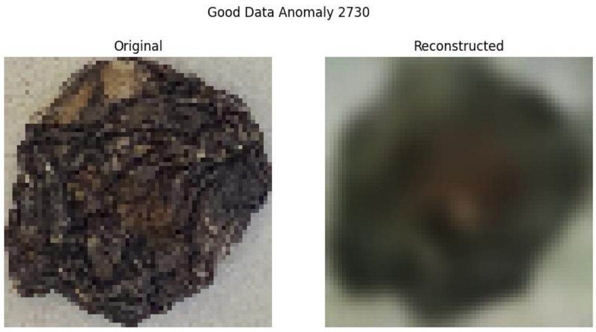


Fig.9 Misclassification of Good Quality Items with High Noise

Fourth Trial:
Accuracy: Good Quality 20.1%, Defective 83.5%
Issues: While overfitting caused color reproduction to deteriorate, the shapes were accurately generated (Fig.10).

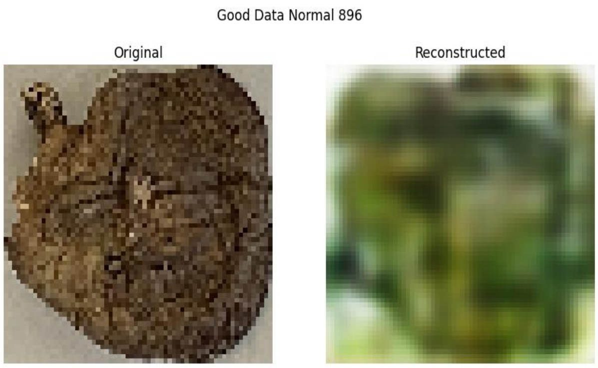


Fig.10 Output Result

Improvement:
By reviewing the output results during overfitting, we recognized the importance of early stopping.

5. Future Work

In the future, we will address the following issues:

Utilization of CNN with Annotations:
By using data labeled with details of defective parts, we aim to improve identification accuracy. We plan to test multiple CNN architectures such as ResNet and EfficientNet.

Enhancement of Data Collection:
We will request video data of good quality products from companies in Takachiho, Miyazaki Prefecture, to create a large-scale dataset. Additionally, we will utilize public datasets to compare identification performance.

Comparison with Other Methods:
We will try methods other than AE and CNN (for example, decision trees and support vector machines) to identify the best model for the specific characteristics of shiitake mushrooms.

The insights gained from this research can be applied to the identification of other natural objects besides shiitake mushrooms, and further development is expected.

6. Conclusion

This paper describes the basic study on automatic sorting of high-quality dried shiitake mushrooms. In the future, we would like to increase the amount of detailed learning data and strive to improve the recognition rate. Finally, this work was supported by JSPS KAKENHI Grant Numbers JP24K0792901 and JP24K15516.

Reference

1. Dried shiitake mushroom grade recognition using D-VGG network and machine vision [Accessed: 2024/12/22].
2. Development of Automated Visual Inspection Technology Using Deep Learning [Accessed: 2024/12/22].

Authors Introduction

Ms. Leona Kimura

She is a master student at Department of Computer Science and System Engineering, University of Miyazaki. His current research interests are image processing, machine learning, and so on.

Mr. Ota Hamasuna

He is a master student at Department of Computer Science and System Engineering, University of Miyazaki. His current research interests are augmented reality, computer graphics, and so on.

Dr. Kaoru Ohe

She received her Ph.D. degrees from University of Miyazaki, Japan, in 2014. Currently she is an Associate Professor of the Center for Science and Engineering Education, Faculty of Engineering. Her research is separation engineering especially adsorption hazardous heavy metals and oxyanions

Dr. Satoshi Ikeda

He received PhD degree from Hiroshima University. He is an associate professor in the Faculty of Engineering, University of Miyazaki. His research interest includes graph theory, probabilistic algorithm, fractal geometry and measure theory.

Dr. Kenji Aoki

He received Ph.D. of Engineering from Kagoshima University in 2010. He is currently working in Information Technology Center at University of Miyazaki as Associate Professor, since 2010. His research interests include bio-informatics, evolutionary computation, information system and Intelligent systems. He is a member of IPSJ and JSET.

Dr. Amane Takei

He is working as a professor for Department of Electrical and systems Engineering, University of Miyazaki. His research interest includes high performance computing for computational electromagnetism, iterative methods for the solution of sparse linear systems, domain decomposition methods for large-scale problems. Prof. Takei is a member of IEEE, an expert advisor of IEIC, a delegate of Kyushu branch of IEEJ, a director of JSST.

Dr. Akihiro Kudo

He received Ph.D. degree from Nagaoka University of Technology He is a professor in the Department of engineering for innovation, National Institute of Technology, Tomakomai college. He is a member of Acoustical society of Japan, Information and Communication Engineers (IEICE).

Mr. Kazuhide Sugimoto

He was Born in Takachiho town, Miyazaki Prefecture. After working in sales in the food service and apparel industries, he joined SUGIMOTO Co., Ltd., a wholesaler of dried shiitake mushrooms produced in Takachiho town, in 2011 after the Great East Japan Earthquake. In response to the current harsh situation, such as aging contract farmers and sluggish demand for dried shiitake mushrooms, he decided to protect the producers by promoting new business development, which he had experienced in sales. New items developed using shiitake mushrooms from Takachiho Township have become a standard item at supermarkets and department stores outside the prefecture. In March 2020, he was appointed as Representative Director, and in 2021, he was selected as a Small and Medium Enterprise Supporter and GFP Ambassador.

Dr. Makoto Sakamoto

He is a professor in the Faculty of Engineering, University of Miyazaki. He is a theoretical computer scientist, and his current main research interests are computer science and information processing.

Development of a plant growing experience application for physically challenged children using VR

Masatoshi Beppu[1], Masatomo Ide[1], Kaoru Ohe[1], Satoshi Ikeda[1], Kenji Aoki[1], Amane Takei[1], Akihiro Kudo[2], Makoto Sakamoto[1*]
([1]Graduate School of Engineering, University of Miyazaki, Japan),
([2]National Institute of Technology, Tomakomai Collage, Japan)
[*]Corresponding Author

Abstract

In 2016, the "first year of VR," many VR platforms emerged, making VR technology more accessible. Currently, technology is expected to be applied and utilized in various fields. Application to the education sector is being promoted as part of the educational use of ICT. However, it is difficult to get the benefits of implementing VR due to lack of technology and equipment for teachers. Therefore, it is necessary to limit the scope of coverage. This research will focus on limb-challenged children and develop a VR application that allows them to experience plant growing. We believe that this will solve the problems that have been a concern for children with physical disabilities, such as the inability to perform exercises using soil and the lack of opportunities for trial-and-error. In this study, we also asked men and women in their teens to 40s to experience the apps we developed and obtained their evaluations through questionnaires. Within the survey, we received certain evaluations in areas such as trial and error. As for future issues, the application will be improved based on the feedback received from the survey. In addition, we believe it is necessary to evaluate the long-term effects of the application by having children with physical disabilities use it.

Keywords: VR, physically handicapped child, education, plant breeding

1. Introduction

1.1. What is a physically challenged child?

The term "physically handicapped children" refers to "children who are born or born with impairments, or who have motor disabilities in their hands, feet, or spine due to illness or accident at an early age" [1].
According to the Ministry of Health, Labor and Welfare, there were 50,100 physically handicapped children in Japan in 2006[2]. In the Children's Charter established by the Ministry of Education, Culture, Sports, Science and Technology, it is stated that "every child shall be given the protection of proper treatment and education in case of physical incapacity or inadequate mental functioning.

1.2. Background

Many physically handicapped children spend time in hospitals. This makes it difficult for them to participate in out-of-school learning and experiential learning. In addition, when students go on field trips or experience learning, teachers and others around them help them to experience success because it is a valuable opportunity, and as a result, there are few opportunities for them to make mistakes and trial-and-error experiences. Other problems include the difficulty of long-term experiential learning due to physical condition and other issues, and the limited activities that can be experienced because of the hospitalization. The impact of these problems has made it difficult to achieve the Children's Charter established by the Ministry of Education, Culture, Sports, Science and Technology. In this study, we will develop a VR application that can simulate off-campus learning, which is difficult to do in hospitals, in order to solve the above problems.

1.3. Virtual Reality (VR)

VR is a technology that creates a virtual environment on a computer and gives the illusion of an actual experience. In 2016, VR equipment became more readily available to the public, and it has been called the "first year of VR". VR technology is expected to play an active role in the field of education. Examples of actual use of VR technology in the field of education will be presented.

1.3.1. Google for education

A comprehensive solution offered by Google Inc. for educational institutions that, with the Expeditions app and a head-mounted display (HMD), allows students to experience natural phenomena and other experiences while in the classroom.

1.3.2. Science experiment simulation application [3]

This application allows users to simulate four fields: "oblique projection," "how to use a gas burner," "the moon and the sun," and "all-sky planetarium. Experiments on events that are difficult to conduct with large experimental equipment or experiments can be conducted in a virtual environment.

2. Methodology

2.1. Development environment

The applications developed in this study were developed in the environment shown in Table 1.

Table 1 Development environment

OS	Windows 10 Pro
Unity	2021.3.0f1
VR Devices	Meta Quest 2
CG Model Creation	Blender 3.1.2

2.2. Prior research

In order to understand what functions are actually needed in the development process, we conducted a survey of teachers working at Kiyotake Seiryu Support School.

The following issues were identified within the research.

- Practical training using soil and other materials is not possible due to the long time spent in the hospital.
- During the training, there is a lot of help, and the individual is not able to experience trial-and-error.
- Difficulty in conducting long-term practical training

2.3. Preliminary experiment

In addition to the problems identified in the preliminary survey, a VR trial session was held as a preliminary experiment for students enrolled in the Kiyotake Seiryu Support School in order to confirm the problems when the target students use the system.

The following problems were identified as a result of the implementation.

- Difficult to perform complex operations
- Difficulty in securing VR equipment to the head

2.4. Development objectives

The application to be developed in this study will implement the following functions to solve the problems identified in the preliminary survey and preliminary experiments.

- Plant growth status changes depending on the behavior of the experiencer.
- Short experience and can quickly move on to the next experience.
- Simple operation, no need to move the body significantly

2.5. VR application

The newly developed application allows users to experience growing mini tomatoes. The experience can be divided into four growing stages, and the growing stage changes depending on the actions taken. The following is a summary of each stage of development and the actions that can be taken. In all phases, a condition is imposed that the water be done once or twice. Other conditions exist for each stage. A start screen is displayed at the beginning of each stage, and a failure screen is displayed when a failure occurs.

2.5.1. Growing stage

2.5.1.1. Seeding period

The stage of sowing mini-tomato seeds. Failure will occur if the condition is not achieved. An additional condition is that the seeds must be sown (Fig 1).

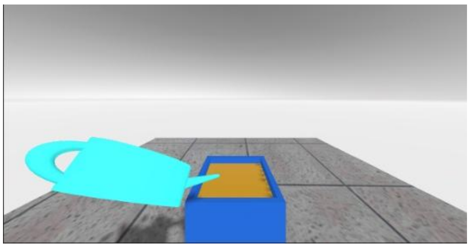


Fig 1 Seeding period

2.5.1.2. Germination period

The stage when mini tomatoes have germinated (Table 2). Failure occurs when conditions are not met. The stage of growth that is transferred changes depending on the percentage of weeds remaining (Fig 2).

Table 2 Germination period transition destination

watering	Weed cutting	Transition Destination
1~2	pull out everything	highest growth period
1~2	pull out a little	middle growth period
1~2	Not pulling out at all.	least growth period
0,3~	All conditions	failure

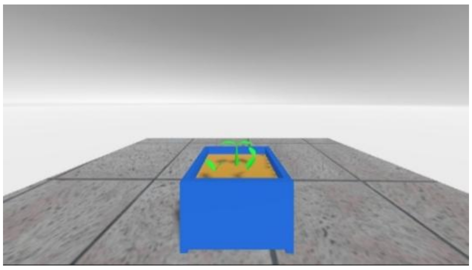


Fig 2 Germination period

2.5.1.3. Growth period

The stage in which mini tomatoes grow and flowers. It is divided into three stages depending on its behavior during the germination stage.

2.5.1.3.1. Highest growth period

The stage of greatest growth during the growth phase (Table 3). Failure to achieve the conditions will result in failure. The growth stage to be transitioned changes depending on the percentage of weeds remaining (Fig 3).

Table 3 Highest growth period transition destination

watering	Weed cutting	Transition Destination
1~2	pull out everything	highest harvesting period
1~2	pull out a little	middle harvesting period
1~2	Not pulling out at all.	least harvesting period
0,3~	All conditions	failure

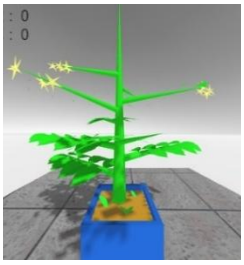


Fig 3 Highest growth period

2.5.1.3.2. Middle growth period

The second largest growth stage in the growing season. Failure occurs if the conditions are not met or if all the weeds remain unexcluded. The percentage of weeds remaining changes the stage of growth to which the weeds are transferred (Table 4).

Table 4 Middle growth period transition destination

watering	Weed cutting	Transition Destination
1~2	pull out everything	middle harvesting period
1~2	pull out a little	least harvesting period
1~2	Not pulling out at all.	Failure
0,3~	All conditions	Failure

2.5.1.3.3. Least growth period

The smallest growing stage of the growing season. Failure occurs when conditions are not achieved or weeds remain (Table 5).

Table 5 Least growth period transition destination

watering	Weed cutting	Transition Destination
1~2	pull out everything	least harvesting period
1~2	Not pulling out at all. pull out a little	Failure
0,3~	All conditions	Failure

2.5.1.4. Harvesting period

The stage at which mini tomatoes have grown and are ready to be harvested. The number of mini tomatoes that can be harvested varies depending on the stage of growth (Table 6,Fig 4).

Table 6 Harvesting period transition destination

growth stage	Number of harvestable
Highest	10
Middle	5
Least	3

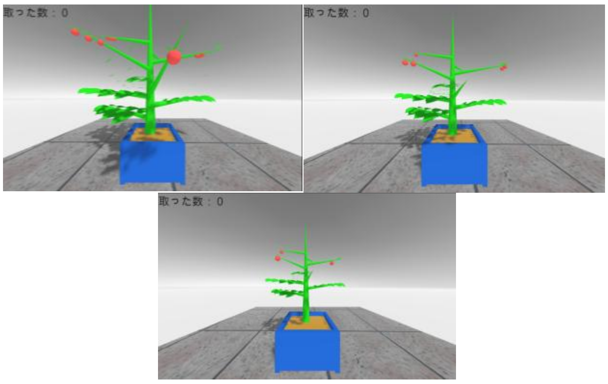


Fig 4 Germination period (upper right:highest, upper left:middle, under:least)

2.5.2. Method of operation

Actions in the application are performed with the controller that comes with Meta Quest2. All operations are performed with buttons, and the tilt function and other functions of the controller are not used.

3. Evaluation experiment

To confirm the effectiveness of the application developed this time, an evaluation experiment was conducted with men and women in their teens to 40s.

After the two trials, the subjects were asked to answer a questionnaire to evaluate the effectiveness of the application.

3.1. Questionnaire content

The questionnaire has eight items, and in addition to age and gender, the respondents are asked to answer on a five-point scale about the feeling of operation (i), whether it was enjoyable (ii), whether it could be considered an alternative to the experience of growing mini-tomatoes (iii), and screen sickness (iv). They were also asked to respond to the results of the first and second growing experience.

3.2. experimental results

We asked 19 men and women, ranging in age from under 10 to 40, to use the application and answer a questionnaire (Table 7).

Table 7 Questionnaire results

age	sex	i	ii	iii	iv	1st	2nd
20	M	4	4	5	2	0	10
10	M	3	3	2	5	0	0
40	M	2	2	2	5	0	3
20	M	2	4	2	5	0	5
20	M	5	2	3	4	0	0
20	M	4	4	4	5	0	0
U10	W	5	5	5	3	0	10
10	M	3	5	5	5	0	3
U10	M	2	5	5	5	5	10
U10	W	2	5	5	5	5	10
10	W	5	5	5	5	0	10
10	W	5	5	5	5	0	3
10	M	4	5	3	5	0	3
U10	W	5	5	5	3	0	3
20	M	5	5	5	5	10	10
10	M	2	4	1	4	0	10
10	M	4	5	2	5	0	10
10	M	4	5	5	5	0	10
20	M	3	4	3	5	0	10

4. Consideration

The results of the second run were better than the results of the first run for all ages. Question 3, which asked whether the program would be a substitute for the actual mini tomato growing experience, also received many favorable responses. Based on the above, we believe it is possible to develop an application that meets the objectives.

5. Conclusion

In this study, a plant growing experience application was developed with the goal of providing educational support to children with physical disabilities. Based on the results of the questionnaire, we believe that the goal was achieved. As a future issue, it is necessary to investigate the long-term effect of the art collection using limb-challenged children as subjects.

Finally, this work was supported by JSPS KAKENHI Grant Numbers JP24K0792901 and JP24K15516.

References

1. Japanese Association for the Physically Handicapped. https://www.nishikyo.or.jp/about/index.html
2. Cabinet Office "white paper on disability Appendix 8 Number of children and persons with disabilities 2013 Edition https://www8.cao.go.jp/shougai/whitepaper/h25hakusho/zenbun/furoku_08.html
3. Kouta Miyamoto, バーチャル技術による小学生向け教育支援に関する基礎的研究, 2022 https://cir.nii.ac.jp/crid/1050015520409327616

Authors Introduction

Mr. Masatoshi Beppu

He is a master student at Department of Computer Science and System Engineering, University of Miyazaki. His current research topic is educational support using VR for children with physical disabilities.

Mr. Masatomo Ide

He is a graduate student at the Graduate School of Engineering, University of Miyazaki. His recent main research activity is to utilize various digital technologies such as web development to contribute to the local community.

Dr. Kaoru Ohe

She received her Ph.D. degrees from University of Miyazaki, Japan, in 2014. Currently she is an Associate Professor of Center for Science and Engineering Education, Faculty of Engineering. Her researches are separation engineering especially adsorption hazardous heavy metals and oxyanions.

Dr. Satoshi Ikeda

He received PhD degree from Hiroshima University. He is an associate professor in the Faculty of Engineering, University of Miyazaki. His research interest includes graph theory, probabilistic algorithm, fractal geometry and measure theory

Dr. Kenji Aoki

He received Ph.D. of Engineering from Kagoshima University in 2010. He is currently working in Information Technology Center at University of Miyazaki as Associate Professor, since 2010. His research interests include bio-informatics, evolutionary computation, information system and Intelligent systems. He is a member of IPSJ and JSET.

Dr. Amane Takei

He is working as a professor for Department of Electrical and systems Engineering, University of Miyazaki. His research interest includes high performance computing for computational electromagnetism, iterative methods for the solution of sparse linear systems, domain decomposition methods for large-scale problems. Prof. Takei is a member of IEEE, an expert advisor of IEIC, a delegate of Kyushu branch of IEEJ, a director of JSST.

Dr. Akihiro Kudo

He received Ph.D. degree from Nagaoka University of Technology He is a professor in the Department of engineering for innovation , National Institute of Technology, Tomakomai college. He is a member of Acoustical society of Japan, Information and Communication Engineers (IEICE).

Dr. Makoto Sakamoto

He is presently a professor in the Faculty of Engineering, University of Miyazaki. His first interests lay in hydrodynamics and time series analysis, especially the directional wave spectrum. He is a theoretical computer scientist, and his current main research interests are automata theory, languages and computation. He is also interested in digital geometry, digital image processing, computer vision, computer graphics, etc.

Exploring Social Media's Role in Predicting Stock Market Trends

Masatoshi Beppu[1], Masatomo Ide[1], Seita Nagashima[2], Satoshi Ikeda[1]*, Amane Takei[1], Makoto Sakamoto[1], Tsutomu Ito[3], Takao Ito[4]
([1]Graduate School of Engineering, University of Miyazaki, Japan),
([2]MEITEC CORPORATION, Japan),
([3]National Institute of Technology, Ube College, Japan),
([4]Hiroshima University, Japan)

*Corresponding Author
E-mail: bisu@cs.miyazaki-u.ac.jp

Abstract

This study analyzes tweets from the official Twitter accounts of NHK News and Nikkei to incorporate sentiment data into a predictive model for the Nikkei Stock Average. Adding sentiment data improved the R^2 score from 45.1% to a maximum of 70.5%, indicating the potential of SNS data in forecasting social indicators. However, no strong correlation between sentiment data and stock prices was observed. Challenges include the short data collection period and the difficulty of sentiment analysis in Japanese. Future work should focus on employing more effective methods for extracting sentiment.

Keywords: Sentiment Analysis, Nikkei Stock Average, Social Indicators, Natural Language Processing

1. Introduction

1.1. Research background

The rapid proliferation of social media in recent years has heightened interest in text mining using SNS. Among them, Twitter, one of the largest SNS platforms, is believed to hold substantial potential value. A study by Bollen et al. [1] utilized Twitter data to conduct sentiment analysis based on the psychological index POMS and successfully predicted the Dow Jones Industrial Average three days ahead with an accuracy of 86.7%. This suggests a correlation between Twitter data and social indicators.

1.2 Research Objective

This study aims to analyze tweets from the official Twitter accounts of NHK News and Nikkei using two sentiment analysis methods. By incorporating sentiment data into Nikkei 225 stock price data, the study seeks to predict closing prices and verify the extent to which SNS tweet data can contribute to forecasting social indicators.

2. Methodology

2.1 *Acquisition of Tweets from Twitter*

TwitterAPI [2] was employed to collect tweet data. StreamingAPI was used to retrieve real-time tweets, including tweet IDs, timestamps, and contents, and store them in a database. By executing the API, tweets from the past 15 minutes were collected, and this process was repeated every 15 minutes.

2.2 Sentiment Analysis and Time-Series Conversion

Two methods were used for sentiment analysis: MeCab+PN Table and Google Natural Language API (GNL). Both methods return sentiment scores where positive sentiment approaches 1 and negative sentiment approaches -1. The average sentiment score per day was calculated based on tweet timestamps, creating time-series data referred to as sentiment data (Fig. 2.1).

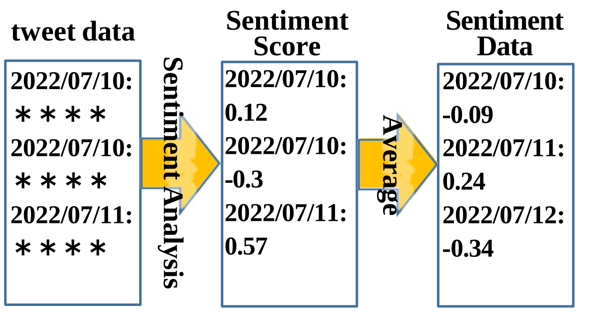


Fig. 2.1 Creation of Sentiment Data

2.2.1 Sentiment Analysis Using MeCab and PN Table

MeCab [3] was used for morphological analysis to decompose text into words. Then, corresponding sentiment scores were assigned using the PN Table [4] developed by Takamura et al. The sentiment value for each tweet was calculated by dividing the total sentiment score by the number of words.

2.2.2 Sentiment Analysis Using Google Natural Language API

The GNL [5], a text analysis tool by Google, was utilized for sentiment analysis. This API provides sentiment scores (-1 to +1) and magnitudes using a pre-trained model. The score was adopted as the evaluation metric in this study.

2.3 Building the Learning Model Using Sentiment Data

A machine learning model was developed using the LSTM algorithm, originally proposed by Hochreiter et al. [6], to learn from sentiment data and the closing prices of the Nikkei 225 index. Since the sentiment data and stock price data have different units, the data were normalized (Eq. (1)) to scale them between 0 and 1. The explanatory variables consisted of sentiment data and the closing prices of the Nikkei 225 for the past 3 days (look_back), while the target variable was the closing price of the Nikkei 225.

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \qquad (1)$$

The structure of the LSTM was based on the approach of Mhamed et al. [7]. The number of nodes, activation functions (ReLU, Softmax, Tanh), and optimization algorithms (Adam, RMSprop) were selected, and approximately 100 tuning tests were performed to finalize the parameter settings.

The performance of the model was evaluated using two key metrics: the coefficient of determination (R^2 score) and Root Mean Squared Error (RMSE).

2.4 Evaluation of the Learning Model

2.4.1 Coefficient of Determination (R^2 score)

The R^2 score [8] measures how well the model explains the data, with values closer to 1 indicating better performance. The R^2 score is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_{\text{true},i} - \overline{y_{\text{true}}}\right)^2}{\sum_{i=1}^{N}\left(y_{\text{true},i} - y_{\text{pred},i}\right)^2}$$

where $\overline{y_{\text{true}}}$ is the mean of the actual values.

2.4.2 Root Mean Squared Error (RMSE)

RMSE [9] quantifies the average prediction error, with smaller values indicating better accuracy. It is calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}\left(y_{\text{true},i} - y_{\text{pred},i}\right)^2}$$

In this study, the training data was divided into 60% for the early part of the time series and 40% for the later part, and model evaluation was performed using this split.

3. Experiment and Discussion

From July 10, 2022, to December 10, 2022, a total of 22,850 tweets from the official NHK News account (@nhk_news) and 3,688 tweets from the official Nikkei account (@nikkei) were collected using the Twitter API. Sentiment analysis was conducted on the collected tweets using MeCab+PN Table and GNL, assigning evaluation scores ranging from -1 to +1 to each tweet and creating time-series sentiment data (Fig.3.1). Sentiment data evaluated using MeCab+PN Table for NHK News is referred to as "NHK_PN," and data evaluated using GNL is referred to as "NHK_GNL." Similarly, the data for Nikkei are termed "Nikkei_PN" and "Nikkei_GNL."

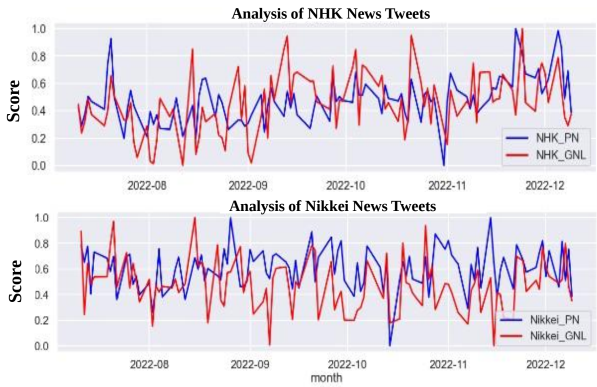


Fig. 3.1 Sentiment Data for @nhk_news and @nikkei

Table 3.1, which presents the correlation coefficients, lists the strength and direction of linear relationships between sentiment data and the Nikkei Stock Average, as well as among the sentiment datasets themselves. A correlation coefficient measures the linear association between two variables, where values range from -1 to 1. Positive values indicate a direct relationship, negative values indicate an inverse relationship, and values close to 0 indicate little to no linear relationship.

Table 3.1 Correlation of Sentiment Data with Nikkei 225 Closing Prices.

	NHK_PN	Nikkei_PN	NHK_GNL	Nikkei_GNL	Nikkei_close
NHK_PN	1.000	0.088	0.398	0.028	0.024
Nikki_PN	0.088	1.000	0.024	0.349	0.002
NHK_GNL	0.398	0.024	1.000	0.142	-0.049
Nikkei_GNL	0.028	0.349	0.142	1.000	0.191
Nikkei_close	0.024	0.002	-0.049	0.191	1.000

As shown in Table 3.1, a weak positive correlation was found within the same account, but there was almost no correlation between different accounts. The sentiment data most correlated with the Nikkei Stock Average was "Nikkei_GNL," though the correlation was very weak.

Nikkei_PN and Nikkei_GNL were combined to create 16 types of learning models. These models were labeled with indices: 0 for the Nikkei Stock Average closing price (Nikkei_close), 1 for NHK_PN, 2 for NHK_GNL, 3 for Nikkei_PN, and 4 for Nikkei_GNL. For instance, when the explanatory variable was limited to the Nikkei closing price, the model was labeled as F0. The learning models were applied to the last 40% of the dataset as validation data to predict the Nikkei closing price. The predictive performance of each model was evaluated using the R^2 score and RMSE.

Table 3.2 Results for R^2 Score and RMSE.

Explanatory variable	**R2_score**	**RMSE**
F0	**0.4514**	**0.0143**
F01	**0.5824**	**0.0139**
F02	**0.6110**	**0.0101**
F03	**0.6366**	**0.0095**
F04	**0.6460**	**0.0092**
F012	**0.6374**	**0.0094**
F013	**0.6252**	**0.0098**
F014	**0.6574**	**0.0089**
F023	**0.6370**	**0.0095**
F024	**0.6914**	**0.0080**
F034	**0.6233**	**0.0089**
F0123	**0.6127**	**0.0101**
F0124	**0.7047**	**0.0077**
F0134	**0.6229**	**0.0098**
F0234	**0.6709**	**0.0086**
F01234	**0.5954**	**0.0105**

As shown in Table 3.2, the F0 model achieved an R^2 score of only 45.1%, indicating poor performance as it fell below 50%. However, by adding sentiment data as explanatory variables, the performance improved significantly, with the F0124 model achieving an R^2 score of 70.5%. Fig. 3.2 and 3.3 illustrate the training and prediction results for model F0124.

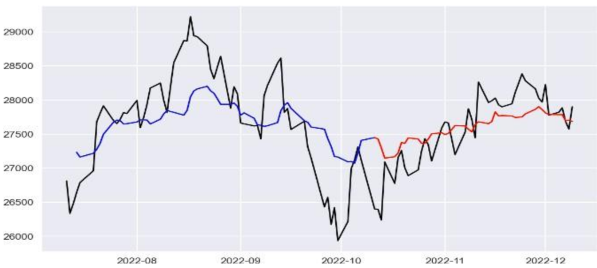


Fig. 3.2 Training and Prediction Results for F0.

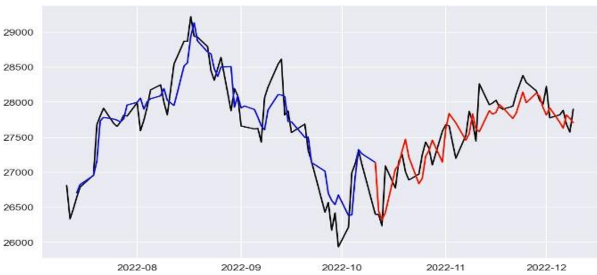


Fig. 3.3 Training and Prediction Results for F0124

In Fig. 3.2 and 3.3, the horizontal axis represents dates, while the vertical axis indicates the Nikkei closing price. The black line denotes the actual closing price, the blue line represents the predicted values from the training data, and the red line corresponds to the predicted values from the validation data. In the F0 model, the blue line fails to capture the black line adequately. In contrast, the F0124 model shows a strong resemblance between the blue and black lines, with the red line also capturing the characteristics more effectively than the F0 model.

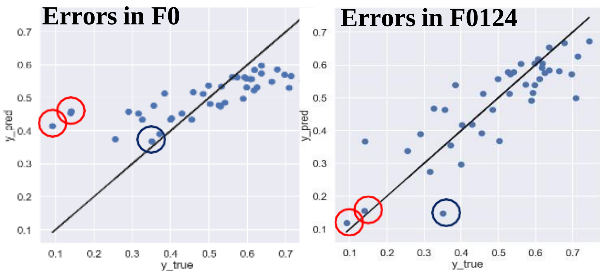


Fig. 3.4 Errors associated with the F0 and F0124 models.

Fig. 3.4 illustrates the prediction errors for the validation data. Each point represents normalized data, with the x-coordinate indicating the actual data and the y-coordinate indicating the predicted data. The black line represents y=x, and points closer to this line indicate lower errors. In the F0124 model, points are more closely aligned with the y=x line compared to the F0 model. Notably, points highlighted in the red box (2022/10/12 and 2022/10/13) represent cases where the F0124 model performed significantly better than the F0 model, while points in the blue box (2022/10/14) indicate instances of degraded performance.

Additionally, the "Crimean Bridge Explosion" on October 8, 2022, a critical event in the Ukraine War, likely influenced NHK News articles, leading to a drop in three predicted values.

4. Conclusion

This study demonstrated that incorporating sentiment analysis data from the official Twitter accounts of NHK News and Nikkei into machine learning models significantly improved the prediction accuracy of the Nikkei Stock Average. Using MeCab+PN Table and GNL for sentiment analysis, the R2 score of the F0 model (which used only the Nikkei closing price as an explanatory variable) was 45.1%, whereas the F0124 model, which included sentiment data, achieved an R2 score of 70.5%. This highlights the substantial predictive value of sentiment data.

However, unlike the findings of Bollen et al., no strong correlation between tweet data and stock prices was observed. This may be due to the nature of the collected tweets, which primarily consisted of news articles and might not adequately reflect the emotions of the Japanese populace. NHK News articles, in particular, tended to focus on events and were often biased toward negative content. Additionally, the news articles themselves were not highly emotive in their language.

Future research should explore more efficient methods for extracting emotional information from Twitter data. Extending the analysis period and increasing the data volume would also likely enable the construction of more accurate predictive models.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers JP24K0792901 and JP24K15516.

References

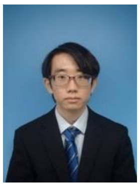
1. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 18.
2. Toriumi, F. (2015). Twitter-based big data collection and analysis (in Japanese). Organizational Science, 48(4), 47-59.
3. MeCab Official. (2022, December 6). MeCab: What is MeCab. Retrieved from https://taku910.github.io/mecab
4. Takamura, D. (2022, December 6). Word sentiment polarity correspondence table. Tokyo Institute of Technology. Retrieved from http://www.lr.pi.titech.ac.jp/~takamura/pndic_en.html
5. Google Cloud. (2022, December 6). Natural Language AI. Retrieved from https://cloud.google.com/natural-language?hl=ja
6. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.
7. Moghar, A., & Hamiche, M. (2020). Stock market prediction using LSTM recurrent neural network. Procedia Computer Science, 170, 1168-1173.
8. Kvalseth, T. O. (1985). Cautionary note about R^2. The American Statistician, 39(4), Part 1.
9. Hyndman, R. J. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679-688.

Authors Introduction

Mr. Masatoshi Beppu

He is a master student at Department of Computer Science and System Engineering, University of Miyazaki. His current research topic is an educational support using VR for children with physical disabilities.

Mr. Masatomo Ide



He is a graduate student at the Graduate School of Engineering, University of Miyazaki. His recent main research activity is to utilize various digital technologies such as web development to contribute to the local community.

Mr. Seita Nagashima

He graduated from the Faculty of Engineering at Miyazaki University in 2023 and is currently employed at MEITEC CORPORATION.

Dr. Satoshi Ikeda

He received Ph.D. from Hiroshima University and is currently an associate professor in the Faculty of Engineering at the University of Miyazaki. His research interests include time series data forecasting, graph theory, probabilistic algorithms, combinatorial optimization, and measure theory.

Dr. Amane Takei

He is working as Professor for Department of Electrical and systems Engineering, University of Miyazaki, Japan. His research interest includes high performance computing for computational electromagnetism, iterative methods for the solution of sparse linear systems, etc. Prof. Takei is a member of IEEE, an expert advisor of IEICE, a delegate of the Kyushu branch of IEEJ, a director of JSST.

Dr. Makoto Sakamoto

He is a professor in the Faculty of Engineering, University of Miyazaki. His current main research interests are computer science and information processing. In particular, he deals with automata, language theory, computation, CG, image processing, virtual technologies, data science, artificial intelligence, and so on.

Dr. Tsutomu Ito

He is Associate Professor of the Department of Business Administration at National Institute of Technology, Ube College, Japan. His current research interests include internet of things (IoT), mechanical engineering, artificial intelligence (AI), automata theory, quantitative analysis of Japanese Keiretsu. Dr. Ito earned his doctor degree of Engineering from Hiroshima University, Japan in 2018.

Dr. Takao Ito

He (Ph.D., Kyoto University) is a professor of Management of Technology, Graduate School of Engineering, Hiroshima University.

Development of a Shrine Festival Support Application with Non-Technical Management Features: Functional Evaluation and Sustainability for Future Generations

Masatomo Ide[1], **Masatoshi Beppu**[1], **Satoshi Ikeda**[1], **Kaoru Ohe**[1], **Kenji Aoki**[1], **Amane Takei**[1], **Akihiro Kudo**[2], **Makoto Sakamoto**[1*]
[1]*Graduate School of Engineering, University of Miyazaki, Japan,*
[2]*National Institute of Technology, Tomakomai Collage, Japan*
[*]*Corresponding Author*
Email: saka2000@cc.miyazaki-u.ac.jp

Abstract

This study developed a web application and video content to help raise awareness and interest in the myths surrounding the local shrine, Yakudo Shrine, and the annual festival, Chibikko Sumo Tournament, as well as to help participants establish memories of the event, and evaluate its usefulness. The application is designed to be used not only this year but also over the long term, and incorporates features that will allow easy maintenance and updating by non-technical personnel and the next generation of management members. The main functions are as follows: (1) accurate and convenient provision of detailed event information, (2) AR photo function utilizing original character illustrations to capture memorable photos, and (3) administrator-only functions to update and edit necessary information. web application can be used from publication to The web application delivered accurate and detailed information to a large number of people from the day of the festival to the day of the festival, helping to increase the number of festival participants by 2.5 times compared to the previous year. The video content, in particular, increased awareness of and interest in shrine-related myths.

Keywords: shrine, shrine-related mythology, web application, easy maintenance

1. Introduction

Yakudo Shirine(Fig. 1) is a small shrine in my hometown, but since 2017 it has had an exciting annual festival that both the "Chibikko Sumo Tournament" (Fig. 2)entrants and audience get excited about, with the goal of revitalizing the community and developing it as a tourism resource. In addition, the festival is volunteer-run by local residents, and this management structure has an important significance in enhancing local attachment and intergenerational cohesion, which will be passed on to the next generation. In anticipation of further development and an increase in the number of participants, a joint stamp

Fig. 1: Yakudo Shrine

Fig. 2: SumouTounament

rally event was held at the same time this year with the festival committee from the neighboring district.

The purpose of this study was to increase awareness and interest in this shrine and festival, as well as to promote a memorable and memorable experience for participants.

In terms of increasing awareness and interest,

・Development and release of a new official website with detailed information about the festival (contents, neighborhood information, sponsoring partners, stamp rally).

・Created video content in the form of a conversational drama that introduces the myths and gods associated with the shrine. By bringing the cultural value of myths and gods to the forefront, we expect to increase the shrine's attractiveness as a tourist resource.

In the promotion of memorable experiences,

・The AR photo shooting function using the original characters appearing in the video was implemented. With this function, the experience of the festival was recorded in the form of a photograph, with the aim of creating opportunities to recall the event at a later date.

In addition, previous studies and existing AR applications supporting local events have generally been developed by researchers[1] or companies[2], and introduced only for the year of the experiment, maintained and operated by the companies. On the other hand, since the festival to be supported emphasizes the fact that local residents are responsible for the operation of the event, we designed the application as a Web application that can be used by local residents over the long term by implementing functions that allow non-technical personnel to easily modify and update the content.

2. Support Method

2.1 Official Home Page

The first objective of the study was to increase awareness of and interest in the event by creating and publishing an official website. By posting detailed information on the website that could not be included in the flyers, the appeal of each festival and stamp rally was promoted and the interest of visitors was increased. In addition, by posting the information on the SNS[3] of the local city hall, accurate and detailed information was spread to many people, which is expected to increase awareness of the event. The list of information (Fig. 3) and a screenshot of the website are shown below. (Fig. 4, 5)

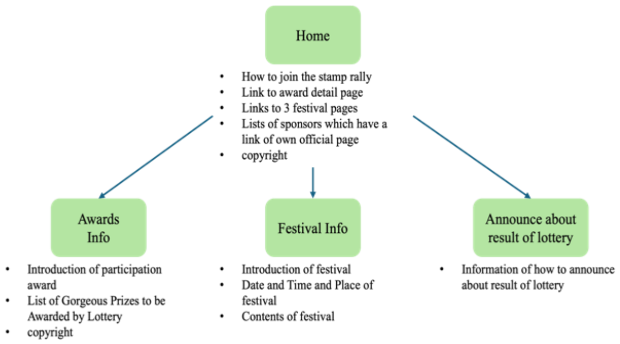


Fig. 3: component of web site

Fig. 4, Fig. 5: screenshots of website

2.2 Video Contents

The video content is structured as shown in the (Fig. 6).It introduces the mythology of the shrine, its connection to festivals, and the fact that festivals are made possible by the power of local residents. It also serves to strengthen participants' interest in the shrine and its festivals, raise awareness of the significance of little-known myths and festivals, and increase their value as a tourism resource over the long term. In addition, the content will be permanently installed at the shrine as a digital interpretation board from December 8, 2024, as shown in the (Fig. 7).

1. About Yakudo Shrine and SumouFestival
2. Awards and Women Hip Sumou
3. “Takemikaduchi” history
4. Stuff message
5. Epilogue

Fig. 6: Video Contents

Fig. 7: Video QR in Shrine

2.3 AR Photo

This function was prototyped as the second objective of the research, to promote content that would encourage participants to fix their memories and memories of the event. The function of displaying and photographing characters without the need for markers, which improves on the limitations of Web AR, has been realized, but the degree of freedom in character placement is low, and there is still room for improvement, so other approaches need to be considered. However, the degree of freedom in character placement is low, and there is still room for improvement, so other approaches need to be considered.

Fig.8: Sample of ARPhoto

2.4 Editing Function (useable only stuff)

Fig. 9 shows an overview of the editing function. With this function, the aim is to realize continuous page management by local residents who run the festival. The editable information was targeted at text and images whose content may change from year to year. Specifically, this includes the introduction and contents of festivals and stamp rallies, date and time information, and sponsor information. In addition, an authentication function using Firebase [4] was implemented to prevent people without appropriate account information from accessing the editing function, thereby preventing unexpected tampering.

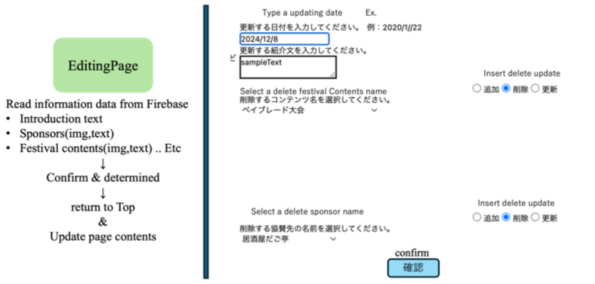


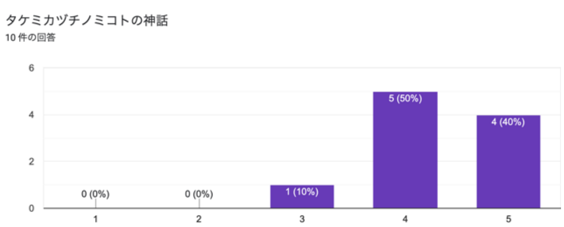
Fig. 9: Editing Function

3. Result

The number of participants in the stamp rally and sumo tournament (Table 1) suggests that the official website made a significant contribution to the research objectives of increasing awareness and communicating accurate and detailed information. The editing functions, from entering numbers to reflecting updated information, functioned well. It is also expected to speed up the response to urgent data updates, such as the change of sponsors' official SNS links, which was an actual case in FY2024.

However, the AR Photo did not perform well in promoting participants' memories and memory retention; as discussed in section 2.3, the low degree of freedom in character placement was a major factor in lowering the level of satisfaction with the AR Photo.

Although the number of viewers of the video content is still small (45 at this time), the responses to the questionnaire (Graph 1) indicate that the content was highly rated, especially the value of myths associated with the shrine. Therefore, the potential to help create value as a tourism resource can be expected. Further efforts should be made to increase the number of viewers in the future.



Graph 1: Degree of increased interest in myth

Table 1: Number of festival participants

	SumouTournament participants	Stamp rally applicants
2023	61	/
2024	153	183

4. Future Works

For the official website, we will introduce analysis services such as Google Analytics [5] to collect detailed data on the number of accesses, pages viewed, etc., to create better pages. We will also expand the contents that can be edited.

For AR photos, we would like to verify and improve multiple improvements in the form of A/B testing, such as providing hybrid web and apps and remaking them using a different approach.

For video content, we will devise ways to increase the number of viewers by using animation and abundant materials, and by using production techniques and cut-outs.

Acknowledgements

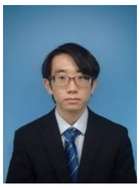
This study was conducted in cooperation with the Chibikko Sumo Tournament Executive Committee, the Sanzai Navel Festival Executive Committee, and the Tonokoori Joshi Festival Executive Committee.

References

1. Ai Funayama, Akari Yoshuda, Makoto Nakazawa https://www.jc.u-aizu.ac.jp/department/management/youshi/2013/06.pdf
2. https://momijihachimangu.or.jp/goshuinar/
3. https://www.facebook.com/100064358814975/photos/954793113342640/?_rdr
4. https://firebase.google.com/?hl=ja
5. https://developers.google.com/analytics?hl=ja

Authors Introduction

Mr. Masatomo Ide



He is a graduate student at the Graduate School of Engineering, University of Miyazaki.
His recent main research activity is to utilize various digital technologies such as web development to contribute to the local community.

Mr. Masatoshi Beppu

He is a master student at Department of Computer Science and System Engineering, University of Miyazaki. His current research topic is an educational support using VR for children with physical disabilities.

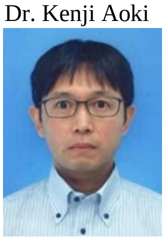
Dr. Satoshi Ikeda

He received PhD degree from Hiroshima University. He is an associate professor in the Faculty of Engineering, University of Miyazaki. His research interest includes graph theory, probabilistic algorithm, fractal geometry and measure theory.

Prof. Kaoru Ohe

She received her Ph.D. degrees from University of Miyazaki, Japan, in 2014. Currently she is an Associate Professor of the Center for Science and Engineering Education, Faculty of Engineering. Her research is separation engineering especially adsorption hazardous heavy metals and oxyanions

Dr. Kenji Aoki



He received Ph.D. of Engineering from Kagoshima University in 2010. He is currently working in Information Technology Center at University of Miyazaki as Associate Professor, since 2010. His research interests include bio-informatics, evolutionary computation, information system and Intelligent systems. He is a member of IPSJ and JSET.

Dr. Amane Takei

He is working as a professor for Department of Electrical and systems Engineering, University of Miyazaki. His research interest includes high performance computing for computational electromagnetism, iterative methods for the solution of sparse linear systems, domain decomposition methods for large-scale problems. Prof. Takei is a member of IEEE, an expert advisor of IEIC, a delegate of Kyushu branch of IEEJ, a director of JSST.

Prof. Akihiro Kudo

He received Ph.D. degree from Nagaoka University of Technology He is a professor in the Department of engineering for innovation , National Institute of Technology, Tomakomai college. He is a member of Acoustical society of Japan, Information and Communication Engineers (IEICE).

Prof. Makoto Sakamoto

He is presently a professor in the Faculty of Engineering, University of Miyazaki. His first interests lay in hydrodynamics and time series analysis, especially the directional wave spectrum. He is a theoretical computer scientist, and his current main research interests are automata theory, languages and computation. He is also interested in digital geometry, digital image processing, computer vision, computer graphics, etc.

A Novel Approach to Reducing Ranking Discrepancies in Tennis Based on Tournament Choices

Masatomo Ide[1], Masatoshi Beppu[1], Kousei Yano[2], Satoshi Ikeda[1]*, Amane Takei[1], Makoto Sakamoto[1], Tsutomu Ito[3], Takao Ito[4]

([1]Graduate School of Engineering, University of Miyazaki, Japan),
([2] GENBASUPPORT Co., Japan),
([3]National Institute of Technology, Ube College, Japan),
([4]Hiroshima University, Japan)

[*]Corresponding Author
E-mail: bisu@cs.miyazaki-u.ac.jp

Abstract

In tennis, discrepancies between rankings and head-to-head results arise because players select different tournaments for ranking. Existing methods using directed graphs cannot address discrepancies caused by varying tournament choices. This study proposes a ranking aggregation method that considers players' selected tournaments to reduce these inconsistencies. The method aggregates all chosen tournaments to form a collective ranking. Experimental results show a reduction in overall discrepancies, although some players saw an increase. This approach offers a partial solution to ranking inconsistencies caused by differing tournament selections in tennis.

Keywords: Tennis Rankings, Ranking Discrepancies, Tournament Selection

1. Introduction

1.1 Research background

In sports, consider two players, A and B, and the creation of a ranking that includes them. Despite A being ranked higher than B, there may be situations where B achieves better results than A. Such a situation is referred to as a "contradiction" in this study. In sports like baseball and basketball, the relative strengths (rankings) of players can be represented using directed graphs, and the contradictions with match results can be counted. Rankings that minimize the total number of contradictions can then be created [1]. However, in tennis, players participate in different tournaments, leading to situations where one context recognizes a contradiction, while another does not. In other words, existing formulations cannot address cases where contradictions are perceived differently between two players.

1.2 Objective of the study

The objective of this research is to develop a generalized ranking system that addresses discrepancies between competition outcomes and rankings, using tennis as a case study. While acknowledging the inherent challenges in perfectly aligning results and rankings, the proposed system aims to provide a more consistent and fair evaluation framework applicable across various sports or competitive domains.

2. Formulation

2.1 ATP Rankings

The ATP ranking is the global ranking system for tennis. Rankings are determined based on the total points from the top 19 tournaments in which a player earned the most points within the past 52 weeks [2]. The ATP determines points based on tournament categories and placements.

2.2 Proposed Formulation

In ATP rankings, players participate in different tournaments. In this study, we propose creating a new ranking by aggregating match results from all tournaments played by the top 10 ranked players. Let $\mathrm{N} = \{\mathbf{1},\mathbf{2},\cdots,n\}$ represent the set of players, and $T = \{t_1, t_2, \cdots, t_m\}$ the set of tournaments. For a player i, the tournaments they select (number of tournaments I) are $T_i = \{t_i^1, t_i^2, \cdots, t_i^I\}$. Match tables are created for each tournament selected by player i. The match table is an $n\times n$ antisymmetric matrix, where the performance of element i against element j is represented by w_{ij} determined as follows:

$$w_{ij} = \begin{cases} \mathbf{1} & win\ for\ i, \\ \mathbf{0} & drow, \\ \mathbf{-1} & loss\ for\ i. \end{cases}$$

Here, a "win" means achieving better results than the opponent in a tournament, even without a direct match (e.g., the tournament champion is considered to have "won" against the players who placed second, third, or fourth based on their relative rankings). Conversely, worse results indicate a "loss," and identical results mean a "draw." To prevent players participating in more tournaments from having an advantage, tournaments not selected during ATP ranking calculation are treated as losses. If two players did not participate in the same tournament, it is treated as a draw.

Under these conditions, match tables are created for all tournaments selected by the top 10 players.

These match tables are then aggregated. The aggregation method proposed in this study is based on the method devised by Yoshitsugu Yamamoto et al. [3].

The similarity constant c_{ij} between rankings when elements i and j exist is defined as follows:

$$c_{ij} = i \text{ higher} - j \text{ higher}$$

(Net tournaments where i ranks higher than j)

Using the matrix C, whose elements are c_{ij}, a comprehensive ranking is created by maximizing the consistency between rankings. The decision variable x_{ij} is defined as follows:

$$x_{ij} = \begin{cases} 1 & i \textit{ is ranked higher than } j, \\ 0 & \textit{otherwise}. \end{cases}$$

The objective function and constraints are:

$$\begin{array}{|ll} \text{Maximize} & \sum_{i=0}^{n} \sum_{j=0}^{n} c_{ij} x_{ij} \\ \text{Subject to} & x_{ij} + x_{ji} = \mathbf{1}, x_{ij} + x_{jk} + x_{ki} \le 2, \\ & x_{ij} \in \{0,1\}. \end{array}$$

Once the X-matrix is determined, the sum of each row in the X-matrix is sorted in descending order to determine the final comprehensive ranking.

3. Experiments

3.1 Data Used

Table 3.1 lists the top 10 players in the ATP rankings and their points as of December 4, 2023 [2].

Table 3.1 ATP Rankings

Ranking	**Player**	**Points**
1	**Djokovic**	**11,245**
2	**Alcaraz**	**8,855**
3	**Medvedev**	**7,600**
4	**Sinner**	**6,490**
5	**Rublev**	**4,805**
6	**Tsitsipas**	**4,235**
7	**Zverev**	**3,985**
8	**Holger Rune**	**3,660**
9	**Hurkacz**	**3.245**
10	**Fritz**	**3,100**

3.2 Experimental Method

A match-up table was created for the tournaments used to determine the ATP rankings of the 10 players in Table 3.1. The total match-up data was aggregated to compute the constants c_{ij}. Based on the resulting C-matrix, the decision variable x_{ij} was determined using Gurobi [4]. By summing each row of the X-matrix and sorting the values in descending order, the final rankings were created.

3.3 Experimental Results

The C-matrix obtained by summing the match-up tables

$$C = \begin{pmatrix} 0 & 33 & 5 & 10 & 11 & 27 & 43 & 11 & 42 & 39 \\ -33 & 0 & -12 & -24 & 22 & 42 & 9 & 47 & 42 & 77 \\ -5 & 12 & 0 & -14 & 34 & 48 & 62 & 64 & 60 & 64 \\ -10 & 24 & 14 & 0 & -19 & 35 & 59 & 33 & 45 & 59 \\ -11 & -22 & -34 & 19 & 0 & 6 & 52 & 37 & 78 & 42 \\ -27 & -42 & -48 & -35 & -6 & 0 & 39 & 10 & 32 & 16 \\ -43 & -9 & -62 & -59 & -52 & -39 & 0 & -24 & -1 & 10 \\ -11 & -47 & -64 & -33 & -37 & -10 & 24 & 0 & 17 & 21 \\ -42 & -42 & -60 & -45 & -78 & -32 & 1 & -17 & 0 & 13 \\ -39 & -77 & -64 & -59 & -42 & -16 & -10 & -21 & -13 & 0 \end{pmatrix}$$

is shown below:

The optimization results obtained using Gurobi for the C-matrix are as follows:

$$X = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

By summing each row of the X-matrix and arranging them in descending order, the rankings were determined, as shown in Table 3.2:

Table 3.2 Rankings Based on the Proposed Method

Ranking	**Player**
1	**Djokovic**
2	**Sinner**
3	**Alcaraz**
4	**Medvedev**
5	**Rublev**
6	**Holger Rune**
7	**Tsitsipas**
8	**Hurkacz**
9	**Zverev**
10	**Fritz**

The ranking changes between the ATP rankings and those derived from the proposed method are illustrated in Fig. 3.1.

In Figure 3.1, the vertical axis represents the ranks, while the horizontal axis denotes the ranking methodology. The figure shows that, compared to the ATP rankings, Sinner, Holger Rune, and Hurkacz improved their rankings, while Alcaraz and Zverev dropped in rank. Djokovic, Medvedev, Rublev, Tsitsipas, and Fritz retained their positions.

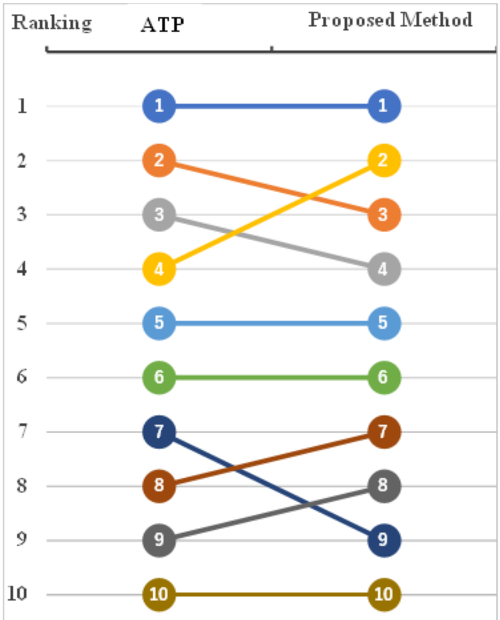


Fig.3.1 Ranking Changes Based on the Methodology

3.4 Evaluation

The ATP rankings and the rankings generated by the proposed method were evaluated based on the "number of ranking discrepancies relative to match results" from the perspective of each player. The match results for each tournament were calculated using w_{ij} as defined in Section 2.2. The ranking discrepancies were classified into the following four patterns:

Pattern A: Discrepancies are counted from both the player's and the opponent's perspectives.
Pattern B: Discrepancies are counted from the player's perspective but not from the opponent's perspective.
Pattern C: Discrepancies are not counted from the player's perspective but are counted from the opponent's perspective.
Pattern D: No discrepancies are counted from either perspective.

The number of discrepancies between match results and ranking positions in tournaments selected by each player is shown in Tables 3.3 and 3.4.

Table 3.3 Ranking Discrepancies in ATP Rankings (from each player's perspective)

Player	Pattern A	Pattern B	Pattern C	Pattern D	Total Discrepancies
Djokovic	13	0	0	95	13
Alcaraz	23	9	0	121	32
Medvedev	38	17	0	125	55
Sinner	37	23	0	111	60
Rublev	29	28	0	114	57
Tsitsipas	29	31	0	102	60
Zverev	24	38	0	118	62
Rune	32	52	0	96	84
Hurkacz	27	47	0	97	74
Fritz	26	67	0	69	93
Total	278	312	0	1048	590

Table 3.4 Ranking Discrepancies in Proposed Method (from each player's perspective)

Player	Pattern A	Pattern B	Pattern C	Pattern D	Total Discrepancies
Djokovic	13	0	0	95	13
Alcaraz	24	19	0	110	43
Medvedev	37	14	0	129	51
Sinner	33	12	0	126	45
Rublev	29	28	0	114	57
Tsitsipas	29	31	0	102	60
Zverev	22	54	0	104	76
Rune	30	44	0	106	74
Hurkacz	27	42	0	102	69
Fritz	26	67	0	69	93
Total	270	311	0	1057	581

From Tables 3.3 and 3.4, the number of discrepancies increased for Alcaraz and Zverev, whose rankings dropped. In contrast, the discrepancies decreased for Medvedev, Sinner, Rune, and Hurkacz, whose rankings improved. The increase in discrepancies for Alcaraz and Zverev is likely due to a tendency for lower-ranked players to exhibit more discrepancies. Furthermore, both the total of Pattern A and Pattern B discrepancies and the overall number of discrepancies decreased.

3.5 Discussion

Comparing the ATP rankings and the rankings generated by the proposed method, positional changes are observed between ranks 2-4 and ranks 7-9. Sinner's rise in ranking can be attributed to his head-to-head winning record against Alcaraz and Medvedev. Additionally, Alcaraz participated in fewer tournaments than Medvedev and Sinner, leading to a drop in his ranking.

For ranks 7-9, the aggregated C matrix of head-to-head match results indicates that Zverev has negative scores against both Rune and Hurkacz. This means Zverev lost more matches than he won against these players, contributing to his lower ranking.

Moreover, while the original rankings lacked a clear definition for the matchup matrix, the formulation proposed in this study treats tournaments not selected as "unattended." This definition eliminates situations where both players simultaneously win or lose, thus removing discrepancies counted only from one perspective (Pattern C). Consequently, the number of discrepancies due to differing perspectives was reduced, addressing a key issue in this study.

4. Conclusion

4.1 Summary

This study proposed a ranking method to reduce "discrepancies between rankings and match results" in sports where the tournaments considered in the rankings differ for each player. As a result, the integration of all tournaments selected by the 10 players partially resolved the issue of tournament selection differences. Evaluation showed that while some players experienced an increase in discrepancies, the overall number of discrepancies decreased.

4.2 Future Work

The proposed formulation resulted in lower rankings for players with fewer selected tournaments (i.e., tournaments used to generate ATP rankings). Future work should develop a formulation that prevents players with fewer selected matches from being disadvantaged.

Acknowledgements

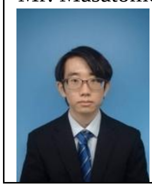
This work was supported by JSPS KAKENHI Grant Numbers JP24K0792901 and JP24K15516.

References

1. Iqbal Ali, Wade D.Cook and Moshe Kress, “On the Minimum Violations Ranking of a Tournament”,Management Science, 32,pp.660-672,1986.
2. ATP 公式サイト https://www.atptour.com/en/
3. Amy N. Langville, Kathryn Pedings, and Yoshitsugu Yamamoto, “A minimum violation ranking method”, Optimization and Engineering, 13, 349-370, 2011.
4. Gurobi 公式サイト https://www.gurobi.com/

Authors Introduction

Mr. Masatomo Ide



He is a graduate student at the Graduate School of Engineering, University of Miyazaki. His recent main research activity is to utilize various digital technologies such as web development to contribute to the local community.

Mr. Masatoshi Beppu

He is a master student at Department of Computer Science and System Engineering, University of Miyazaki. His current research topic is an educational support using VR for children with physical disabilities.

Mr. Kousei Yano

After graduating from the Faculty of Engineering at Miyazaki University in 2024, he is currently working at GENBASUPPORT Co., Japan.

Dr. Satoshi Ikeda

He received Ph.D. from Hiroshima University and is currently an associate professor in the Faculty of Engineering at the University of Miyazaki. His research interests include time series data forecasting, graph theory, probabilistic algorithms, combinatorial optimization, and measure theory.

Dr. Amane Takei

He is working as Professor for Department of Electrical and systems Engineering, University of Miyazaki, Japan. His research interest includes high performance computing for computational electromagnetism, iterative methods for the solution of sparse linear systems, etc. Prof. Takei is a member of IEEE, an expert advisor of IEICE, a delegate of the Kyushu branch of IEEJ, a director of JSST.

Dr. Makoto Sakamoto

He is a professor in the Faculty of Engineering, University of Miyazaki. His current main research interests are computer science and information processing. In particular, he deals with automata, language theory, computation, CG, image processing, virtual technologies, data science, artificial intelligence, and so on.

Dr. Tsutomu Ito

He is Associate Professor of the Department of Business Administration at National Institute of Technology, Ube College, Japan. His current research interests include internet of things (IoT), mechanical engineering, artificial intelligence (AI), automata theory, quantitative analysis of Japanese Keiretsu. Dr. Ito earned his doctor degree of Engineering from Hiroshima University, Japan in 2018.

Dr. Takao Ito

He (Ph.D., Kyoto University) is a professor of Management of Technology, Graduate School of Engineering, Hiroshima University.

Sound field evaluation on acoustical experiment with several loudspeaker locations

Akihiro Kudo, Shun Kubota
National Institute of Technology, Tomakomai College, Japan

Amane Takei, Makoto Sakamoto
Graduate School of Engineering, University of Miyazaki, Japan
Email: kudo@tomakomai-ct.ac.jp

Abstract

When a laboratory at a technical college is used as a site for psychoacoustic experiments, there are objects in the experimental environment such as walls, floors, ceilings, and chairs that obstruct the propagation of sound waves, which may change the accuracy of sound direction localization by altering the sound waves reaching the listener's ears. Therefore, the purpose of this study was to clarify the effects of these environments on sound localization characteristics using simulations. In this paper, we evaluated the effect of the reflection of sound waves on a chair to keep posture of the subject on the sound image localization for several loudspeaker positions. Results showed that it isn’t strongly possible that the early reflections due to the chair will clearly shift the direction of the sound image.

Keywords: large-scale simulation, acoustical sound field, non-steady state analysis, interaural time difference

1. Introduction

When a laboratory at a technical college is used as a site for psychoacoustic experiments, there are objects in the experimental environment such as walls, floors, ceilings, and chairs that obstruct the propagation of sound waves, which may change the accuracy of sound direction localization by altering the sound waves reaching the listener's ears. Therefore, the purpose of this study is to clarify the effects of these environments on sound localization characteristics using simulations. We determined the parameter values of the non-stationary analysis method of the sound field needed for the above evaluation at last year's ICAROB [1]. In this paper, we evaluated the effect of the reflection of sound waves by a chair set up to immobilize the subject on the subject's sound image localization for loudspeaker positions placed every 45 degrees in the horizontal plane.

2. Adventure Sound

2.1. Brief introduction

ADVENTURE_Sound is an open-source software for sound simulation. This allows us to simulate parallel finite element analysis software for the sound fields in huge space. Analyses are performed by solving the large-scale linear system with parallel computing based on the Mandel’s domain decomposition method (DDM) [2].

2.2. Simulation procedures with ADVENTURE_Sound

The procedure to use ADVENTURE_Sound is as follows; 1) generate the mesh data, 2) add boundary conditions (BCs), 3) subdivide the mesh data into subdomains, 4) simulate the sound field distributions, 5) convert to the results to vtk data for visualization. On the fourth procedure, it needs to calculate Helmholtz equation in the time domain. For this, Newmark β method is adopted to the equation. As the boundary conditions, input wave function and first and second derivatives of that function.

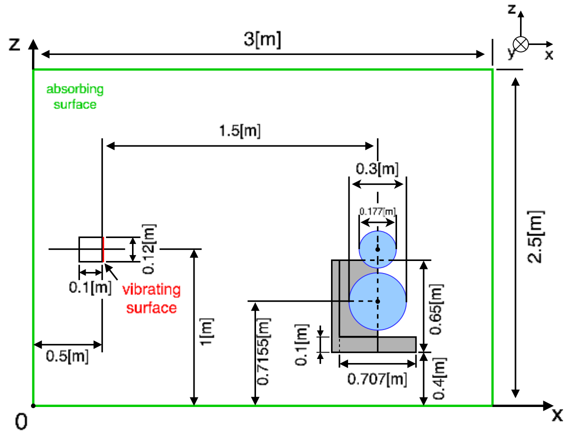
3. Simulation on a typical psychoacoustic experimental environment

3.1. simulation model

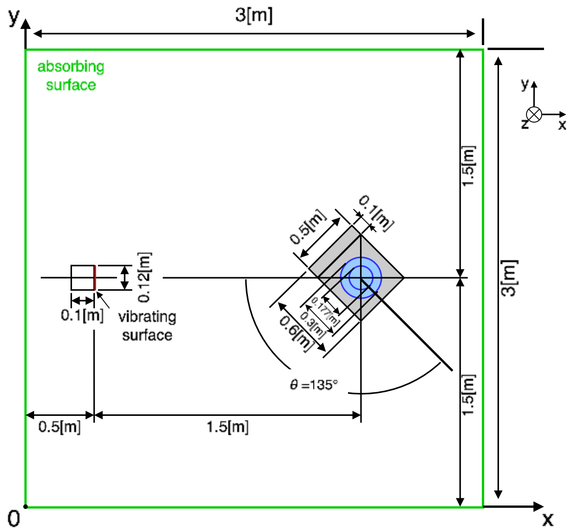
Fig. 1 shows that a simulation model that includes a loudspeaker-listener pair with a chair that holds listener’s body. This is designed for typical psychoacoustical experiment. The shape of the listener is a snowman consisted of two spheres, their diameters approximating the average head and body [3]. The loudspeaker is located at 135° which is behind it as seen by the listener.

The loudspeaker is assumed to be a small-diameter, full-range product often used in sound image localization experiments. This model is used to evaluate the time waveform caused by sound waves emitted from a loudspeaker and reflected or diffracted by a chair.

In this model, the acoustic impedance of the room walls, ceiling, and floor is set to a value of 445.9kg/m^{2}s to ensure that sound waves are completely absorbed. On the other hand, the acoustic impedance of the surface of the loudspeaker and the chair is set to 4.56×10^9 kg/m^{2}s to ensure that sound waves are completely reflected.



(b) side view



(a) Bird’s view

Figure 1. simulation model of psychoacoustical

Table 1 shows that the dimensions of the simulation model. The distance between the loudspeaker and the subject is 1.5 m, which is known as the transitional distance from spherical waves to plane waves in the audible bandwidth.

Table 1. dimension of the simulation model

Room, width[m] ×depth[m]×height[m]	3×3× 2.5
Room volume [m^3]	22.5
Loudspeaker, width[m] ×depth[m] ×height[m]	0. 1 ×0. 12 × 0.12
Diameter of the listener head, torso [m]	0.177, 0.3
Distance from loudspeaker to listener [m]	1.5
Backrest, width[m] ×depth[m] ×height[m]	0. 5 ×0. 1 × 0.55
Seat, width[m] ×depth[m] ×height[m]	0. 5 ×0. 5 × 0.1

3.2. Specs of the parallel computer

Our parallel computer is called as TOMAHAWK. Table 2 shows that the detailed specs of the parallel computer.

Table 2. Specs of parallel computer

CPU	Intel Xeon E5-2650L,1.8GHz with 16 cores L2 20480[KB]
Main memory/node [GB]	32
Number of nodes	7
Total cores	112
OS	CentOS 6.2
Compiler	gcc-4.4.7, mipch-3.2

3.3. Simulation conditions

Table 3 shows that analysis conditions. These parameter values come from previous our study [4].
Because of the mesh size is 0.025m and the time resolution is 1 µs and the Courant number equals to 0.0238, therefore the Courant condition is satisfied.

Table 3. Analysis conditions for the test model

Mesh size[m]	0.025
Number of elements	16,264,673
Acoustic impedance of the Room wall, floor, ceiling [kg/m^{2}s]	445.9
Acoustic impedance of the Loudspeaker, chair, listener	4.56×10^9
Medium(air) r_0[kg/m^3]	1.3
Speed of sound c_0[m/s]	343
Time resolution Δt [µs]	1.0
Duration time to simulate [µs]	8000
Courant number$\sqrt{3}$ ×c× $\Delta t/\Delta x$	0.0238≪ 1
Velocity potential $\{\Phi\}_t$ Full width at half maximum[ms] Cut-off frequency f_c[Hz]	Gaussian pulse 0.264 646

For boundary conditions as input wave, the single wave is useful to analysis the transient sound field, because sound reflections are observed clearly. In this study, a gaussian pulse as velocity potential is adopted as follows,

$$\{\Phi\}_t = e^{-\{\alpha(t-t_0)\}^2} \quad \dots \quad (1)$$

In the Eq(1), decay parameter α is set to 3141, delay parameter t_0 is set to 0.001[s].

3.4. Simulation results

Fig. 2 shows the simulation results for every 1000 microseconds on xz plane at y=1.5m. In this figure, the red color indicates a high intensity of velocity potential proportional to sound pressure. It can be seen that sound waves are emitted from the loudspeaker, and as time passes, the sound waves travel toward the listener, and at approximately 5049 µs, the direct waves reach the listener's head. After 6049µs, the sound waves are seen to be distributed in a complex manner due to reflections from the chair.

Fig. 3 shows the simulation results on xy plane at z=1m. In this figure, the direct waves emitted from the loudspeaker arrive at the backrest of chair between 3049µs and 4049µs. At t=6049µs, the first reflected waves occur around the listener.

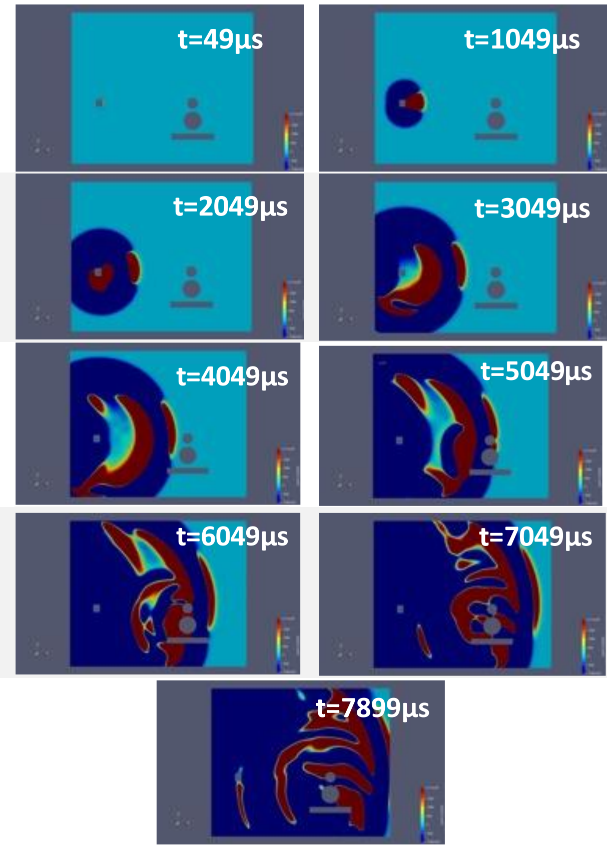


Figure 2. Visualization of the sound field at several times on xz plane at y=1.5m

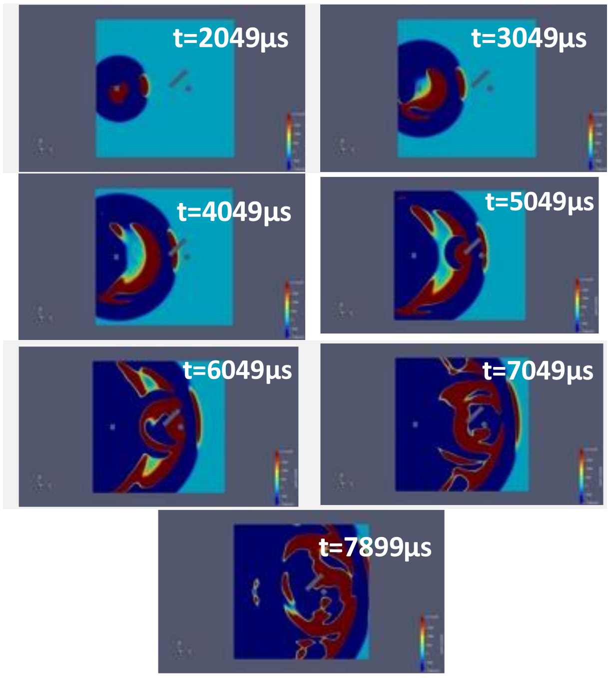


Figure 3. Visualization of the sound field at several times on the xy plane at z=1m

3.5. Discussions

Fig. 4 shows the simulation results on the xz plane at y7=1.5m when the sound wave arrived at the listener.

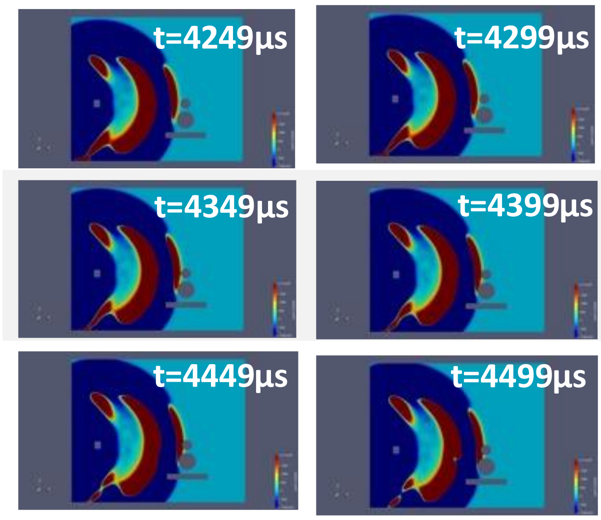
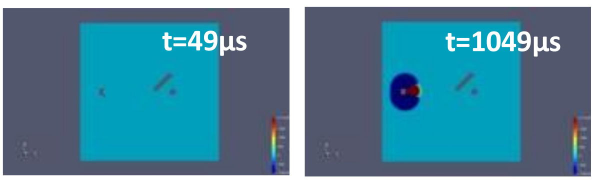


Figure 4. Visualization of the sound field when the sound waves arrived at listener on the xy plane at y=1.5m

The timing obtained from this figure, is approximately 4399µs. Because the sound velocity was 343m/s and the elapsed time was 4399 µs, the distance between the loudspeaker and the listener is estimated at 1.50m. This is consistent with the distance set in the simulation model, indicating the validity of this simulation.

Fig. 5 shows that the simulation results on the xy plane at z=1m when the sound waves arrived at the listener. The



time at which the sound waves reach the listener's right ears and left ears is 4499μs and 4799μs, respectively.

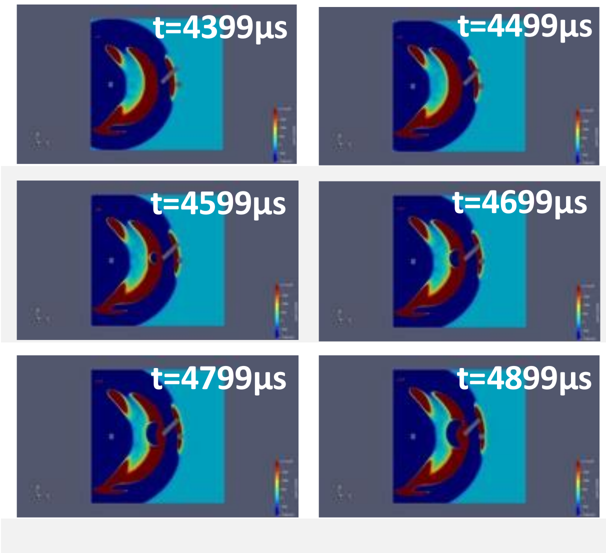


Figure 5. Sound field when the first reflection waves arrived at both ears on the xy plane at z=1m

Fig. 6 shows that the first reflections (called as early reflections) arrived at both ears of listener. The time at which the reflected wave reaches the listener's two ears is 5499μs and 5999μs, respectively.

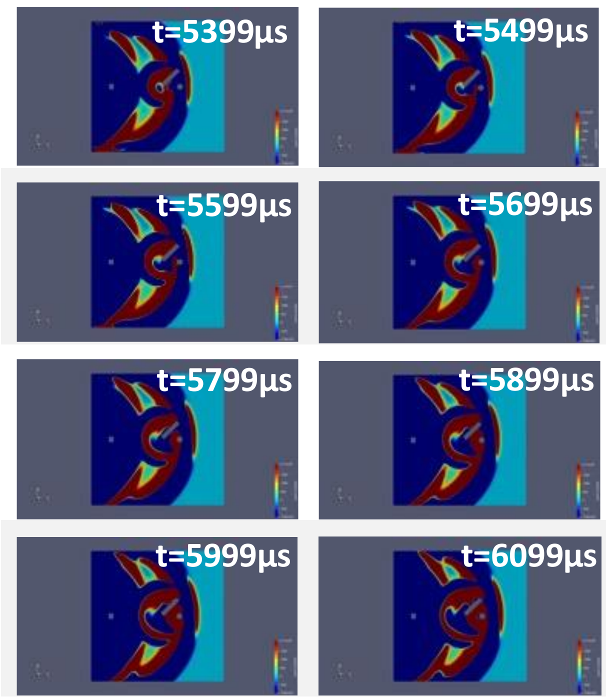


Figure 6. Sound field when the first reflection waves arrived at both ears on the xy plane at z=1m

The time profile under different loudspeaker locations is shown in fig. 7. In this figure, (b) is the results from previous study [4].

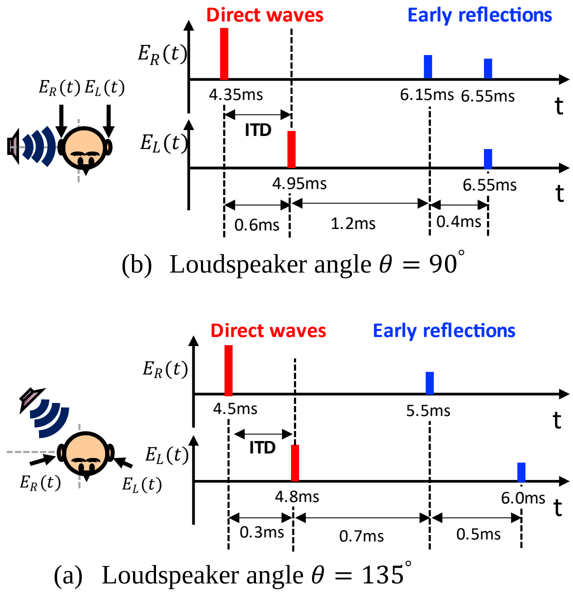


Figure 7. Time profile of sound waves at both ears at different loudspeaker locations

At $\theta = 135°$, the time difference between the arrival of sound waves at the two ears is known as ITD. The ITD was estimated to be approximately 0.3ms. The time difference between right direct waves and left early reflections is 1.5ms. Another time difference between left direct waves and right early reflections is 0.7ms. Because these differences are greater than the maximum of theoretical ITDs whose value is 0.67ms obtained from $\Delta s/c = d(\theta + \sin\theta)/2\,c$ [5] where the Δs is the path difference when sound waves reach both ears, c is the sound velocity, d is the diameter of listener's head and θ is the azimuth angle with respect to the front of the listener.

From fig. 7(b), at $\theta = 90°$, also, both time differences between the direct waves and the early reflections on the facing side exceed the maximum of ITD.

Therefore, it isn't strongly possible that the early reflections will clearly shift the direction of the sound image, since these time differences are slightly greater than the maximum ITD.

4. Conclusion

Purpose of this paper is to clarify the effect of these environments on the human auditory characteristics, especially sound image localization ability. The simulation target is loudspeaker and snowman model with chair, was adopted as a typical psychoacoustical experimental environment.

The simulation results showed that 1) The arrival times of sound waves at both ears when emitted from a loudspeaker placed at 135° behind the listener were 4.5ms and 4.8ms, respectively, 2) The arrival times of the early reflections at both ears were 5.5ms and 6.0ms, respectively, 3) The time differences from the direct wave in the right ear to the early reflections in the left ear and from the direct wave in the left ear to the early reflections in the right ear, were 1.5ms and 0.5ms, respectively.

Our conclusion is that it isn't strongly possible that the early reflections will clearly shift the direction of the sound image as perceived by the listener, because these time differences between the direct waves and early reflections slightly exceeds the range in the time domain that affects sound image localization.

Acknowledgements

This research was supported by Grant-in-Aid for Scientific Research 24K15516.

References

1. A.Kudo, A. Takei, M.Sakamoto, "Sound field evaluation on acoustical experiment using non-steady state analysis", ICAROB2024, OS22-3.
2. J. Mandel, "Balancing domain decomposition", Communications on Numerical Methods in Engineering, Vol.9, pp.233-241, 1993.
3. S. Shimada, K. Sugiyama, H. Hokari, Wave equation and Head related transfer functions model (in Japanese, Kameda book service, 2011).
4. A. Kudo, S. Kubota and A. Takei, "Evaluation of ITDs in a loudspeaker-snowman model based on non-stationary sound field analysis using ADVENTURE_Sound", JSST2024, OS9-3.
5. B. C. J. Moore, An introduction to the Psychology of Hearing (ELSEVIER ACEDEMIC PRESS, 2007).

Authors Introduction

Prof. Akihiro Kudo

He received ph.D. degree from Nagaoka University of Technology He is a professor in the Department of engineering for innovation, National Institute of Technology, Tomakomai college. He is a member of Acoustical society of Japan, Information and Communication Engineers (IEICE).

Mr. Shun Kubota

He is a student in the Department of Engineering for innovation at Tomakomai National College of Technology. He is engaged in research on large-scale acoustic analysis using parallel computers.

Prof. Amane Takei

He is working as Professor for Department of Electrical and systems Engineering, University of Miyazaki, Japan. His research interest includes high performance computing for computational electromagnetism, iterative methods for the solution of sparse linear systems, domain decomposition methods for large-scale problems. Prof. Takei is a member of IEEE, an expert advisor of The Institute of Electronics, Information and Communication Engineers (IEICE), a delegate of the Kyushu branch of Institute of Electrical Engineers of Japan (IEEJ), a director of Japan Society for Simulation Technology (JSST).

Prof. Makoto Sakamoto

He is presently a professor in the Faculty of Engineering, University of Miyazaki. His first interests lay in hydrodynamics and time series analysis, especially the directional wave spectrum. He is a theoretical computer scientist, and his current main research interests are automata theory, languages and computation. He is also interested in digital geometry, digital image processing, computer vision, computer graphics, virtual reality, augmented reality, entertainment computing, complex systems, artificial intelligence, data science, and so forth.

Parallel high-frequency electromagnetic field analysis based on hierarchical domain decomposition method

Amane Takei*
Faculty of Engineering, University of Miyazaki, Japan

Akihiro Kudo
National Institute of Technology, Tomakomai college, Japan

Makoto Sakamoto
Faculty of Engineering, University of Miyazaki, Japan

＊Corresponding Author

E-mail:takei@cc.miyazaki-u.ac.jp
http://www.miyazaki-u.ac.jp/

Abstract

In this presentation, a parallel high-frequency electromagnetic field analysis code based on an iterative domain decomposition method is explained that is named ADVENTURE_FullWave. A stationary vector wave equation for the high-frequency electromagnetic field analyses is solved taking an electric field as an unknown function. Then, to solve subdomain problems by the direct method, the direct method based on the LDL^T decomposition method is introduced in subdomains. The simplified Berenger's PML is introduced which these eight corners are given the average value of all PML's layers. And, we show a numerical example of a microwave. More detail will be shown in the conference.

Keywords: Electromagnetic field analysis, Finite element method, Domain decomposition method, Huge-scale analysis.

1. Introduction

Electromagnetic field analysis based on a numerical analysis method, such as the finite element method, has become widespread [1] due to recent improvements in computer performance and numerical calculation technology. In the case of accurately reproducing an analysis model of complicated shape, it is necessary to use many small the elements. In the case of analyzing the state of electromagnetic waves propagation in a wide range, a wide analysis domain is examined. Furthermore, to perform a high-accuracy analysis, it is necessary to model the analysis domain with a sufficiently small element for the wave-length, and, in this case, the number of elements also increases. Increasing the number of elements increases the scale of the problem. Therefore, a method that can calculate large-scale problems has come to be demanded. Moreover, large-scale problems must be solved with high accuracy. In the presentation, a large-scale analysis code: ADVENTURE_FullWave is introduced, and detail of the parallel algorism is shown.

2. Governing equations and algorithm for parallel computing

In ADVENTURE_FullWave, the full-wave analysis based on an E method [1] is considered. $\mathbf{E}_\mathrm{h}$ and $\mathbf{J}_h$ are finite element approximations of electric field $\mathbf{E}$ [V/m] and current density $\mathbf{J}$ [A/m^2], respectively. The permeability is given by $\mu = \mu_0\mu_r$ [H/m], μ_0 is the vacuum permeability [H/m], and μ_r is the relative permeability. The complex permittivity is given by $\varepsilon = \varepsilon_0\varepsilon_r - \sigma/j\omega$ [F/m], ε_0 is the vacuum permittivity [F/m], ε_r is the relative permittivity, and ω is the angular frequency [rad/s]. The following equation is the finite element equation to be solved:

$$\iiint_\Omega (1/\mu)\mathrm{rot}\mathbf{E}_\mathrm{h} \cdot rot\mathbf{E}_\mathrm{h}^* dv - \omega^2 \iiint_\Omega \varepsilon\mathbf{E}_\mathrm{h} \cdot \mathbf{E}_\mathrm{h}^* dv = j\omega \iiint_\Omega \mathbf{J}_h \cdot \mathbf{E}_h^* dv. \quad (1)$$

The equation contains complex numbers and becomes a complex symmetric matrix. In the present study, the electric field **E**, which is unknown, is obtained using the conjugate orthogonal conjugate gradient (COCG) method. The finite element approximation (1) is rewritten as $Ku = f$ by the coefficient matrix K, the unknown vector u, and the right-hand side vector f. Next, Ω is divided into N subdomains (Eq. (2)). Eq. (3) and (4) are obtained from Eq. (2).

$$\begin{bmatrix} K_{II}^{(1)} & 0 & 0 & K_{IB}^{(1)}R_B^{(1)\mathrm{T}} \\ 0 & \ddots & 0 & \vdots \\ & & K_{II}^{(N)} & K_{IB}^{(N)}R_B^{(N)\mathrm{T}} \\ 0 & 0 & & \\ R_B^{(1)}K_{IB}^{(1)\mathrm{T}} & \cdots & R_B^{(N)}K_{IB}^{(N)\mathrm{T}} & \sum_{i=1}^{N} R_B^{(i)}K_{BB}^{(i)}R_B^{(i)\mathrm{T}} \end{bmatrix} \begin{bmatrix} u_I^{(1)} \\ \vdots \\ u_I^{(N)} \\ u_B \end{bmatrix} = \begin{bmatrix} f_I^{(1)} \\ \vdots \\ f_I^{(N)} \\ f_B \end{bmatrix} \quad (2)$$

$$K_{II}^{(i)}u_I^{(i)} = f_I^{(i)} - K_{IB}^{(i)}u_B^{(i)} \quad (i = 1, \dots, N) \quad (3)$$

$$\left\{\sum_{i=1}^{N} R_B^{(i)}\left\{K_{BB}^{(i)} - K_{IB}^{(i)\mathrm{T}}\left(K_{II}^{(i)}\right)^{-1}K_{IB}^{(i)}\right\}R_B^{(i)T}\right\}u_B = \sum_{i=1}^{N} R_B^{(i)}\left\{f_B^{(i)} - K_{IB}^{(i)T}\left(K_{II}^{(i)}\right)^{-1}f_I^{(i)}\right\} \quad (4)$$

where ${f_B}^{(i)}$ is the right-hand vector for u_B, and $\left(K_{II}^{(i)}\right)^{-1}$ is the inverse matrix of $K_{II}^{(i)}$. Equation (4) is referred to as an interface problem and is an equation for satisfying the continuity between domains in the domain decomposition method. For simplicity, rewrite equation (5) as follows:

$$Su_B = g,$$
$$S = \sum_{i=1}^{N} R_B^{(i)}S^{(i)}R_B^{(i)T}, \quad S^{(i)} = K_{BB}^{(i)} - K_{IB}^{(i)T}\left(K_{II}^{(i)}\right)^{-1}K_{IB}^{(i)}. \quad (5)$$

3. PML

3.1. *Berenger's PML*

The PML can be used to create an absorbing boundary by surrounding the analysis domain with a PML. From the viewpoint of the accuracy of the obtained solution, the PML is currently the most effective absorbing boundary condition. Although Berenger's PML is originally proposed as an absorbing boundary condition for the FDTD method, in the present study, we apply a finite element method dealing with an unstructured grid, we propose a simplified method omitting the directionality of electric conductivity given to the PML and confirm its effectiveness.

Berenger's PML [2] stacks several PMLs outside the analysis domain and gradually sets a large value of electric conductivity according to the outer layer so that the outermost wall can be surrounded with a perfect conductor wall without reflecting electromagnetic waves.

Figure 1 shows a schematic diagram of Berenger's PML absorbing boundary.

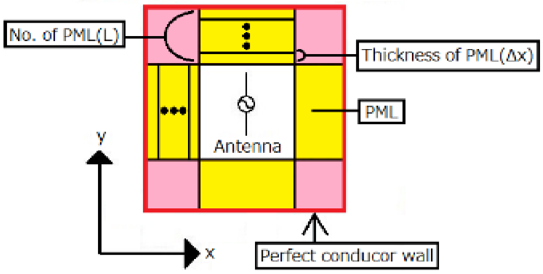


Fig. 1. PML absorbing boundary

In this paper the distribution of the electric conductivity for PML is expressed as follows:

$$\sigma = \sigma_{max}\left[\frac{\left(L - \hat{L}(x)\right)\Delta x}{L\Delta x}\right]^M \quad (6)$$

where Δx is the thickness of PML 1, L is the number of layers of the PML, $\hat{L}(x)$ is a coefficient determined by position x, and $\hat{L}(x)$ = 0 at the position of the Lth layer, $\hat{L}(x)$ = 1 at the position of the (L-1)th layer, and $\hat{L}(x)$ = L -1 at the position of the first layer.

Moreover, σ_{max} is the maximum value of the electric conductivity for the PML, and M is the degree distribution of electric conductivity. This equation is used to determine the electric conductivity of each layer of the PML.

The parameters to be determined as the parameters of the PML are the thickness Δx of PML 1, the number L of PML layers, the maximum electric conductivity σ_{max} of the PML, the degree M distribution of the electric conductivity, the reflection coefficient R [dB] between the PML of the outermost layer, and the perfect conductor wall. The reflection coefficient R is approximated as follows:

$$|R(\phi)| \cong exp\left[-\frac{2\sigma_{max}L\Delta x}{(M+1)\varepsilon_0 c}\cos\phi\right] \quad (7)$$

where ϕ is the incident angle of the electromagnetic wave, and c is the speed of light. Since we cannot decide the incident angle for an arbitrary incident wave, ϕ = 0, a reflection coefficient for perpendicular incidence is used as a reference. Moreover, since the M that gives the distribution of the electric conductivity causes the calculation accuracy to deteriorate if the change of the electric field in the PML is too steep, M is approximately 2 to 4. If the number of layers L is too large, more memory will be required, and if L is too small, it will not function adequately as an absorbing boundary. There are many cases where the concrete number of L is set to 4 to

16. The thickness Δx of PML 1 is a constant thickness of all layers.

We set the reflection coefficient $R(0)$ according to the required accuracy. Upon determining the above parameters, the maximum electric conductivity σ_{max} is given as follows:

$$\sigma_{max} = -\frac{(M+1)\varepsilon_0 c}{2L\Delta x}\ln|R(0)| \quad (8)$$

In the present study, we construct a PML using (6) through (8) with L = 9, M = 4, and Δx = λ/10. However, in order to reduce the analysis scale, we examine the optimum value of L in the next section.

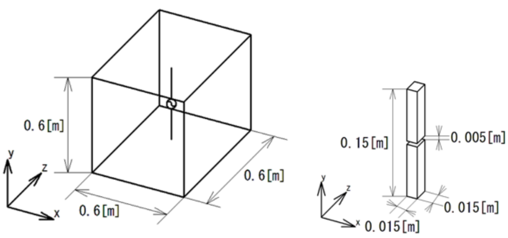
3.2. *Numerical results*

We assign the PML to the dipole antenna model. The analysis domain is a cube of length 0.6 [m] so that the distance from the antenna to the innermost PML matches the wavelength. The current density is applied to the antenna as a current source as follows:

$$I(y) = I_0 \cos\left(\frac{2\pi}{\lambda} y\right) \quad : -l \leq y \leq l \quad (9)$$

where I_0 = 0.08 [A/m^2], λ is the wavelength, and l is the length from the feeding point to the antenna tip.

The analysis frequency is 1 [GHz], and the length of the antenna is 0.15 [m], which is the half wavelength. Here, mesh division is performed so that the maximum side length of the element is 1/20 of the wavelength. The analysis domain's boundary is a perfect conductor. Figure 4 shows a schematic diagram of the dipole antenna model.



(a) Analysis domain (b) Antenna

Fig. 2. Dipole antenna model

We assign PMLs to the domain boundary as shown in Fig. 2(a). The plane portion of the PML at the domain boundary overlaps a number of flat plates according to the number of layers, and the corner portion of the PML is one rectangular parallelepiped or cube. The boundary of the outermost layer of the PML is a perfect conductor wall. We perform performance evaluation by setting the thickness of one layer to be 0.03 [m] and the PML to have L = 9 (hereinafter a PML with L layers is abbreviated as PML(L)). Table 1 lists the number of elements and the degree of freedom of the analysis model.

Table 1. Number of elements and DOFs of the dipole antenna model

	PML(0): Perfect conductor wall	PML(9)
No. of Elements	4,669,759	26,899,669
DOFs	5,506,368	31,703,550

In (8), we set L = 9, Δx = 0.03, M = 4, and $R(0)$ = -120 [dB], which yields the maximum electric conductivity σ_{max} to PML(9). In addition, we decide the electric conductivity of each layer using (6). In this study, we set the average value of each layer to the electric conductivity of the corner portion. We evaluate the performance of the PML based on the reflection coefficient obtained using the S_{11} parameter[3]. The observation point of the S_{11} parameter is on the x-axis 1 cm inside of the PML. The computing environment in the present study is a 25-PC cluster with Intel Core i7-2600K multi-core CPUs (total: 100 cores) and 32 GB memory. Table 2 lists the reflection coefficient, the CPU time, and the memory size.

Table 2. Results for reflection coefficient, CPU time, and memory size

	PML(0): Perfect conductor wall	PML(9)
Reflection coefficient [dB]	0	-18.65
CPU time [s]	1,278	18,787
Memory size [MB/core]	44.3	227.3

Acknowledgements

The present study was supported in part by a JSPS Grant-in-Aid for Scientific Research (Basic Research (A), 24H00695). The computer environment used in the present study was supported in part through a JST Adaptable and Seamless Technology transfer Program (A-STEP, JPMJTR24RF).

References

1. Takei, S. Sugimoto, M. Ogino, S. Yoshimura, H. Kanayama: Full Wave Analyses of Electromagnetic Fields with an Iterative Domain Decomposition Methodm, IEEE Transactions on Magnetics, 46:8 (2010), 2860-2863.
2. Takei, I. Higashi, M. Aikawa and T. Yamada, "Microwave analysis based on parallel finite element method," Journal of Advanced Simulation in Science and Engineering, 6:1 (2019), 215-233.

Authors Introduction

Prof. Amane Takei

He is working as Professor for Department of Electrical and systems Engineering, University of Miyazaki, Japan. His research interest includes high performance computing for computational electromagnetism, iterative methods for the solution of sparse linear systems, domain decomposition methods for large-scale problems. Prof. Takei is a member of IEEE, an expert advisor of The Institute of Electronics, Information and Communication Engineers (IEICE), a menber of Institute of Electrical Engineers of Japan (IEEJ) and Japan Society for Simulation Technology (JSST).

Prof. Akihiro Kudo

He received ph.D. degree from Nagaoka University of Technology He is a professor in the Department of engineering for innovation, National Institute of Technology, Tomakomai college. He is a member of Acoustical society of Japan, Information and Communication Engineers (IEICE).

Prof. Makoto Sakamoto

He is presently a professor in the Faculty of Engineering, University of Miyazaki. His first interests lay in hydrodynamics and time series analysis, especially the directional wave spectrum. He is a theoretical computer scientist, and his current main research interests are automata theory, languages and computation. He is also interested in digital geometry, digital image processing, computer vision, computer graphics, virtual reality, augmented reality, entertainment computing, complex systems, artificial intelligence, data science, and so forth.

Adsorption Equilibrium of Selenium Oxyanions Using FeY Mixed Oxides

Kaoru Ohe*, Amu Wakamatsu, Tatsuya Oshima
University of Miyazaki, 1-1 Gakuenkibanadai-Nishi,, Miyazaki, 889-2192, Japan
Email: okaoru@cc.miyazaki-u. ac.jp, amu.wakamatsu@gmail.com, oshimat@cc.miyazaki-u.ac.jp

** Corresponding author*

Abstract

Selenium oxyanions (Se(IV) and Se(VI)) in wastewater are of concern as serious pollutants due to their easy bioaccumulation and toxicity to living organisms. In this study, mixed oxides with various Y/Fe molar composition ratios (x=0.5, 1, 2) were investigated in the adsorption properties of selenium oxyanions, and were studied their adsorption mechanism. The experimental data best fits the Langmuir adsorption model, which is characteristic of monolayer adsorption (r^2 >0.991). The maximum adsorption capacity (q_{max}) of Se(IV) and Se(VI) increased in the order of FeY2 > FeY1 > FeY0.5 and of Fe2 > FeY1, respectively. It was suggested that FeY2 adsorbed Se(IV) by formation of inner-sphere complexes and outer-sphere complexes, and adsorbed Se(VI) by outer-sphere complexes.

Keywords: selenium oxyanions, mixed oxides, adsorption, iron, yttrium

1. Introduction

Selenium is a toxic element that is released into the environment due to human-induced pollution, such as mining activities and irrigated agriculture, and natural pollution, such as the dissolution of minerals [1], [2]. Because ingesting these elements can cause serious health problems in the human body, their removal from environmental water is an issue that needs to be urgently addressed. The World Health Organization (WHO) strictly regulates the guideline drinking water standard value at 0.01 mg dm^3. Selenium in environmental water exists as selenite Se(IV) and selenate Se(VI) . The pK_a values indicate that Se(IV) exists predominantly as H_2SeO_3, $HSeO_3^-$ and SeO_3^{2-} (pK_{a1}= 2.61 and pK_{a2}=8.46) [11] and Se(VI) as $HSeO_4^-$ and SeO_4^{2-} (pK_a= 1.70) [11]. Both oxyanions could result in health problems and are known to bioaccumulate in tissues.

Generally, the methods used to remove harmful metal ions from water include coagulation and coprecipitation with iron or aluminum, but these methods require a long time for treatment and produce a large amount of sludge. Adsorption using metal oxides is a useful technique for removing these pollutants from environmental water due to its effectiveness and simplicity. Magnetic adsorbents have attracted attention because they can be separated from water more easily and quickly under an external magnetic field [3]. Previous studies have reported that adsorbents with a high point of zero charge (PZC) and a large specific surface area exhibit high adsorption performance for arsenic and selenium [4], [5]. Since iron has a high affinity for selenium and it is known that the compounds of yttrium have a high isoelectric point (PZC Y_2O_3 : 9.3 [6], $Y(OH)_3$: 10.3 [7]). We prepared an adsorbent by mixing iron and yttrium, and obtained a higher PZC and higher specific surface area than conventional iron-based adsorbents [8].

In this study, we investigated the adsorption behavior of selenium oxyanions using FeY mixed oxides prepared using the method of Ohe *et al* [8]. We studied the adsorption mechanism of selenium oxyanions by measuring the effect of ionic strength on selenium adsorption and the zeta potential before and after selenium adsorption.

2. Experimental

2.1. *Adsorbents*

The specific surface area and PZC of FeYx (x is molar ration of Y(III)/Fe(II) = 0.5, 1, 2) prepared by the method of Ohe *et al* [8] increased with increasing x value. The adsorbents used for comparison, Fe_3O_4 and amorphous yttrium compound (AYA), were prepared using $FeCl_2$ and YCl_3, respectively, by the same method as FeYx. The specific surface area and PZC of FeYx, Fe_3O_4 and AYA were 140 m^2/g and 8.8 (FeY0.5), 166 m^2/g and 9.8 (FeY1), 183 m^2/g and 10.3 (FeY2), 29.4 m^2/g and 5.7 (Fe_3O_4), 59.1 m^2/g and 7.8 (AYA), respectively.

2.2. *Selenium adsorption experiments*

All adsorption experiments were carried out using batch method. For the pH dependence experiments, 1.0 ×10^{-4} mol dm^{-3} metal solution was adjusted to the specified pH by adding 1.0 ×10^{-2} mol dm^{-3} HCl solution or 1.0 ×10^{-2} mol dm^{-3} NaOH aqueous solution. In the adsorption isotherm experiments, Se(IV) and Se(VI) concentrations were used to 5.0×10^{-5}- 3.0×10^{-3} mol dm^{-3} and 1.0×10^{-5}- 4.0×10^{-3} mol dm^{-3}, respectively, and the pH was adjusted to the specified value using 1.0×10^{-2} mol dm^{-3} N, N-Bis(2-hydroxyethyl)-2-aminoethane sulfonic acid (BES) buffer and 1.0×10^{-2} mol dm^{-3} NaOH aqueous solution. In the experiment for the effect of ionic strength on selenium adsorption, metal solution of 1.0×10^{-4} mol dm^{-3} was prepared, containing 1.0×10^{-3} - 1.0×10^{-1} mol dm^{-3} NaCl.

The pH was adjusted using 1.0×10^{-2} mol dm^{-3} HCl solution or 1.0×10^{-2} mol dm^{-3} NaOH aqueous solution. FeYx 10 mg and 15 cm^3 of the metal solution were added to a sample tube, and the mixture was shaken at 303 K, 120 rpm, for 24 h in water bath. After shaking, the mixture was filtered through 0.45 µm hydrophilic PTFE membrane filter, and the equilibrium pH of the filtrate was measured. The metal concentrations of the filtrate were measured using atomic absorption spectrophotometer (AAS, HITACHI Z-2310, SHIMADZU AA-7000) or ICP emission spectrometer (ICP-AES, SHIMADZU ICPS-8100). The percentage of adsorption (*A*%, Eq.(1)) and the amount of selenium adsorbed (*q*, Eq.(2)) were calculated as follows:

$$A\% = 100\cdot(C_i - C_{eq})/ C_i \quad (1)$$

$$q = (C_i - C_{eq})\cdot v/w \quad (2)$$

where C_i is initial concentration of selenium, C_{eq} is equilibrium concentration of selenium, q is the amount of selenium adsorbed [mmol/g], w is the weight of adsorbents [mg], v is the volume of the solution [cm^3]

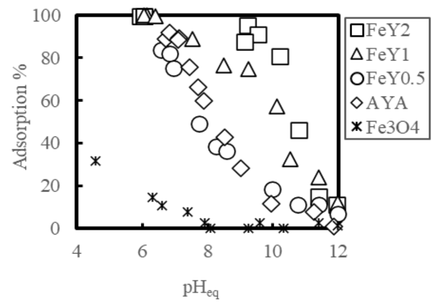
3. Results and Discussion

3.1. *Effect of pH on selenium oxyanions adsorption*

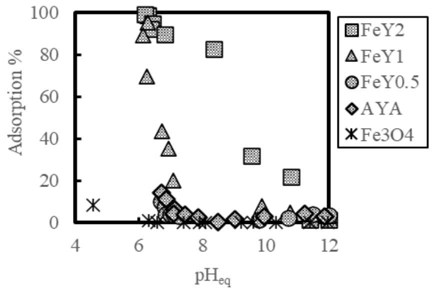
Fig.1a and Fig.1b show the effect of pH on selenite (Se(IV) and selenate(VI) adsorption. The maximum adsorption percentage for Se(VI) on FeY2 (pH 5.9) and FeY1 (pH 6.1) showed 100% (Fig.1a). At pH 4.0 -12, FeYx showed an adsorption rate equal to or higher than that of the comparative samples Fe_3O_4 and amorphous yttrium adsorbent (AYA). In addition, FeY2 showed an adsorption rate of over 90% in a wide range of pH 5.9-9.5. From Fig. 1b, it can be seen that FeY1 and FeY2 adsorbed Se(VI) in the pH range of 6.0-12, while the other adsorbents hardly adsorbed any. FeY2 showed an adsorption rate of over 90% at pH 6.0 exhibited an adsorption rate of over 90%, but it was found that the adsorption rate decreased as the pH increased. FeY1 showed a significant decrease in adsorption rate at pH 6.0-7.0, and at pH 7.0, the adsorption rate was 10%. FeY2 adsorbed Se(VI) over a wider pH range than FeY1. Se(IV) exists as $HSeO_3^-$ or SeO_3^{2-} in the pH range 4<pH<12 (Fig.2a), and Se(VI) exists as SeO_4^{2-} in aqueous solution(Fig.2b). Thus, at pH >PZC, the adsorption rate is thought to have decreased due to the electrostatic repulsion between the negatively charged adsorbent surface and Se(IV) and Se(VI).

3.2. *Adsorption isotherms of selenium oxyanions*

Adsorption isotherms for Se(IV) and Se(VI) at pH_{eq} 6.3-6.7 (303 K) were investigated. All adsorption isotherms (Fig.3) fit well Langmuir adsorption isotherms as shown by Eq. (3)

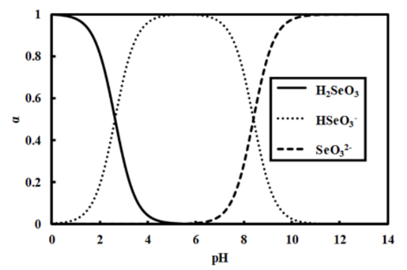


a. Se(IV) adsorption

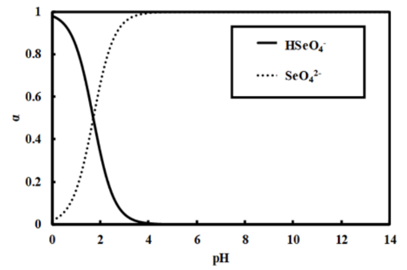


b. Se(VI) adsorption

Fig. 1. The pH dependence for Se(IV) and Se(VI) adsorptions



a. Se(IV)

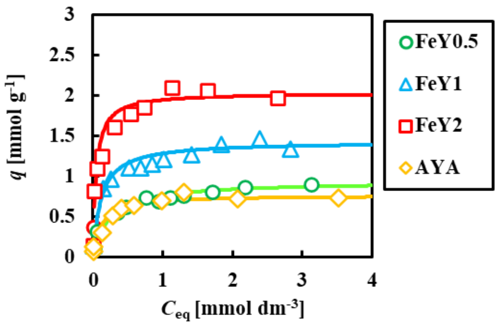


b. Se(VI)

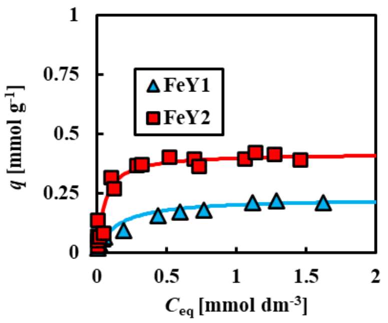
Fig. 2. Distribution ratio (α) of chemical species of Se(IV) and Se(VI).

$$q = q_{max}\cdot K_L\cdot C_{eq} / (1+K_L\cdot C_{eq}) \quad (3)$$

where q_{max} is amount adsorbed at saturation (mmol/g), and K_L is adsorption equilibrium constant (dm^3/mmol). The parameters for the adsorption isotherms of Se(IV) and Se(VI) are summarized in Table 1. Se(IV) and Se(VI) weren't almost adsorbed on Fe_3O_4 at pH_{eq} 6.1. For the adsorption of Se(IV), it was found that q_{max} increased in the order of FeY0.5 < FeY1 < FeY2. Furthermore, the q_{max} of FeYx was higher than that of



a. Se(IV) adsorption



b. Se(VI) adsorption

Fig. 3. Adsorption isotherms of Se(IV) and Se(VI) at 303K with FeYx and AYA. Solid lines are calculation curves of Langmuir adsorption

Table. 1. The maximum adsorption capacity (q_{max}) for Se(IV) and Se(VI) at 303K.

Adsorbents	q_{max} Se(IV) [mmol g^{-1}]	q_{max} Se(VI) [mmol g^{-1}]	References
FeY2	2.04 (pH 6.3-6.5)	0.418 (pH6.5-6.7)	this study
FeY1	1.43 (pH 6.2-6.4)	0.230 (pH6.5-6.7)	this study
FeY0.5	0.936 (pH 6.3-6.5)	Almost no adsorption	this study
AYA	0.761 (pH6.6-6.7)	Almost no adsorption	this study
Fe-Mn	0.0832 (pH 4.0)	0.00974 (pH 7.0)	12
Fe-Cu	0.179 (pH 7.4)	0.0756 (pH 7.4)	2

AYA. These results show that FeYx adsorbs more Se(IV) than Fe-Mn [12] , Fe-Cu [2] adsorbents and AYA, and that FeY2 is the most suitable for Se(IV) adsorption. The q_{max} of Se(VI) by FeY1 and FeY2 was 0.230×10^{-3} mol g^{-1} and 0.418×10^{-3} mol g^{-1}, respectively. On the other hand, FeY0.5, Fe_3O_4 and AYA did not adsorb Se(VI) at all. These results show that adsorbents with high PZC can effectively adsorb Se(VI). In other words, it was found that FeY2 is most suitable for the adsorption of Se(IV) and Se(VI). The q_{max} of Se(VI) with FeY2 were 2.4 times higher than that of Se(VI), respectively. FeY2 has very high adsorption performance for Se(IV).

3.3. *Adsorption Mechanism*

The following experiments were conducted using FeY2, which showed a high saturation adsorption capacity for selenium. PZC shift refers to shift of PZC to lower pH, which is used as evidence of strong adsorption with specific ions and the formation of inner-sphere complexes. Adsorption by inner-sphere complexes is a direct coordination bond formed between the adsorbed ion and the metal ion on the surface of the adsorbent without the inclusion of water molecules. On the other hand, adsorption by outer-sphere complexes is adsorption by electrostatic attraction involving one or more water molecules between the metal ion on the adsorbent surface and the adsorbed ion. Assuming that the outer-sphere complex is located outside the slip plane, it is possible to distinguish between outer-sphere complexes and inner-sphere complexes by measuring the zeta potential [9].

Fig. 4 shows zeta potential of FeY2 before and after adsorption of selenium oxyanions. As there is no PZC shift before and after selenium adsorption, it is thought that Se(VI) forms an outer-sphere complex and is adsorbed onto FeY2. Whereas the PZC of FeY2 after Se(IV) adsorption shifted to lower pH than before the adsorption. Se(IV) was adsorbed inner-sphere complex on the surface of FeY2.

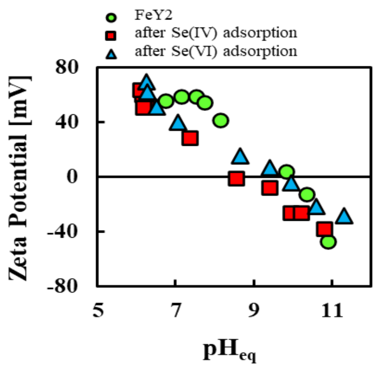


Fig. 4. Zeta potential of FeY2 after Se(IV) and Se(VI) adsorption.

By investigating the effect of ionic strength using a neutral electrolyte, it is possible to indirectly determine whether surface complex of adsorbed selenium oxyanion and adsorbent is an inner-sphere complex or an outer-sphere complex [10]. If selenium adsorption is due to formation of outer-sphere complex, adsorption decreases as the ionic strength of the solution increases due to competition with the anion of neutral electrolyte. On the other hand, Selenium adsorption is due to the formation of inner-sphere complexes, selenium is adsorbed by coordination bonding with the adsorption site on adsorbent, so it is thought that it is not affected by ionic strength. Fig.5 shows the effect of ionic strength on Se(IV) adsorption with FeY2. It was found that the adsorption percentage decreased significantly with the addition of NaCl. The adsorption of Se(VI) was also significantly affected on ionic strength and its adsorption behavior indicated the same as Se(IV) adsorption. Thus, Se(IV) wa inner-sphere complexes and outer-sphere complexes and Se(VI) was adsorbed on FeY2 by forming outer-sphere complexes.

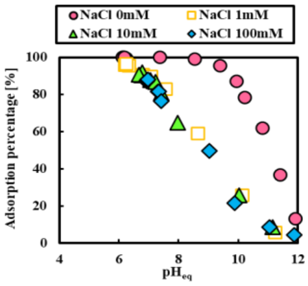


Fig. 5. Effect of ionic strength on Se(IV) adsorption with FeY2.

4. Conclusions

In this study, mixed iron and yttrium oxides (FeYx, where x is the molar ratio of Y(III)/Fe(II)) were prepared as adsorbents for selenium anions in water. FeY2 was found to have the largest specific surface area and the highest PZC among the FeYx, these values improved on magnetite and AYA. FeY2 adsorbed selenium over a wider pH range than conventional adsorbents, and increased the maximum adsorption capacities of Se(IV) and Se(VI) were increased. The influence of zeta potential and ionic strength showed that Se(IV) was adsorbed on FeY2 by the formation of inner-sphere complexes and harmful-sphere complexes, and that Se(VI) was adsorbed on FeY2 by the formation of outer-sphere complexes. FeY2 is expected to be applied to the treatment of high-contamination water.

References

1. N. Chubar, M. Szlachta, Static and dynamic adsorptive removal of selenite and selenate by alkoxide-free sol–gel-generated Mg–Al–CO_3 layered double hydroxide: Effect of competing ions, Chemical Engineering Journal 279, 2015, pp. 885-896, doi.org/10.1016/j.cej.2015.05.070
2. W. Sun, W. Pan, F. Wang, N. Xu, Removal of Se(IV) and Se(VI) by MFe_2O_4 nanoparticles from aqueous solution, Chemical Engineering Journal 273, 2015, pp. 353-362, doi.org/10.1016/j.cej.2015.03.061
3. L. C. A. Oliveria, R. V. R. A. Rios, J. D. Fabris, K. Sapag, V. K. Grag, R. M. Lago, Clay–iron oxide magnetic composites for the adsorption of contaminants in water Applied Clay Science 22, 2003, pp. 169-177, doi.org/10.1016/S0169-1317(02)00156-4
4. Ohe, K., R. Tomimatsu R., T. Oshima, Y. Baba, Removal of As (III) and As (V) in Groundwater Using Ferrite Adsorbents, Journal of Ion Exchange 18(4), 2007, pp. 550-553. doi.org/10.5182/jaie.18.550
5. Y. T. Chan, W. H. Kuan, T. Y. Chen, M. K. Wang, Adsorption mechanism of selenate and selenite on the binary oxide systems, Water Research 43, 2009, pp. 4412-4420, doi:10.1016/j.watres.2009.06.056
6. R. Sprycha, J. Jablonski, E. Matijević, Zeta potential and surface charge of monodispersed colloidal yttrium(III) oxide and basic carbonate, Journal of Colloid and Interface Science 149, 1992, pp. 561-568, doi.org/10.1016/0021-9797(92)90443-P
7. S.-H. Lee, K.-W. Kim, B.-T. Lee, S. Bang, H. Kim, H. Kang, A. Jang, Int. Enhanced Arsenate Removal Performance in Aqueous Solution by Yttrium-Based Adsorbents, International Journal of Environmental Research and Public Health 12, 2015, pp. 13523-13541, doi.org/10.3390/ijerph121013523
8. Ohe, K., A. Wakamatsu, T. Oshima, Development of yttrium(III)-incorporated magnetite for adsorptive removal of arsenic and a study of the adsorption mechanism, Kagaku Kogaku Ronbunshu 45, 2019, pp. 168-176, doi.org/10.1252/kakoronbunshu.45.168
9. S. Goldberg, C. T. Johnston, Mechanisms of Arsenic Adsorption on Amorphous Oxides Evaluated Using Macroscopic Measurements, Vibrational Spectroscopy, and Surface Complexation Modeling, Journal of Colloid and Interface Science 234, 2001, pp. 204-216, doi.org/10.1006/jcis.2000.7295
10. M. B. McBride, A Critique of Diffuse Double Layer Models Applied to Colloid and Surface Chemistry, Clays and Clay Minerals 45(5), 1997, pp. 598-608, doi:10.1346/CCMN.1997.0450412
11. D. A. Dzombak and F. M. M. Morel, Surface complexation modeling: Hydrous Ferric oxide, Wiley, 1990, ISBN: 978-0-471-63731-8
12. C.-M. Gonzalez, J. Hemandes, J. G. Parsons, J. L. Gardea-Torresdey, A study of the removal of selenite and selenate from aqueous solutions using a magnetic iron/manganese oxide nanomaterial and ICP-MS Microchemical Journal 96, 2010, pp. 324-329, doi.org/10.1016/j.microc.2010.05.005

Authors Introduction

Dr. Kaoru Ohe

She received her Ph.D. degrees from University of Miyazaki, Japan, in 2014. Currently she is an Associate Professor of Center for Science and Engineering Education, Faculty of Engineering. Her researches are separation engineering especially adsorption hazardous heavy metals and oxyanions.

Ms Amu Wakamatsu

She received Master’s degree in Engineering in 2018 from Faculty of Engineering, University of Miyazaki in Japan.

Prof. Dr. Tatsuya Oshima

He received his Ph.D. degrees from Kyushu University, Japan, in 2001. Currently he is an Professor of the Department of Applied Chemistry, Faculty of Engineering. His researches are separation engineering (especially solvent extraction for precious metal ions) and formulation techniques for poorly water-soluble drugs and nutrients.

Influence of CNN Layer Depth on Spiral Visual Illusions

Kenji Aoki
University of Miyazaki, 1-1 Gakuenkibanadai-nishi, Miyazaki-city, Miyazaki, 889-2192, Japan

Makoto Sakamoto
University of Miyazaki, 1-1 Gakuenkibanadai-nishi, Miyazaki-city, Miyazaki, 889-2192, Japan
Email: aoki@cc.miyazaki-u.ac.jp, saka2000@cc.miyazaki-u.ac.jp

Abstract

Understanding the mechanism of visual illusion generation through Convolutional Neural Networks (CNNs) that mimic the receptive fields of the visual cortex can contribute to elucidating the mechanisms of visual information processing in the brain. In our previous research, we demonstrated the potential for Fraser's spiral illusion to manifest in CNNs. In this study, we focused on the depth of the CNN layer structure and examined the impact of the number of layers on the manifestation of the visual illusion. We provided 14 types of spiral illusion images to three different CNN patterns with varying layer structures and tasked them with distinguishing between concentric circles and spirals. The results indicated that CNNs with fewer layers were more prone to the illusion, whereas CNNs with more layers were less likely to exhibit the illusion. These results suggest that the number of layers in a CNN influences the manifestation of visual illusions.

Keywords: Convolutional Neural Network, CNN, Visual Illusion, Spiral Illusion

1. Introduction

Convolutional Neural Networks (CNNs), which mimic the local receptive fields of the visual cortex, have been extensively studied and utilized in various fields [1], including image recognition [2]. In human vision, under certain conditions, phenomena known as "visual illusions" occur, where shapes, sizes, lengths, colors, and directions are perceived differently from their physical reality. The occurrence of visual illusions suggests that visual information from the retina is not transmitted directly to the brain but undergoes some processing. Therefore, investigating whether CNNs can exhibit illusions and understanding the mechanisms behind them, could contribute to elucidating the mechanisms of visual information processing.

In the study by Watanabe et al., experiments were conducted using deep neural networks to observe motion from "rotating snakes" images, a type of motion illusion, to determine if the same illusions observed in humans could be replicated [3]. Motion illusions are phenomena where stationary images are perceived as moving. In the case of "rotating snakes," disk-shaped images resembling snakes appear to rotate. The results showed that the neural network predicted rotational motion from the "rotating snakes" images, indicating that deep neural networks can exhibit illusions.

Our previous research also demonstrated the potential for Fraser's spiral illusion to occur in CNNs [4]. Fraser's spiral illusion is a phenomenon where concentric circles are perceived as a spiral. This study suggested that the structure of the CNN model influences the occurrence of visual illusions. Horikawa et al. compared CNNs with the human brain and reported homology in information representation between each layer of the CNN and each region of the human visual cortex [5]. This implies that altering the number of convolutional layers in CNNs could affect the occurrence of illusions. Additionally, Simonyan et al. reported that the deeper the layers of a CNN, the higher the accuracy of image classification [6]. Therefore, in our research, we hypothesized that increasing the depth of the layers would enhance image discrimination accuracy, leading to the correct identification of physically concentric images as concentric, thereby reducing the occurrence of visual illusions. In this study, we focused on the depth of the CNN layer structure and examined the influence of the number of layers on the occurrence of visual illusions.

2. Methodology

2.1. *CNN models*

The CNN constructed in this study is composed of several layers, including convolutional layers, normalization layers, and pooling layers. In the input layer, the pixel values (0-255) of the input image are provided. The convolutional layers extract local features by multiplying the input values with the filter values and summing the neighboring output values. The results of the convolution are converted into output signals using an activation function. The ReLU (Rectified Linear Unit) function was used as the activation function. The ReLU function outputs the input value directly if it is greater than 0, and outputs nothing if it is 0 or less.

The pooling layers perform down-sampling to reduce the complexity of the subsequent layers, which is equivalent to reducing the resolution in image processing [7]. This helps to mitigate the effect on the output results when the input image is slightly shifted. In this study, max pooling was used. Max pooling divides the image into small rectangular regions and outputs only the maximum value within each region to the next layer.

Batch Normalization, developed by Serger et al., is a technique to accelerate deep learning [8]. It also reduces dependency on initial values and helps prevent overfitting. Dropout is a technique that randomly deletes neurons during training to prevent overfitting.

These layers were combined to construct the CNN model. The structure of the three models constructed in this study is shown in Fig. 1. The dimension of the input layer is 150 x 150 for all models. Grayscale images of 150 x 150 pixels were provided as input, and the images passed through multiple convolutional layers (Conv2d in Fig. 1), Batch Normalization layers (BatchNormalization in Fig. 1), and pooling layers (MaxPooling in Fig. 1). Finally, the images passed through a global average pooling layer (GlobalAveragePooling in Fig. 1), Dropout (Dropout in Fig. 1), and a fully connected layer (Dense in Fig. 1) to output either "0" or "1". An output value of "0" indicates a concentric circle, while "1" indicates a spiral.

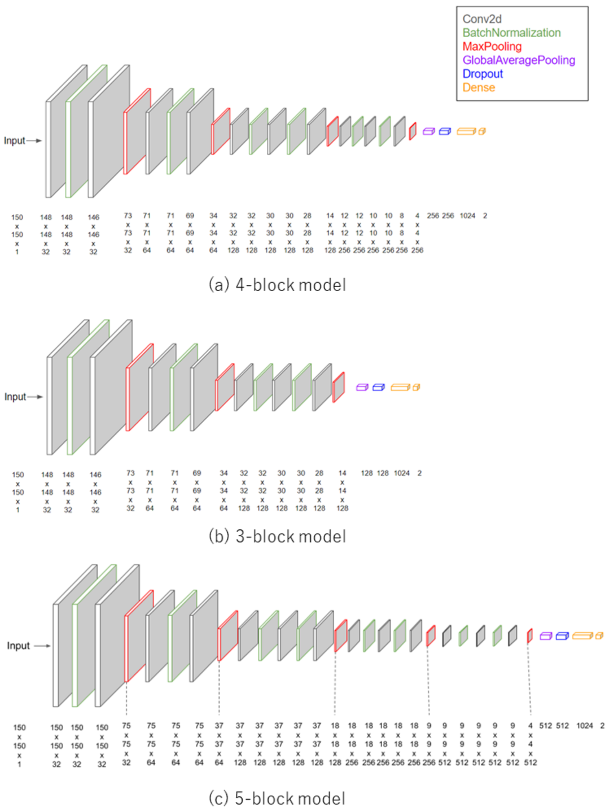


Fig.1 Composition of the 3 CNN models

The convolutional layers, Batch Normalization layers, and Max Pooling layers were combined into a "block," and models with different depths were constructed by combining four, three, and five blocks. The configurations of these models are shown in Fig. 1 (a), (b), and (c), respectively. Each section separated by dotted lines in Fig. 1 (c) represents one block. In our previous research, we demonstrated the potential for spiral illusion occurrence using a model composed of four blocks (4-block model in Fig. 1 (a)) [4]. We investigated whether visual illusions occurred in these three models and whether there were any changes in the occurrence of visual illusions.

2.2. *Computing Environment*

In this study, we utilized Python (ver. 3.9.6), a programming language rich in machine learning libraries, and TensorFlow (ver. 2.5.0), an open-source machine learning software library developed by Google, for machine learning using CNNs. The computer used for this study had the following specifications: OS - Windows 10, CPU - Intel Xeon X3480, GPU - NVIDIA GeForce RTX3060, and RAM - 20.0GB.

2.3. *Training Datasets*

To train the constructed CNN to distinguish between spiral and concentric circle images, we created a training dataset. First, we generated 50 images each of 150 x 150 pixels for both concentric circles and spirals. Additionally, we augmented these images by applying horizontal and vertical shifts, flips, and scaling. Including the original images, we prepared 250 images each for concentric circles and spirals. From these, 200 images of each type were randomly selected as training data, and 50 images of each type were selected as test data. Fig. 2 shows some of the created training data images. Fig. 2 (a) shows concentric circle images, and Fig. 2 (b) shows spiral images. In addition to images of concentric circles and spirals drawn with solid and dashed lines, we created images with features seen in spiral illusions. The left side of Fig. 2 shows images drawn with solid lines, the center with dashed lines, and the right side shows images combining black and white lines and triangular endpoints, similar to those seen in spiral illusion images (e.g., Fig. 3 06). We also prepared patterns with different combinations of background and line colors, such as white and black, black and white, and gray and black. Note that the training dataset does not include spiral illusion images.

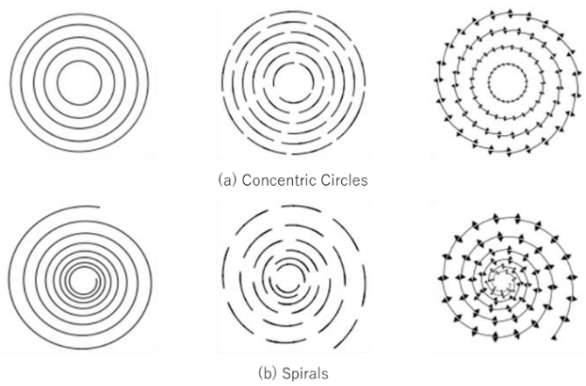


Fig.2 Example of training data set

2.4. *Validation Images*

After training the CNN model, we provided 14 spiral illusion images as input and checked the output. The spiral illusion images used for validation are shown in Fig. 3. These images were prepared with reference to those published on the "Akiyoshi Kitaoka's Illusion Pages" [9]. The images were standardized to a size of 150 pixels by 150 pixels and converted to 256-level grayscale. Images 01 to 14 in Fig. 3 are actually concentric circles, but they exhibit a visual illusion phenomenon where they appear spiral when viewed by humans.

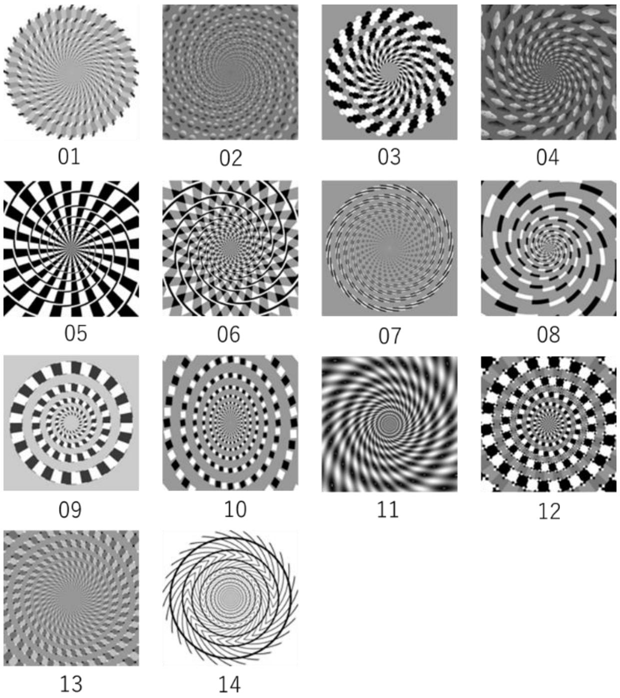


Fig.3 Spiral illusion images for verification

3. Results and Discussion

We checked the changes in the loss function values and accuracy with respect to the number of training iterations during the training process of the constructed CNN models. Cross-entropy error was used as the loss function. For all three models shown in Fig. 1, it was confirmed that the loss function values and accuracy for both the training and test data stabilized after 300 epochs, so the models were trained for up to 500 epochs for validation.

To eliminate the possibility of random classification as spirals, we randomly changed the initial values and repeated the training and validation 10 times. Fig. 4 shows the number of times each validation image was classified as a visual illusion. The vertical axis represents the number of times the image was classified as a spiral out of 10 validations, and the horizontal axis corresponds to the image names shown in Fig. 3. The results for the 4-block model, 3-block model, and 5-block model are shown in blue, red, and yellow, respectively.

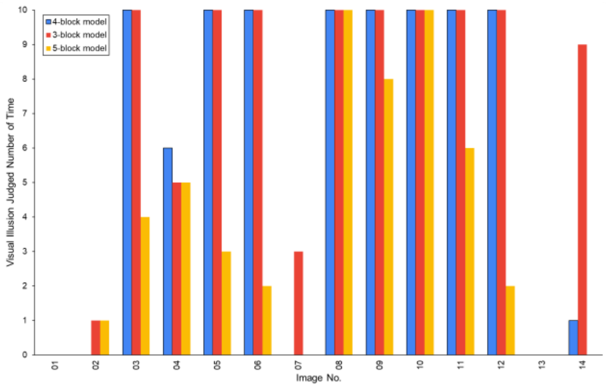


Fig.4 Comparison of the number of spiral illusion occurrences among three models

For the 4-block model (blue in Fig. 4), 8 out of 14 images (57%) were consistently classified as spirals in all 10 trials, indicating the occurrence of spiral illusions. Additionally, 4 out of 14 images (29%) were consistently classified as concentric circles, indicating no illusion. The remaining images had mixed classifications.

For the 3-block model (red in Fig. 4), 8 out of 14 images (57%) were consistently classified as spirals in all 10 trials, indicating the occurrence of spiral illusions. Additionally, 2 out of 14 images (14%) were consistently classified as concentric circles, indicating no illusion.

For the 5-block model (yellow in Fig. 4), 2 out of 14 images (14%) were consistently classified as spirals in all 10 trials, indicating the occurrence of spiral illusions. Additionally, 4 out of 14 images (29%) were consistently classified as concentric circles, indicating no illusion.

The 3-block model showed a similar rate of spiral classification as the 4-block model. However, images "02" and "07", which were not classified as spirals in the 4-block model, were classified as spirals in the 3-block model. Additionally, image 14 was classified as a spiral more frequently in the 3-block model than in the 4-block model. The 5-block model classified images as spirals less frequently than the 4-block model. These results suggest that deeper layers may reduce the occurrence of visual illusions. However, it was not clearly demonstrated whether deeper layers lead to more physically accurate classifications.

4. Conclusion

In this study, we constructed CNNs with varying depths, trained them with concentric circle and spiral images, and used the models to test for the occurrence of spiral illusions. By comparing the results across different models, we found that deeper layers might reduce the occurrence of visual illusions. In the future, we plan to investigate the influence of changing the training images and further modifying the model depth on the occurrence of illusions. Additionally, we aim to examine and compare the features extracted by each layer to understand the differences.

References

1. A. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, K. Kavukcuoglu, WaveNet: A Generative Model for Raw Audio, 2016. doi:10.48550/arXiv.1609.03499
2. Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition. Proceedings of the IEEE, vol. 86 (11), 1998, pp. 2278-2324. doi:10.1109/5.726791
3. E. Watanabe, A. Kitaoka, K. Sakamoto, M. Yasugi, K. Tanaka, Illusory Motion Reproduced by Deep Neural Networks Trained for Prediction. Frontiers in Psychology, vol. 9, 2018. doi:10.3389/fpsyg.2018.00345
4. K. Aoki., T. Togo., M. Sakamoto. Investigating Visual Illusions in Convolutional Neural Networks Using Spiral Illusion Images, Proceedings of the Sixteenth International Conference on Genetic and Evolutionary Computing, Genetic and Evolutionary Computing, Lecture Notes in Electrical Engineering 1322, 2024, (in press).
5. T. Horikawa, Y. Kamitani, Generic decoding of seen and imagined objects using hierarchical visual features. Nature Communications, vol. 8 (1), 2017, p. 15037. doi:10.1038/ncomms15037
6. K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition., 2014. doi:10.48550/arXiv.1409.1556
7. S. Albawi, T. A. Mohammed, S. Al-Zawi, Understanding of a convolutional neural network. 2017 International Conference on Engineering and Technology (ICET), 2017, pp.1-6. doi:10.1109/ICEngTechnol.2017.8308186
8. S. Ioffe, C. Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, 2015. dio:10.48550/arXiv.1502.03167
9. A. Kitaoka, Akiyoshi’s illusion pages. http://www.psy.ritsumei.ac.jp/~akitaoka/index-j.html, last accessed 2022/02/07.

Authors Introduction

Dr. Kenji Aoki

He received a Ph.D. in Engineering from Kagoshima University. He is working in the Information Technology Center at University of Miyazaki as Associate Professor. His research interests include visual information processing, artificial intelligence, campus information systems and educational information systems. He is a member of IPSJ and JSET.

Dr. Makoto Sakamoto

He received the Ph.D. degree in computer science and systems engineering from Yamaguchi University. He is a professor in the Faculty of Engineering, University of Miyazaki. He is a theoretical computer scientist, and his current main research interests are computer science and information processing.

Evaluation of Passive Interaction in XR Chakra Meditation Application Based on Behavioral Biometrics

P. I. A. Gayathri Bimba
Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi City, Shikawa, 923-1211, Japan

Chie-Tung Lin, **R.P.C. Janaka Rajapakse**
Tainan National University of the Arts, No. 66, Daqi, Guantian Dist. Tainan City, 72045, Taiwan

Kazunori Miyata
Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi City, Shikawa, 923-1211, Japan
Email: g.illanperuma@yahoo.com, feyin.tw@gmail.com, janakaraja@gmail.com, miyata@jaist.ac.jp

Abstract

This study explores the effectiveness of an XR chakra meditation application developed using behavioral biometrics. A direct comparison with a similar application was impossible since no commercial XR-based chakra meditation apps exist. Instead, the study compared the XR application with popular YouTube chakra meditation video, which are widely used for meditation, as indicated by their high viewership and subscriber counts. Participants provided feedback through questionnaires and brainwave data to evaluate its strengths and weaknesses of use. The experiment focused on several aspects, including the level of interference with meditation, ease of learning and use, physical and emotional effects on users, user preferences, and the perception of vibrational realism between the two methods.

Keywords: Evaluation, Passive Interaction, XR Chakra Meditation, Behavioral Biometrics, EEG

1. Introduction

Extended reality (XR) includes a spectrum of immersive technologies that mix physical and virtual environments, including virtual reality (VR), augmented reality (AR) and mixed reality (MR). This vast area creates a transformative platform for various applications, including mental health, education, and well-being practices [1]. In the discipline of meditation, these technologies provide users with a novel method to commence mindfulness practices with immersive experiences [2], [3], [4]. The meditation of the chakra, originating from ancient Indian spiritual traditions, emphasizes the proper balance and alignment of the body's energy places of learning, known as chakras [5]. These seven energy centers are associated with the different mental, physical, and spiritual levels of an individual's wellness. Each chakra regulates particular psychological functions, encompassing creativity, intuition, self-expression, and spiritual connection. The cultural importance of chakra meditation is significant, interwoven with disciplines such as Yoga, Ayurveda, and traditional healing techniques [6]. As such, the chakra meditation serves as a therapeutic exercise and a holistic commitment to the physical and spiritual self. More specifically, a meditation application of XR Chakra can improve traditional practices by creating a passive interaction environment that promotes a deeper concentration and self-awareness, thus potentially increasing the global meditation experience [7]. The integration of behavioral biometrics into XR meditation applications is particularly relevant.

Fig. 1. XR Chakra meditation application of virtual and vibrational realities based on behavioral biometrics [7].

Traditional meditation practices are often firmly based on the practitioner's internal feedback mechanisms, generally derived from sensory experiences and cognitive states. However, improving these processes by behavioral biometrics allows immediate feedback and, based on data, can adjust the meditation experience according to the user's emotional and cognitive responses. As described in our previous development, monitoring psychological and physiological states, such as brainwave signals and heart rate variability, associated with behavioral analysis concerning passive interaction models in the XR environment, can provide real-time information on the user's state of mind [7]. This multidimensional feedback mechanism supports a more personalized meditation regime, allowing tailor-made interventions and extending the user's commitment to practice (Fig. 1). In addition, the interactive elements of the system enable users to view the energy flows associated with the chakras, increasing their meditative experience by providing educational information on energy centers visually engagingly.

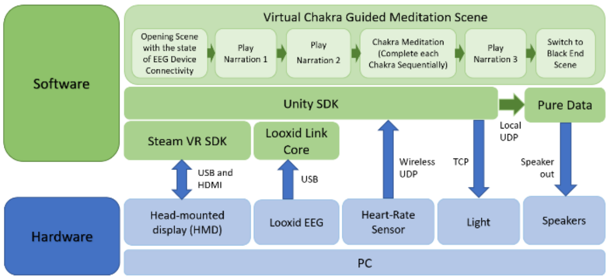


Fig. 2. System overview [7].

Fig. 2 illustrates how the XR Chakra meditation program functions within a complex system architecture that comprehensively integrates virtual and vibrational realities while incorporating behavioral biometrics to improve user experience. By combining visual and auditory components with the neurofeedback that EEG provides, the program could enhance the benefits of meditation, promoting emotional and spiritual well-being as well as relaxation.

This paper aims to assess the effectiveness of a Chakra XR meditation application compared to popular YouTube videos focused on meditation of the chakra. While YouTube serves as a widespread platform to access guided meditations, the nature of its passive visualization could be critical in contrast with the requirements of XR technology for the active engagement of its users. This change in perspective could offer new insights into how technology might enhance meditation practices, fostering situations that motivate users to engage and apply their consciousness in real time. The current investigation employed a mixed-method approach, integrating qualitative feedback with quantitative behavioral biometric measures to evaluate the entire meditation experience. Understanding these processes clarifies the frameworks of engagement, attention, and emotional states, encouraging a more sophisticated knowledge of meditation in digital contexts. The growing demand for novel well-being solutions compels the evaluation of the effectiveness of XR applications in enhancing traditional methods like chakra meditation, highlighting the significant convergence of culture and technology.

2. Methodology

The following section explains the design of the research experiment, the experimental subjects, and the procedure in sequence.

2.1. *Experimental Design*

The following are the questions to be studied in this experiment:

- Compare the effects of the XR Chakra meditation application and YouTube Chakra meditation video for beginners.
- Whether the sound frequency and color light related to Chakra can increase the effect of Chakra meditation.

The experimental design of the questionnaire is to let the subjects watch the YouTube Chakra guided video [8] first, fill in the questionnaire, and then use this XR Chakra meditation application to fill in the questionnaire again so that the differences between the two can be compared. Brain waves can be recorded throughout the process when using these two meditation methods.

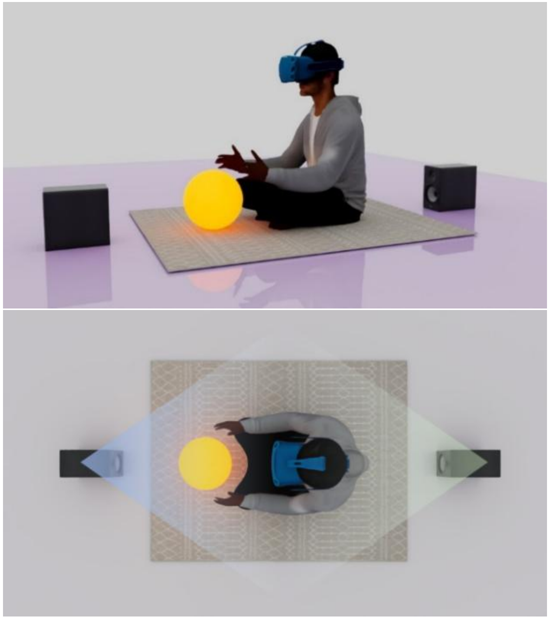


Fig. 3. Experimental environment and experimental conditions.

2.1.1 Experimental Environment

The experiment was held in a classroom with adequate space because it had ample space and less unnecessary interference. Based on its popularity, the Meditative Mind channel's "Quick 7 Chakra Cleansing | 3 Minutes Per Chakra | Seed Mantra Chanting Meditation | Root to Crown" was selected as the experimental stimulus that compared with XR Chakra meditation application in this experiment [8]. As of November 2024, the channel has over 6 million subscribers, and this chakra-guided video has been viewed more than 15 million times. Experimental subjects can watch this YouTube Chakra guided video on the Head Mounted Display (HMD) and use the XR Chakra meditation application as stimuli with the same experimental conditions. The venue layout and experimental conditions are shown in Fig. 3. Since the HTC-VIVE Pro VR system uses at least two base stations at diagonal height for positioning, the venue size is about 2 m x 3m. Two speakers are positioned in front of and behind the user, respectively. As shown in Fig. 3, when the user places their palm on the spherical lamp, the sound frequency emitted from the speakers causes the lamp to vibrate slightly, transmitting the sensation to the body.

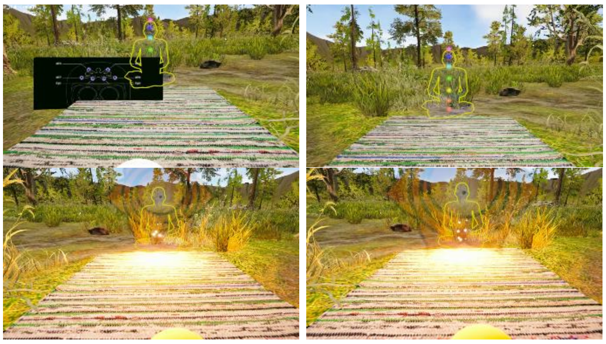


Fig. 4. The scenarios of the XR Chakra experience.

The virtual environment should minimize visual effects and create a sense of relaxation and safety. The virtual scene of XR Chakra was composed of daytime forest scenery, choosing a bright forest in the plains rather than a dense, damp black forest, and placed a meditation mat and light ball, as shown in the button images in Fig. 4. The actual meditation mat and light ball were identically mapped in the virtual environment.

As shown in Fig. 4 (Top left), at the beginning of the XR Chakra program, subjects can see the EEG device icon and the status of its contact points. When the contact points are correctly connected, they will be highlighted in purple, allowing subjects to see the wearing status in HMD. Once the device is worn correctly, the EEG device icon will automatically disappear. If there is poor contact with the EEG device or signal issues during meditation, the EEG device icon will also be displayed, and it will automatically disappear once the issue is resolved.

As shown in the top right of Fig. 4, a humanoid figure is designed in front of the user to help them understand the positions of the chakras. The current chakra position is displayed within the figure during the meditation process. By bringing the body into the scene and achieving a meditative presence, one becomes aware of being in the present moment—whether in a virtual environment or in reality. The spherical lights emit corresponding colors based on the current meditation chakra, illuminating the body with colored light as depicted in the images at the bottom of Fig. 4.

As shown in Fig. 5, the first narration is played when the user properly wears the devices. After the playback, the heart rate and brain waves determine that the user has entered a relaxed state. The subject then enters Muladhara Chakra (the first Chakra) meditation, and the second narration is played. The subject starts meditating at this time, and the system detects the subject's meditation state in real-time through brainwaves. The application can assess the level of attention and accumulate time as the subject maintains the level of attention at 60% to 80%. After the accumulated time reaches one minute, the subject can proceed to the next Chakra.

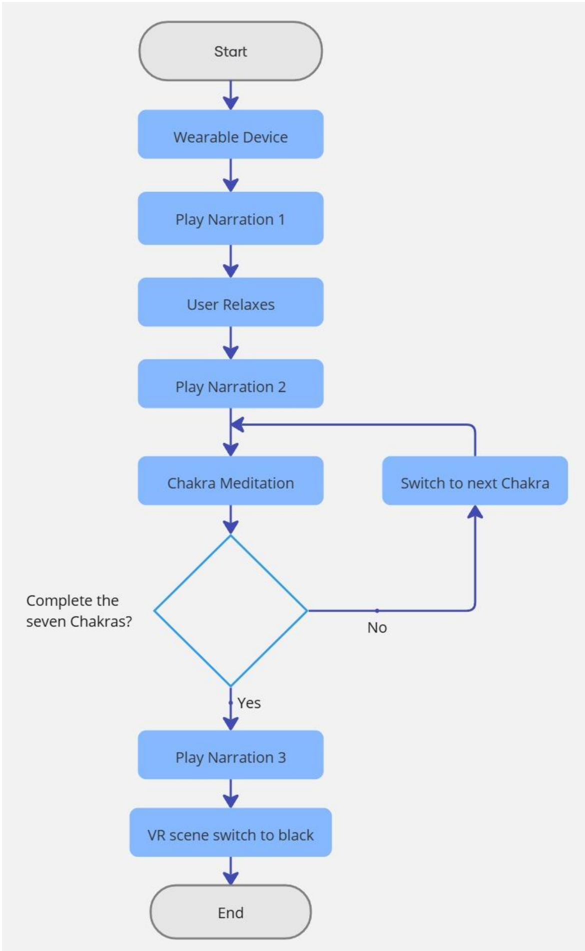


Fig. 5. Scene flow chart of the XR Chakra meditation application; [11] shows a sample screencast of one of the subjects who completed the XR Chakra meditation experience virtual scenario.

As the meditation continues, a totem representing the chakra appears on the screen. When the subject’s attention level is at the right level, the totem gradually grows in size. The subject moves to the next chakra when it reaches its maximum dimension (bottom right of Fig. 4). When distracted or not relaxed enough, the totem stops amplifying. This is an essential guide and encouragement for beginners. The third narration is played after the end of the seven-chakra meditation and switches to a black scene so the subject can close their eyes and meditate. At this time, the chakras are more unblocked and balanced, and the mood is stable. It is easy to enter a more profound meditation. In this state, it is easy to get inspiration and enlightenment, and even your wishes can come true, increasing creativity.

2.1.2 Electroencephalography(EEG)

Brain waves are recorded using the noninvasive EEG system Looxid Link device [9], which is installed together with the VIVE Pro HMD using a unique face gasket with the EEG sensors shown in Fig. 6. Looxid Link EEG electrodes were attached to the forehead. The contact area includes a frontal six-channel EEG system (F3, F4, F7, F8, Fp1, and Fp2) conforming to International System 10-10, based on [10].

Fig. 6. HTC VIVE Pro HMD, a face gasket with the EEG sensors (top) and Looxid Link EEG electrode placements(bottom).

Table 1. 19 Opinion survey questions in the second and third sections of the questionnaire.

#	Question
Q1	I got distracted by my surroundings.
Q2	I noticed loud noises around me.
Q3	I understood how to perform chakra meditation.
Q4	I was able to calm my mind.
Q5	I could focus better on chakra meditation.
Q6	The sounds made me feel calm.
Q7	The sounds helped me concentrate on meditation.
Q8	My mind tended to wander.
Q9	The visuals distracted me.
Q10	I enjoyed meditating using this method.
Q11	I would recommend that others use this method for meditation.
Q12	I felt physical discomfort (dizziness, nausea).
Q13	I felt eye discomfort (soreness, dryness, fatigue).
Q14	I felt sleepy.
Q15	It made me feel bored.
Q16	My mood has improved.
Q17	I gained new inspiration.
Q18	I wanted to close my eyes.
Q19	After meditating, I felt very comfortable.

2.1.3. Evaluation Questionnaire

The questionnaires in the study have five sections. The first section consisted of primary data about the user (age, biological gender, educational background, and eyesight, whether it is normal or corrected-to-normal) and several questions about their previous experience in meditation and using immersive media. The second and third sections aimed to evaluate the effects of the YouTube Chakra meditation video and XR Chakra meditation application, respectively. The second and third sections contained the same opinion survey questions shown in Table 1. The users evaluated these 19 opinion questions separately using a 5-level rating scale. Additionally, the fourth section assessed the impact of lights and sound on the XR Chakra meditation application using a 5-level rating scale. (Table 2). The questions in the final section focused on collecting overall experimental feedback and suggestions for further improvements, as shown in Table 3.

Table 2. The questions for evaluating the impact of lights and sound on the XR Chakra meditation.

#	Question
Q20	It feels better to have sound.
Q21	It feels better with lights.
Q22	Which method do you prefer? Sound and light are available; both are needed, or neither is required.

Table 3. The questions for overall experimental feedback and suggestions.

#	Question
Q23	Which method do you prefer?
Q24	Why do you prefer that method?
Q25	Do you have any unique feelings or experiences?
Q26	Do you think any aspects of this experiment need improvement or any other suggestions?
Q27	Are there any other comments you would like to share?

2.1.4. Compliance with Ethical Standards

The collection of EEG data requires careful examination of the ethical implications and the rights of the participants involved in the research process. Therefore, the review was asked for in advance from the Human Research Ethics Review Committee of National Cheng Kung University, and the experiment was conducted only after approval (NCKU HREC-E-112-541-2). Participants were informed (1) about the purpose of the study, (2) that they had the right to stop the experiment at any time without providing any reason, and (3) that they could stop the experiment if they felt sick or had any discomfort. All the training sessions were performed under a researcher's supervision in case of an emergency.

2.2. *Experimental Subjects*

The experimental subjects were adults aged 18 and above who did not have eye infectious diseases. Although this experiment uses the HTC VIVE Pro system in sitting posture, it is less likely to cause dizziness. However, subjects who often have dizziness symptoms are not recommended to participate. The subjects are mainly used for questionnaire statistics and recording the brainwave state of the subjects when meditating using both methods. There are 21 subjects in total, six males and 15 females. Most of them are college students (57.1%), mainly engaged in art and music creation, and most have experience using immersive media in VR HMD (76.2%). As many as 81% have meditation experience, and most of them do not meditate often. Only 17.6% of subjects have done chakra meditation, which meets the goal of this study: to create an XR chakra meditation program for beginners.

2.3. *Procedure*

The experiment was expected to last 60 minutes but lasted about 40 minutes. During the experiment, the subject needed to wear a wrist heart rate monitor and the HTC VIVE Pro HMD. The HMD contains the EEG device, and its EEG electrodes fit on the forehead, with the contact points shown in Fig. 5. The EEG and HMD were disinfected with alcohol after each use. Beforehand, inform the subjects that because this experiment requires concentration, they are advised to avoid activities that may interfere with physiological indicators 24 hours before the designated experimental time, such as drinking caffeinated beverages, taking medicine, or staying up late. In addition, to better measure brain waves and avoid wiping off cosmetics, washing the forehead before the experiment was recommended. The detailed experimental procedure is explained below.

- *Explain the Content of the Experiment.*
  The subjects were explained the consent form for experimental research participation, which included the name of the project, the purpose of the experiment, the equipment used, the conditions of the participants, the expected experimental time, possible risks and damages, research compensation, data storage, and processing methods, and finally, the experimental process.

- *Fill out the Consent Form and Pre-questionnaire.*
  The subjects completed the consent form for the experimental study, the anonymous basic information, and the pre-questionnaire.

- *Wear a Heart Rate Sensor*
  This experiment used a heart rate sensor worn on the wrist, similar to a watch, to detect the subject's heartbeat. A green light was on the subject's wrist, and no electrical stimulation was to be used. The heart rate was only used as an interactive reference in the program and was not used for recording.

- *Wearing HTC VIVE Pro HMD and EEG Device*
  Before the measurement, the subjects were asked to remove any metal objects, hair bands, and mobile phones and if they needed to go to the bathroom first to avoid the inconvenience of going out during the experiment. As shown in Fig. 5, the EEG device used in this study is connected to the HMD, so it was sufficient to wear the HMD. The EEG electrodes should be aligned with the forehead before wearing, and the virtual screen should show whether it was worn correctly.

- *Watch the Chakra Meditation YouTube Video*
  The subjects watched the YouTube Chakra meditation video to do a Chakra meditation session for 10 minutes.

- *Fill out the Questionnaire*
  The subject completed the second section of the questionnaire (Table 1), which aimed to evaluate the effects of the YouTube Chakra meditation video.

- *Use the XR Chakra Meditation Program*
  The system recorded brainwave data during use. After wearing the HMD, the subjects followed the voice guidance and used the XR Chakra meditation program for 10 minutes to complete the seven-chakra meditation. The EEG device recorded the brain's electrical activity and transmitted it to the computer to save in data files. It was a noninvasive measurement and did not send any energy or substance to the human body.

- *Complete the Questionnaire*
  The subject completed the 27-question questionnaire (Table 1, Table 2, and Table 3), including the fourth and last sections, about the impact of lights and sound on the XR Chakra meditation application and the overall comparison of both meditation methods.

3. Results and Discussion

3.1. *Questionnaire Data Analysis*

Most comparison questions have five options, from left to right: strongly disagree, disagree, neither agree nor disagree, agree, and strongly agree. This section analyzes user opinions on both meditation methods using Likert plots.

3.1.1. Inference Level

The questionnaire survey divided interference into external interference and sound and image interference, checking whether it is possible to meditate.

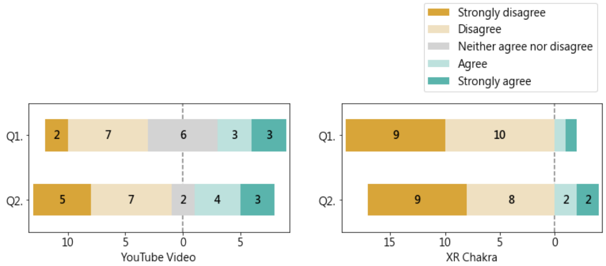


Fig. 7. Degree of external interference.

External interference: More subjects responded as they disagreed with the answers to questions Q1(got distracted by my surroundings) and Q2(noticed loud noises around me). As shown in Fig. 7, watching YouTube videos is more disturbing, but the impact is insignificant in a controlled environment. Even if we want to meditate at home, we will find a more isolated space. When using XR Chakra for meditation, it is less likely to be disturbed by surroundings.

Sound and visual interference: The results of the questionnaire questions Q6 (sounds made me feel calm), Q7 (sounds helped me concentrate on meditation), and Q9 (visuals distracted me) are shown in Fig. 8. Questions Q6 and Q7 examined whether the sound design of the YouTube video and XR Chakra application can make subjects feel calm, relaxed, and helpful. Meditation results show that all have such functions, but XR meditation programs can achieve calmness and focus on meditation.

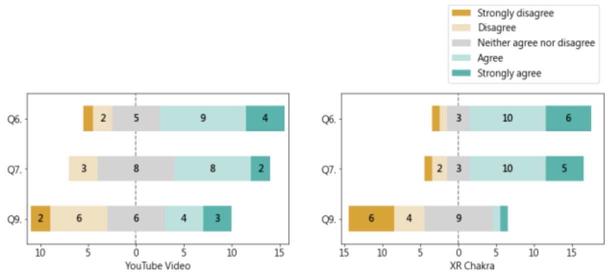


Fig. 8. Degree of interference from sound and visuals.

Question Q9 explored whether the visuals interfered with the subjects' meditation. According to the survey results of the YouTube video, 33.3% (7) of the subjects felt it would have an impact. Only 10% (2 people) thought the XR Chakra program had an effect.

Meditation difficulty level: Fig. 9 shows the results of questionnaire questions Q4 (I was able to calm my mind) and Q5 (I could focus better on chakra meditation). It indicates that the XR Chakra meditation program can help subjects calm down and concentrate on chakra meditation. Question Q8 (mind tended to wander) represents the degree of distraction. The results show that watching YouTube made subjects more likely to have their mind wander.

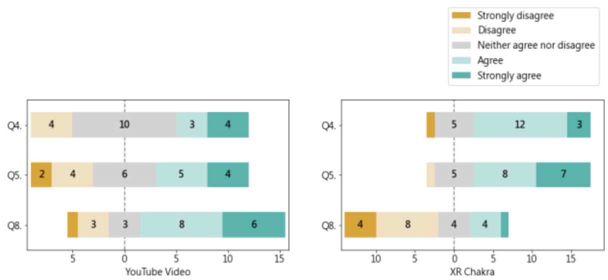


Fig. 9. The level of difficulty in meditation.

3.1.2. Ease of Learning

Question Q3 (I understood how to perform chakra meditation) attempts to determine whether the subject can use these two methods to learn how to perform chakra meditation, as shown in Fig. 10. Since YouTube videos do not have any guidelines, they only play images and chant, so the user cannot learn how to do chakra meditation.

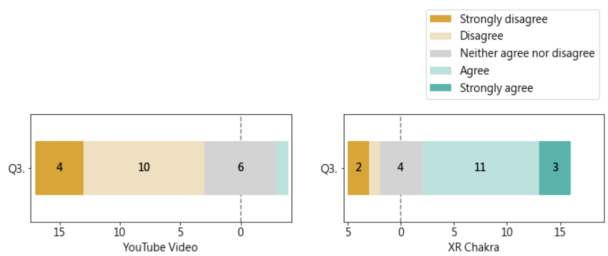


Fig. 10. The level of ease of learning chakra meditation.

However, users can learn to meditate using this XR Chakra application's simple guidance.

3.1.3. Physical Effects

Fig. 11 shows the results of questionnaire questions Q12 (I feel physical discomfort, such as dizziness or nausea) and Q13 (I feel eye discomfort, such as soreness, dryness, or fatigue). Displaying YouTube has less impact on the body, especially HMD, which can easily cause eye fatigue. Questionnaire feedback also shows that the weight of HMDs can easily cause discomfort.

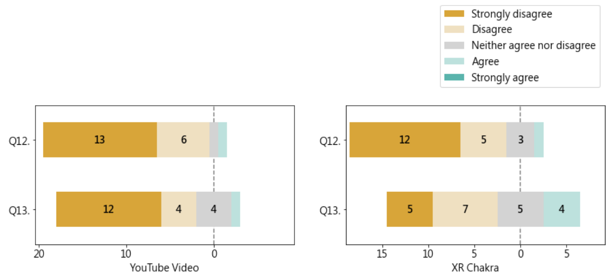


Fig. 11. The level of physical effects on the body.

3.1.4. Psychological and Emotional Effects

The questionnaire Questions Q14 to Q19 considered psychological and emotional states, including wanting to sleep, feeling bored, being in a good mood, getting new inspiration, wanting to close their eyes, and feeling comfortable. The results are shown in Fig. 12. Most subjects felt good and comfortable after meditation. There is not much difference between these two methods, but the XR Chakra meditation method is not dull.

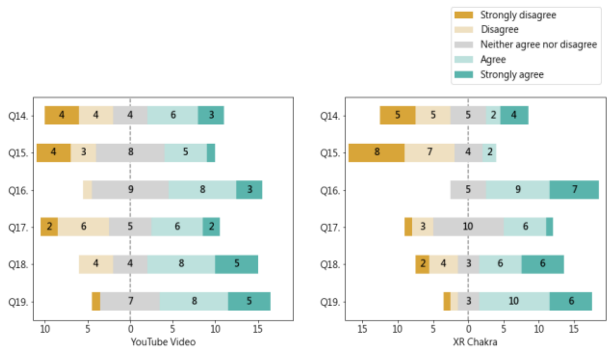


Fig. 12. The level of physical effects on the body.

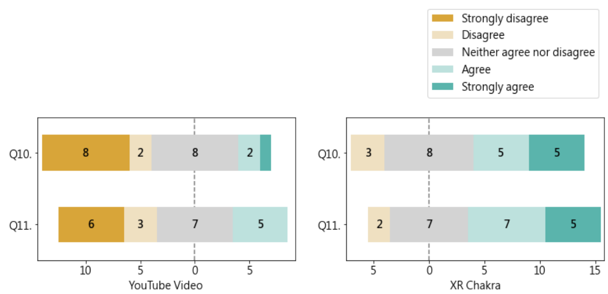


Fig. 13. Comparison of user recommendations on two methods.

3.1.5. Overall Comparison

The results of questionnaire questions Q10 (I enjoyed meditating using this method) and Q11 (I would recommend others meditate this way) are shown in Fig. 13. And Fig. 14 shows the results of Question Q23 (Which method do you prefer?), which shows that most of the subjects prefer the XR Chakra meditation program.

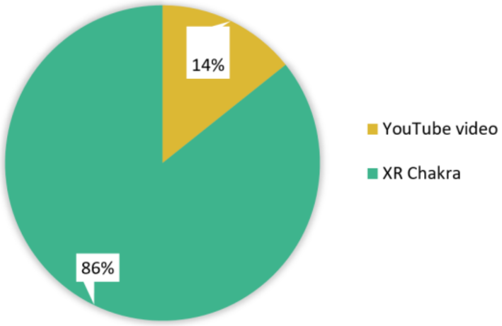


Fig. 14. Overall comparison of user preferences on both meditation methods.

3.2. *Vibrational Realism*

Fig. 15 shows the results of questionnaire questions Q20 (It feels better to have sound) and Q21 (It feels better with lights). Some subjects mentioned that sound makes it easier for them to meditate. Some subjects found that the color of light affects their mood. However, some subjects also felt that dimming or having no lights might be more relaxing.

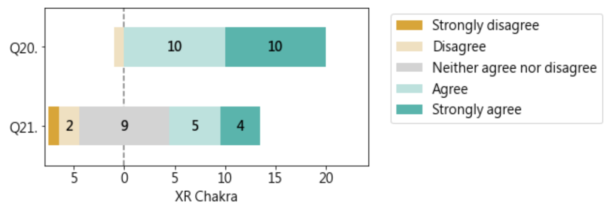


Fig. 15. The preferences for sound and lighting.

Among 52.4% of the subjects prefer to have sound, and 42.9% of the subjects want to have both. The position of the light is on the ground in front, as shown in Fig. 16. Since the design of the HMD has light leakage, the light can shine from the bottom of the HMD to the eyes, causing eye discomfort. In the future, we can design lights at the top and one at the front so that the front and rear chakras and the crown chakra can be illuminated simultaneously without the problem of illuminating the eyes.

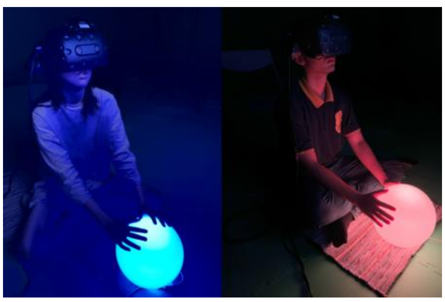


Fig. 16. The position of the light ball.

3.3. *Feedback on Contents of the XR Chakra*

Based on the responses to questions Q24 to Q27, some subjects said they were very excited when they saw the totem getting more extensive and then stopping. We also considered the dynamic state of the totem when designing its motion. On the one hand, we hope that users can quickly know their meditation status, but they do not want it to cause interference. Some test subjects said that they felt perfect when seeing the forest. Virtual content was also researched during the production of the XR Chakra program, which is a suitable virtual scenario that would make users feel safe and secure enough to meditate [7]. In addition to forests, users suggested and wanted to test the future scenes of the seaside, mountains, etc. In addition, some subjects thought that after using the XR chakra meditation program, they felt calmer and brought a sense of freshness.

3.4. *Brainwave Analysis*

This experiment aimed to determine whether the XR Chakra meditation program can help subjects more easily enter a state of chakra meditation. Therefore, the experiment sessions recorded the subjects' brainwaves while watching the YouTube chakra meditation video and used our XR Chakra meditation program. The attention level obtained from brainwave data was divided into one-minute intervals, three to six minutes, and the last minute for comparison and analysis. The median value was calculated and depicted in a box plot, as shown in Fig. 17.

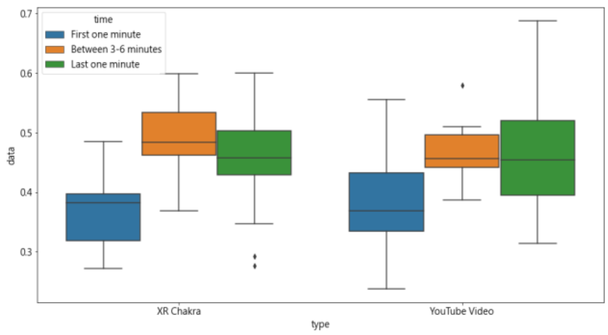


Fig. 17. Comparison of attention between two methods.

At 3-6 minutes, the subjects entered a meditative state, and the attention level was higher. This was even more obvious when using the XR Chakra meditation program. At the last minute, the subject may feel tired, you can no longer concentrate, and become more confused.

4. Conclusion

Based on Chakra meditation, a chakra meditation program was developed in an XR environment using passive interactive technology based on behavioral biometrics. Brainwaves and heart rate were used as passive interaction methods. The user does not need to use gestures or controllers, and the system executes automatically. This research also introduces basic vibration reality into the system through sound waves and light wave vibrations to assist the human body's chakras enter a balanced state.

After comparing watching a YouTube chakra meditation video with using the XR Chakra meditation program developed by this research, it was found that the XR Chakra program is better than YouTube video in terms of interference level, ease of learning, psychological and emotional impact, and physical impact. However, the subjects generally preferred the XR Chakra meditation method. At present, the weight of the HMD is a problem that is difficult to ignore for meditation. However, with the advancement of technology, it can be solved one day. The VIVE Pro weighs about 800 grams, while the latest Bigscreen Beyond HMD is only 127 grams [12], making meditation more comfortable. Regarding the vibration reality, most respondents liked the sound design, while only 42.9% wanted the lighting design. Vibrational reality is an emerging field of study that utilizes various technologies, such as ultrasound, to be developed and integrated into future versions.

Moreover, for the design of the XR Chakra meditation program, it is necessary to complete the seven chakras and then continue to meditate to gain inspiration. Based on the current estimate that most subjects entered the 4th chakra within 10 minutes, it takes approximately 20 minutes. Further research and experiments can be done in the future. In brain wave experiments, this XR Chakra meditation program can keep subjects in a meditative state better than YouTube videos, but not significantly. Perhaps longer, long-term experiments are needed to make a significant difference. Experimental results show that meditation has a positive impact on emotions. For the general public and artistic creators, regular meditation will help stabilize emotions, increase awareness and empathy, and make people more focused. These characteristics can help individuals grow and achieve success.

Acknowledgments

The authors express gratitude to the participants for their time and effort in this research project's experimental sessions.

References

1. R. M. Kelly, E. M. Seabrook, F. Foley, N. Thomas, M. Nedeljkovic, and G. Wadley, "Design considerations for supporting mindfulness in virtual reality," Frontiers in Virtual Reality, Sec. Virtual Reality in Medicine, Vol. 2, 19 pages, 2022, https://doi.org/10.3389/frvir.2021.672556
2. BreatheVR, "Breathevr: A healthy mind has an easy breath.," 2021, https://www.discoverneon.com/breathevr". Accessed: 2024-12-29
3. K. Shrikumar, "Playne: The Meditation Game.," https://www.playne.co. Accessed: 2024-12-29.
4. J. Carline, "Nature Treks VR", 21st May 2019. https://www.oculus.com/experiences/quest/2616537008386430. Accessed: 2024-12-29.
5. D. G. White, Editor, Yoga in Practice. Princeton University Press, 2012. ISBN 10: 8121512727, JSTOR, https://doi.org/10.2307/j.ctvcm4gpf.
6. J. Carmody and R. A. Baer, "Relationships between mindfulness practice and levels of mindfulness medical and psychological symptoms and well-being in a mindfulness-based stress reduction program," Journal of Behavioral Medicine, vol. 31, no. 1, pp. 23-33, 2008.
7. C. -T. Lin and R. P. C. Janaka Rajapakse, "Development of a Chakra Meditation Application of Virtual and Vibrational Realities Based on Behavioral Biometric," 2023 International Conference on Cyberworlds (CW), Sousse, Tunisia, 2023, pp. 290-297, doi: 10.1109/CW58918.2023.00052
8. Meditative Mind, Quick 7 chakra cleansing | 3 minutes per chakra | seed mantra chanting meditation | root to crown. Available:https://youtu.be/NmAHY_tg9Es?si=KNviN7LTBq1A69_8.
9. Looxidlabs.com, "Looxid Link.", https://looxidlabs.com
10. V. Jurcak, D. Tsuzuki, I. Dan, "10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems," NeuroImage, Vol. 34, no. 4, 2007, pp. 1600-1611, ISSN 1053-8119, https://doi.org/10.1016/j.neuroimage.2006.09.024.
11. The virtual scenarios of the XR Chakra experience, https://youtu.be/QNKEqyvIsaw
12. Bigscreen. Bigscreen beyond. Accessed: 2024-05-13, Available: https://www.bigscreenvr.com.

Authors Introduction

Ms. P.I.A. Gayathri Bimba

She received her MBA in Marketing in 2021 from Cardiff Metropolitan University in the UK. She is currently a research student at the School of Knowledge Science in Japan Advanced Institute of Science and Technology. Her research interests include new media, interactive art, therapeutic content, art therapy, user studies, and digital marketing.

Dr. Chien-Tung Lin

He received his B.S. degree from the Department of Computer Science and Information Management, Providence University, Taiwan, 1999 and his MFA degree from the Graduate Institute of Animation and Film Art, Tainan National University of the Arts, Taiwan. His research interests include virtual reality, interactive art, EEG, art therapy, computer animation, and healing methods.

Dr. Janaka Rajapakse

R.P.C. Janaka Rajapakse is an Associate Professor at the Graduate Institute of Animation and Film Art, Tainan National University of the Arts, Taiwan. He received his Ph.D. in Knowledge System Science from the Japan Advanced Institute of Science and Technology in 2008. His research interests include computer animation, motion capture, VR/AR/MR, haptic interfaces, AI, computer graphics, and Kansei Engineering. He is a member of the Motion Capture Society, The Society for Art and Science, ASIAGRAPH, and SIG-Design Creativity.

Dr. Kazunori Miyata

He is a Professor at the School of Knowledge Science in Japan Advanced Institute of Science and Technology. He received a PhD degree (Engineering) from Tokyo Institute of Technology, Japan, in 1996. His research interests include computer graphics, media art, and multimedia applications. He is the 4th president of the Society for Art and Science and a member of ACM, IEEE, IPSJ, IEICE, and others.

ThoughtDiffusion: An Interactive Installation for Exploring Neuro-Art from EEG Data with Stable Diffusion Models

Janaka Rajapakse
Tainan National University of the Arts, No. 66, Daqi, Guantian Dist., Tainan, 720005, Taiwan
Email: janaka@mail.tnnua.edu.tw

Abstract

ThoughtDiffusion is an interactive installation that combines cognitive processes with the generative capabilities of AI to generate artistic images in real time. The installation system uses Stable Diffusion models and state-of-the-art neural decoding techniques that allow the mapping of brainwave patterns into coherent visual representations. The installation is based on a non-invasive commodity EEG headset that records users' brain signals fed into a stable diffusion model to output images corresponding to the intended mental state but unique to the participants. This installation uses a Kinect V2 sensor to capture users' body movements, which advance multimodal interaction significantly relating to relaxation, the state of being calm, and the state of attention.

Keywords: passive interaction, Neuto-art, Stable Diffusion, TouchDesigner, art interaction

1. Introduction

Neuroart is a growing field that combines art and neuroscience. It transforms neural activity into visual expressions, providing a unique perspective on the human mind. The emergence of the dissemination of thought represents a significant intersection between technology and creativity, nourishing an environment where artistic expression is transformed by ideas gleaned from neurological data. In addition, integrating neurotechnology into artistic practices generates discussions on authenticity and the property of creative expression. The question arises: To what extent does technology influence art production? While artists use brain-computer interface (BCI) technology to evoke specific artistic results, they engage in a crucial dialogue between intentionality and inadvertent expressions generated by cognitive processes [1], [2], [3], [4]. This duality can enrich artistic efforts, in which the interaction between brain activity and technological influence leads to innovative forms of creative expression [5].

In the scope of this study, the thought-of thought appears as a revolutionary installation that transcends traditional borders by exploiting electroencephalography (EEG) data to forge visual expressions in real time. This innovative company not only represents a convergence of technological and artistic disciplines but also facilitates an enriching interaction between users and works of art, ultimately improving their emotional experiences. In conjunction with EEG data, ThoughtDiffusion installation uses the capabilities of the Kinect-V2 sensor and the RGB camera to collect relevant user data, which improves the global interactive experience. The Kinect-V2 facilitates real-time movement monitoring, allowing users to engage with installation beyond simple passive observation [6]. The system captures the spatial dynamics of user movements, creating a reactive environment in which visual expressions dynamically correspond to cognitive states and physical interactions.

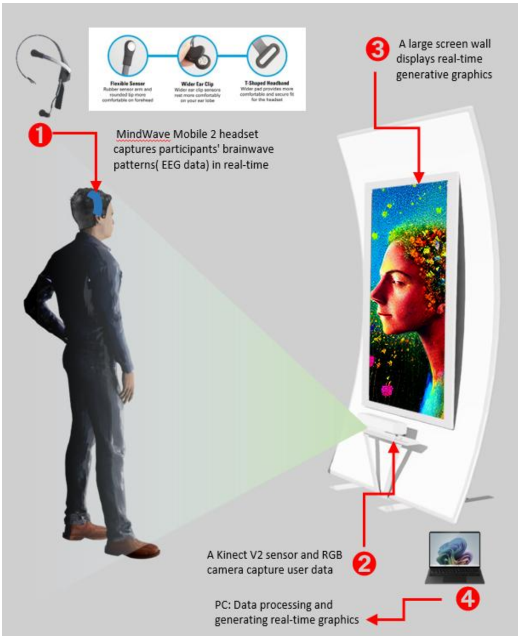


Fig. 1. ThoughtDiffusion Interactive Installation.

The evolution of diffusion models represents significant changes in the scenario of interactive art creation and was accompanied by the emergence of new applications that leverage its general approach [7]. As researchers and professionals navigate the interrelationships between diffusion methodologies and rapid strategies, the robust dialogue will undoubtedly promote future innovations in interactive art [8]. This installation's key aspect is its noninvasive approach. It uses a NeuroSky MindWave Mobile-2 EEG headset [9] to record users' brain signals, which are processed through the Stable Diffusion(SD) models to create unique visual expressions that correspond to the users' intended mental states. This installation proposes a scenario-oriented prompt structure that the user can feed multiple prompts to interact passively to regenerate the thought of thought experience in real-time.

2. Related Works

2.1. *Interactive Art and Diffusion Models*

Recent progress in artificial intelligence has generated a new paradigm in the kingdom of interactive art, in particular through the advent of diffusion models. These models, which create data distributions capable of producing complex and different results, have basically transformed artistic expression, improving multimodal interactions and deepening human involvement [10]. These probabilistic approaches allow the generation of high-dimensional outputs with unprecedented loyalty and varieties, which is particularly relevant in the artistic contexts in which the visual, auditory, and tactile modalities intersect. By facilitating the creation of works of art that can dynamically adapt to user interactions through different sensory ways, artists have the power to invent new forms of involvement [11]. The multi-sensory experience not only fascinates participants but also promotes a deeper emotional connection between the public and art, thus increasing the commitment [12]. In addition to redefining the artistic process, these models also push innovative paths for the involvement of the public. Interactive art that uses diffusion models can be experienced in various ways, inviting a broader audience to commit. This potential for customization allows individual experiences to differ according to the interactions, preferences, and unique answers of a public member.

Since the diffusion models generate interactive experiences of art and influence, the need for a critical discourse on the implications of the machine's involvement in creativity is presented [13]. Although these models can improve aspects of artistic expression, dialogue must also understand the complexities of how human artists navigate their roles within an increasingly automated landscape [14]. In conclusion, new diffusion models have started a transformative movement within interactive art, improving multimodal interactions and enriching human participation, simultaneously unlocking new realities of artistic expression and public involvement. While artists adopt these technologies, they create a more collaborative and engaging experience that redefines the boundaries of artistic creation. ThoughtDiffusion integrates technologies like EEG monitoring and depth sensors, facilitating the development of a creative landscape in which interactivity is deeply rooted in the interaction between human cognition and artistic representation.

2.2. *Prompt Engineering and Generative Expression in Art*

The evolution of rapid engineering in interactive art represents a critical intersection of technology, creativity, and artistic expression. As generative artificial intelligence tools become increasingly integrated into the creative process, the improvement of instructions and the integration of multiple instructions have emerged as vital components to improve artistic ideas [15]. This literature revision explores recent progress in this sector, underlining the meaning of evolving practices in rapid engineering. Rapid engineering, the practice of designing and refining inputs to arouse desired results from AI systems, has witnessed substantial developments that facilitate a deeper exploration of creative concepts. In addition, rapid design's role extends beyond simple refinement; It includes integrating multiple suggestions to improve creative processes. Recent investigations have examined the possibilities that support using different suggestions in artistic education [8].

Prompt-based rapid design practice promotes iterative processes that allow emerging students and artists to unlock their creative potential. Combining various instructive suggestions allows the creators to explore multiple dimensions of a singular artistic idea, thus promoting a richer and more varied creative output [15]. Mizrahi [17] underlines that the mastery of rapid engineering can accelerate the transition of an artist from a beginner to a competent creator, observing that the "*language of creativity*" plays a fundamental role in the generative process. By providing educational paintings focused on the art of creating engaging instructions, instructors can allow students to exploit the generative tools of artificial intelligence more effectively [16]. Artists can improve their ideation and expression through the iterative improvement of generative instructions and the integration of multiple suggestions. This aspect of artistic practice not only expands the creative potential of individuals but also underlines the educational importance of understanding and mastery of prompt engineering as a critical element of contemporary art production. Therefore, the continuous exploration in this domain benefits from both the artistic community and the broader field of creative education. ThoughtDiffusion installation provides a platform for redefining artistic scenarios through users' multiple prompts, allowing for adjustments to the artwork using their physical and cognitive inputs in real time; these models make the creative process more accessible to explore new forms of interaction with more human involvements.

3. System Overview

The system of installation consists of two aspects: hardware and software, as shown in Fig. 2. As Hardware devices were used, both movement data from Kinect-V2 and brainwave data from the MindWave Mobile 2 EEG

Headset reach a multimodal interaction that enriches users' involvement. The Kinect-V2 RGB camera was used as a sensor to detect the user's skeleton data, depth, and silhouette. EEG headset was introduced to capture users' brainwave data in real-time; the device mechanism allows the system to record EEG frequency spectrum (alpha, beta, etc.) and proprietary *eSense* indicators for the level of *attention* and *relaxation* through EEG data. The development environment was Intel i9 PC on Windows 10 Pro system, with the NVIDIA GeForce RTX 2080Ti graphics card.

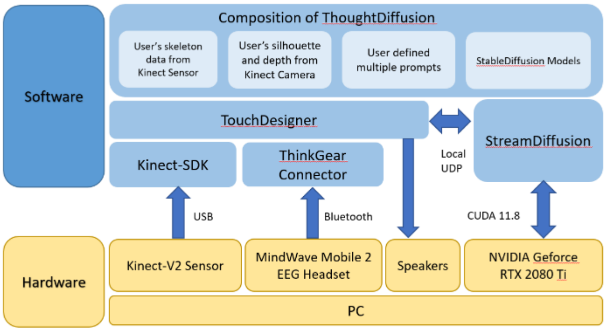


Fig. 2. System Overview

3.1. *Hardware*

The installation is equipped with a large screen wall to display real-time generated graphics, as shown in Fig. 3. The Kinect-V2 sensor is mounted under the large display, which can easily capture the user's full-body skeleton data. Kinect has two built-in sensors: an RGB color video camera and a depth sensor [6], which helps to capture the user's silhouette and depth for video-to-video generation in StreamDiffusion [19] model in real-time. Fig. 3 depicts the installation space. The x-axis corresponds to the horizontal direction across the sensor's field of view, the y-axis represents the vertical direction in the sensor's field of view, and the z-axis represents the normal of the sensor, which can detect depth information, indicating how far away the user is from the sensor.

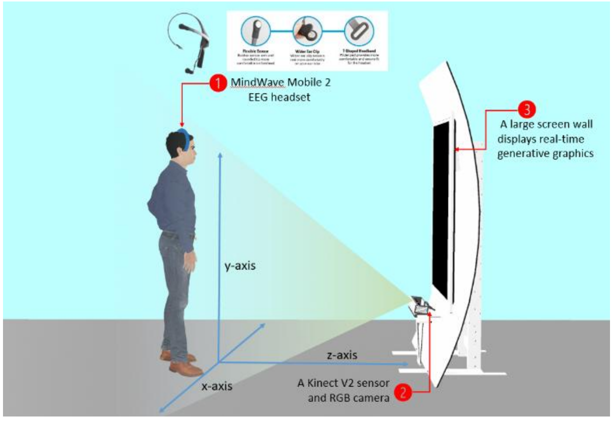


Fig. 3. Affected by pedestrians crossing

This installation mainly uses the functions of Kinect to perform the following three operations:

- Calculate the user's relative position by detecting values of the user's hip-joint coordinates, which control the step counts of the denoise schedule.
- Measure the displacement of the user's hip joint along the x-axis to control the weight of multiple prompts.
- The camera captures the user's silhouette for video-to-video interactive generation in the StreamDiffusion module.

While Kinect captures users' physical movements, Bluetooth connects the user-worn MindWave Mobile 2 EEG headset to the system, which allows the capture of brainwaves for passive interaction.

3.2. *Software*

The system was developed with TouchDesigner [18], a node-based visual programming language for the creation of real-time interactive media content. The Kinect SDK allows integration with TouchDesigner and its built-in channel operator, *Kinect CHOP*, which reads positional and skeletal tracking data. The built-in texture operator, *Kinect TOP*, captures data from the Kinect depth camera and RGB color camera, which is used to compute and isolate the user's silhouette.

The installed NeuroSky's Windows Developer Tools 3.2 provides ThinkGear Connector, which lets TouchDesigner transfer data from the MindWave Mobile 2 EEG headset through a *TCP/IP DAT*(data operator node) socket server (localhost with port 13854) [20]. Bluetooth automatically connects to the device whenever the system needs to read the data from the EEG headset. The recorded serial data is converted to a table (*TouchDesigner Table DAT*), which quickly reads the data into the separate channels for focus levels: *attention* and *relaxation*, and EEG data values: *delta*, *Theta*, *lowAlpha*, *highAlpha*, *lowBeta*, *highBeta*, *lowGamma*, *highGamma*.

The TouchDesigner adopted the submodule StreamDiffusion-TD [19], which is installed with Python 3.6, CUDA 11.8, and the Git desktop client. This submodule supports generating real-time generative AI content based on text2img and img2img. The measured levels of attention and relaxation in EEG data control the *seed* and *seed weight* of StreamDiffusion-TD in real-time. Using multiple Text COMP nodes, set up the multiple Prompt Boxes to read user-given prompts, which can be controlled by users' physical displacement along the x-axis determined by the Kinect CHOPs.

4. Results of the Interaction

The system promotes different modes of interaction.

1. ***Language of creativity***: The user can add their ideation and expression through multiple prompts, interact while moving along the horizontal direction(x-axis), pay attention to their input prompts, and iteratively improve generative prompts with multiple suggestions. Not only side-by-side comparison, but users can also move towards the display (along the z-axis, the normal of the Kinect sensor).

2. ***Pose for prompts***: Users (multiple users up to 3) can pose to the Kinect Camera while paying attention to the added input prompt. The system captures the user's poses for video-to-video interactive generation in the StreamDiffusion module. This scenario-oriented practice not only expands individuals' creative potential but also provides a more collaborative and engaging experience that redefines the boundaries of interactive creation.

Model: txt2img (sd-turbo)

z-distance from the display (denoise schedule)	Prompt: *Gandalf from Lord of the Rings, diffuse lighting, fantasy, intricate elegant highly, detailed lifelike photorealistic digital painting, art station.*	Prompt: *a perfect bonsai.*
0.5m (4 steps)		
1.5m (8 steps)		
2.0m (12 steps)		

Fig. 4. Creative expressions and physical engagement of the body (a sample Moving towards the display(along the z-axis, normal of the Kinect sensor).

4.1. *Creative Expressions*

The ThoughtDiffusion installation offers users an opportunity for an unprecedented transformation in the interaction between creative expression and physical engagement of the body. These modes of interaction, characterized by their multifaceted methods, support a complex interplay of language, user input, and bodily movements in iterative feedback, which collectively fosters richer and more dynamic creative processes. This exploration deepens these interaction ways, examining how they improve the user's experience and promote collaborative creativity. As shown in Fig. 4, language is a fundamental interaction element at the center of the involvement process. When the user moves closer to the display, the number of steps of the *denoise schedule* in StreamDiffusion gradually decreases. Incorporating physical body movements not only modifies the generative content on display but also enhances collective intelligence, as the intuitions lead to richer and more nuanced interactions that modify the prompt and passively adjust the body position to achieve creative results.

Model: img2img (sd-turbo)

Prompt: *A Boy with super powers and curly and brown hair, illustration concept art anime key visual trending pixiv fanbox by wlop and greg rutkowski and makoto shinkai and studio ghibli*

z-distance from (denoise schedule)	input	output
2.0m (12 steps)		
3.5m (16 steps)		

Fig. 5. Posing for the prompts (video-to-video interactive generation).

Fig. 5 illustrates how users can respond to a given prompt. The dynamic nature of creative processes requires a structure, such as *a method of acting or posing*, in which users can perfect their ideas. This iterative approach promotes an experimentation environment in which users are encouraged to take creative risks, knowing they can adjust their production in response to the evolving collaborative scene.

4.2. *Level of Attention and Relaxation*

Neurological feedback on attention and relaxation levels is another important aspect of this installation's complexity interaction. While the levels of attention and relaxation cannot be directly associated with the meaning of seed and seed weight parameters of the StreamDiffusionTD model, our approach attempted to map and trigger their values and intuitively demonstrate the capability of EEG signals to reconstruct dynamic visual perceptions when the user keeps the attention on the given prompt. As shown in Fig. 6, different levels of attention can significantly modulate to generate real-time output using the Stable Diffusion model *txt2img (sd15)* for a given prompt: "*An elegant girl in an urban outfit, cute fine face, round eyes, digital painting, glasses, blonde hair, fashion magazine*." The different levels of relaxation experimented with the Stable diffusion model *img2img (sd15)* for a given prompt: "*A boy in a village outfit, cute fine face, round eyes, digital painting, face paint, curly hair, art magazine*."(Fig. 7). As a new form of human-computer interaction, the ability to effortlessly adjust users' physical postures to achieve desired outcomes in real-time has become a major focus in training attention and relaxation practices, which can also be considered art therapy.

Model: txt2img (sd15)

Prompt: *An elegant girl in an urban outfit, cute fine face, round eyes, digital painting, glasses, blonde hair, fashion magazine*

Attention: 50 Attention: 60 Attention: 65 Attention: 80

Fig. 6. Different levels of attention.

Model: img2img (sd15)

Prompt: *A boy in a village outfit, cute fine face, round eyes, digital painting, face paint, curly hair, art magazine.*

Relaxation: 60 Relaxation: 70 Relaxation:78 Relaxation: 85

Fig. 7. Different levels of relaxation.

5. Conclusion

As users engage in this vibrant form of human-computer interaction, user gestures, and neuro-feedback mechanisms generate new content and co-evolving their creative identities. The interaction of distinct prospects and experiences occurs in the form of collaborative creation richer in the sum of its parts. The installation acts as a co-creator, offering instructions and possibilities that use users to think in addition to their initial ideas and hypotheses. This coevolution of thought and creativity cultivates a sense of collective property of creative production, thus improving the intrinsic motivation to collaborate further.

In summary, this installation system's multifaceted interaction method significantly enriches creative expression and collaboration. By exploiting the dynamics of language, user poses, and iterative neurological feedback, this system creates a rich space for innovative ideas and collective creativity. Through the development of this installation, it is clear that the evolution of technology continues to transform the nature of creative collaboration, allowing users to interact with each other and their creative processes in profoundly transformative ways.

References

1. E. M. Holz, L. Botrel L, and A. Ku¨bler, "Independent home use of Brain Painting improves the quality of life of two artists in the locked-in state diagnosed with amyotrophic lateral sclerosis." Brain-Computer Interfaces, Vol. 2, no. 2–3, pp.117–34, 2015.
2. C-T. Lin, R.P.C.J. Rajapakse, and Y. Tokuyama, "NeuroPhyllotaxis: An Interactive Application for Generative Art Based on EEG Data" Proc. of the International Conference on Artificial Life and Robotics 2021, pp. 724-727, 2021, https://doi.org/10.5954/ICAROB.2021.OS20-1
3. R.P.C.J. Rajapakse, and Y. Tokuyama, Y. (2021, January 21–24). Thoughtmix: Interactive watercolor generation and mixing based on EEG data. Proc. of the International Conference on Artificial Life and Robotics 2021, pp. 728-731, https://doi.org/10.5954/ICAROB.2021.OS20-3
4. C-T. Lin, R.P.C.J. Rajapakse, and Y. Tokuyama, "Development of EEG Data-driven Generative Art Application for Real-time and Dynamic Interaction" Journal of Robotics, Networking and Artificial Life, vol. 8, no. 2, pp. 117-121, -727, 2021, https://doi.org/10.2991/jrnal.k.210713.010
5. K. Yamashiro, N. Matsumoto, and Y. Ikegaya Y, "Diffusion model-based image generation from rat brain activity." PLOS ONE, vol.19, no. 9,: doi: https://doi.org/10.1371/journal.pone.0309709
6. "Kinect for Windows", https://learn.microsoft.com/zh-tw/windows/apps/design/devices/kinect-for-windows
7. R. Po, W. Yifan, V. Golyanik, K. Aberman, J. T. Barron, A. Bermano, E. Chan, T. Dekel, A. Holynski, A. Kanazawa, C.K. Liu, L. Liu, B. Mildenhall, M. Nießner, B. Ommer, C. Theobalt, P. Wonka, G. Wetzstein, "State of the Art on Diffusion Models for Visual Computing," Computer Graphics Forum, Vol. 43, No. 2, : e15063, 2024, https://doi.org/10.1111/cgf.15063
8. J. Hutson, and P. Cotroneo, "Generative AI Tools in Art Education: Exploring Prompt Engineering and Iterative Processes for Enhanced Creativity," Metaverse, vol. 4, no. 1, pp. 1-14, 2023, https://doi.org/10.54517/m.v4i1.2164
9. NeuroSky MindWave Mobile-2 EEG Headset https://store.neurosky.com/pages/mindwave
10. V. Boutin, T. Fel, L. Singhal, R. Mukherji, A. Nagaraj, J. Colin, and T. Serre, "Diffusion Models as Artists: Are we Closing the Gap between Humans and Machines?" https://doi.org/10.48550/arXiv.2301.11722
11. Georgios Marentakis, David Pirrò, and Marian Weger. 2017. Creative Evaluation. In Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17). Association for Computing Machinery, New York, NY, USA, 853–864. https://doi.org/10.1145/3064663.3064710

12. Claudia Rébola Winegarden, Nicholas Komor, and Scott M. Gilliland. 2010. Silent mutations: physical-digital interactions in spaces. In Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction (TEI '10). Associationn for Computing Machinery, New York, NY, USA, 277–278.
13. B. Wang, Q. Chen, and Z. Wang, "Diffusion-Based Visual Art Creation: A Survey and New Perspectives," arXiv:2408.12128, 2024. 4, 2024.
14. H. Cao, T. Cheng, Z. Gao, Y. Xu, G. Chen, P-A. Heng and Z. Li. Stan, "A Survey on Generative Diffusion Models." IEEE Transactions on Knowledge and Data Engineering, vol. 36, pp. 2814-2830, 2022, https://doi.org/10.48550/arXiv.2209.02646
15. J. Hutson and B. Robertson, "Exploring the educational potential of AI generative art in 3D design fundamentals: A case study on prompt engineering and creative workflows," Global Journal of HUMAN-SOCIAL SCIENCE: A Arts & Humanities-Psychology, Vol. 23, No. 2,2023,https://socialscienceresearch.org/index.php/GJHSS/article/view/103669/37415
16. J. Oppenlaender, R. Linder, and J. Silvennoinen, "Prompting AI art: An investigation into the creative skill of prompt engineering," International Journal of Human-Computer Interaction, pp. 1-23, November 2024. https://doi.org/10.1080/10447318.2024.2431761
17. G. Mizrahi and D. Serfaty. "Unlocking the secrets of prompt engineering: Master the art of creative language generation to accelerate your journey from novice to pro," Packt Publishing; 1st edition, January 12, 2024.
18. TouchDesigner, https://derivative.ca/, Accessed on 18th December 2024.
19. K. Kodaira, C. Xu, T. Hazama, T. Yoshimoto, K. Ohno, S. Mitsuhori, S. Sugano, H.g Cho, Z. Liu, and K. Keutzer, "Streamdiffusion: A pipeline-level solution for real-time interactive generation." arXiv preprint arXiv:2312.12491 (2023).
20. NeuroSky's Windows Developer Tools 3.2 and ThinkGear Connector, https://store.neurosky.com/products/pc-developer-tools

Authors Introduction

Dr. Janaka Rajapakse

R.P.C. Janaka Rajapakse is an Associate Professor at the Graduate Institute of Animation and Film Art, Tainan National University of the Arts, Taiwan. He received his Ph.D. in Knowledge System Science from the Japan Advanced Institute of Science and Technology in 2008. His research interests include computer animation, motion capture, VR/AR/MR, haptic interfaces, AI, computer graphics, and Kansei Engineering. He is a member of the Motion Capture Society, The Society for Art and Science, ASIAGRAPH, and SIG-Design Creativity.

PassBy2: Passive Interaction through the Pedestrian Counts and Real-time Weather Information

Chung Chien-Lin
Tainan National University of the Arts, No. 66, Daqi, Guantian Dist., Tainan, 720005, Taiwan

Janaka Rajapakse
Tainan National University of the Arts, No. 66, Daqi, Guantian Dist., Tainan, 720005, Taiwan
Email: 1qazjkj23@gmail.com, janaka@mail.tnnua.edu.tw

Abstract

A proverb in my country says, "Constant dropping wears the stone. " it means even something as small as a drop of water can cut through a stone after a long period of accumulation. It expresses the idea that the small changes we make unconsciously in our daily lives can leave unique traces over time. This project focuses on the impact of people on their surroundings, which concerns a contemporary street scene projected in an open space and captures pedestrians' pass-by counts using a Kinect sensor. Based on the measured counts of pedestrians passing by and real-time weather information, the developed application controls the color of the street scene and gradually decreases as the pedestrians cross. Using the L-system generative algorithm, the trees in the street scene grow progressively as the pedestrians cross.

Keywords: passive interaction, pedestrian counts, L-system, TouchDesigner, art interaction

1. Introduction

The current definition of public art is very vague. It can be government-employed, commercial, voluntary, or even illegal. There are also many creative methods, such as sculpture, dance, lighting, graffiti, etc. Even the recent emergence of digital art, which uses projection technology and expanded reality, etc. However, public art mainly has the common characteristic of being designed for a public audience. However, public art still has some commonalities, such as art forms installed in public spaces and available for public viewing or participation.

With the advancement of science and technology, public art has begun to be created using digital art methods and is presented using technologies such as extended reality and mixed reality. Digital art can use animation to tell stories and show more dynamics than other media. Pass-by uses digital art to combine virtual space with real space through projection and uses digital content to create interactive art [1].

This project, "Pass-by," is a passive interactive public art. It focuses on using passive interactive methods to present elements in life that are easily overlooked, such as the intersections we pass on our way out of our homes or the street trees we see everywhere on the roadside. No matter how small the influence is, it will leave its mark. Even just passing by can have an impact.

In the first version, the visual composition chose the street scene as the primary visual element. A computer-generated tree image is placed in the center of the street (Fig. 1).

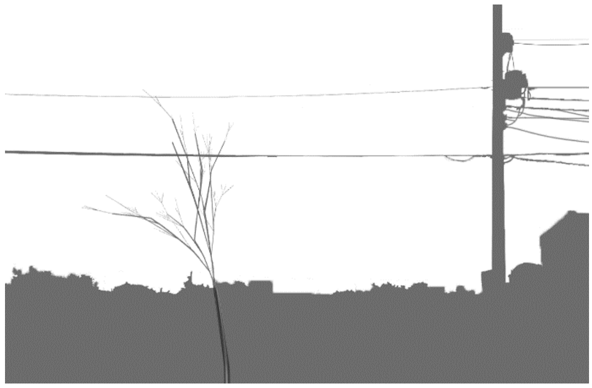


Fig. 1 Streetscape and tree generated from Pass-By [1]

The tree gradually grows through the interaction between the viewer and the work. The viewer only needs to pass by the work during the interaction process. The distance sensors are used to capture the conditions of people passing by, which generates different effects on the growth of the trees. At the same time, the shadows of pedestrians also blend with the street scene due to the light.

After exhibiting the first version, the proposed new version, “PassBy2,” combines the concept of "Site-specific art" and considering location factors. This paper focuses on how local weather affects streetscape and uses a Kinect sensor to explore the accurate behavior of pedestrians. The proposed new version stores the generated image when the exhibition is closed daily. The changes in the streetscape and the growth of trees caused by pedestrians crossing each day will create a unique composition of generated art, just like people working together to create and record their own streets.

2. Related work

The section discusses prior works in two subsections: 1) the technology and development of interactive art and 2) the concept of "PassBy2".

2.1. *Interactive art*

The distinction between passive art and interactive art lies primarily in the audience's engagement level. Passive art invites viewers to appreciate the artwork without altering it, while interactive art encourages active participation, transforming the viewer into a co-creator [2].

Passive art involves artworks that are meant to be observed. That is, in traditional static works, such as Banksy's Minnie Mouse [3], the audience is not involved in the creation of the work. Compared with conventional passive art, interactive art allows the audience to play an important role. The artwork responds to the viewer's size, movement, and even sound.

In contrast, interactive art incorporates technology to foster engagement, allowing viewers to manipulate or influence the artwork. This can range from simple interactions to complex systems where audience input alters the artistic output [4].

Most of the interactive arts focus on timely and active interaction. For instance, "Plane White" by Carina Ow captures viewer's gestures to allow viewers to actively participate in interaction and produce image changes [5]. Or focus on the current changes such as "Shiny Ball Mirror" by Daniel Rozin, which uses a metal ball to create two reflections in front of the viewer.

With the advancement of technology, interactive art has begun to use electronic sensors to create creations, such as Arduino, Kinect, etc. Interactive art, enhanced by technology, fosters deeper engagement and aesthetic appreciation among audiences [6]. Like "Rain Room" by Random International, which uses motion sensors [7], the interactive art installation detects the presence and movement of individuals, creating a zone where the rain stops above them, providing an uncanny experience of controlling the rain and walking through the rain without getting wet. There are also public places, like "The Tree of Resonating Colors of Life" by TeamLab, that change color due to the impact of someone's blow [8], and a color tone resonates.

Compared with the previous two, passive interaction must conform to the fact that interactive art evolves with its audience, making each encounter unique, and the audience does not notice that. Such as Four thirty-three by John Milton Cage Jr., during the piano concert, no notes are played; only the keyboard cover is opened and closed. A tape recorder is used to record the audience's voices during this silent concert. In this project, "PassBy2" is a passive interaction that downplays the current changes and active interactions, focusing on the long-term accumulation caused by the viewer's unconsciousness.

Passive interaction refers to systems or individuals engaging without direct or active involvement, often relying on inherent properties or feedback mechanisms. This concept is particularly relevant in robotics and human-robot interaction. In art, passive interaction encompasses experiences where the audience engages with art forms without active participation, allowing for emotional and aesthetic appreciation. This concept is explored through various mediums, including traditional portrait art and interactive storytelling [9], where the art influences the viewer's emotional state without requiring direct interaction.

In this project, we combine time with passive interaction to weaken the interactivity of the item and strengthen the feedback generated by accumulation.

2.2. *Concept of "PassBy2"*

The idea of "PassBy2" came from a backpack we used regularly from high school days. The backpack in Fig. 6 faded by long-term use due to personal usage habits. It is faded unevenly, creating a crease-like texture. We never notice the changes in backpacks unless someone reminds us. Likewise, many things are changing unnoticeably, thinking that they will freeze now of pressing the shutter just like the photos, but we fail to notice that the images will also turn yellow and fade.

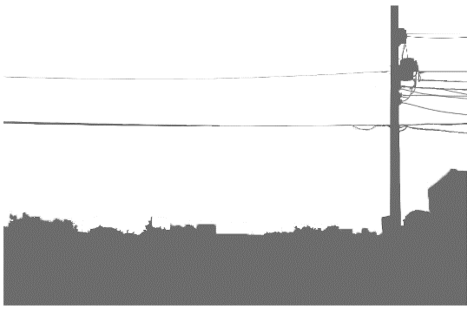
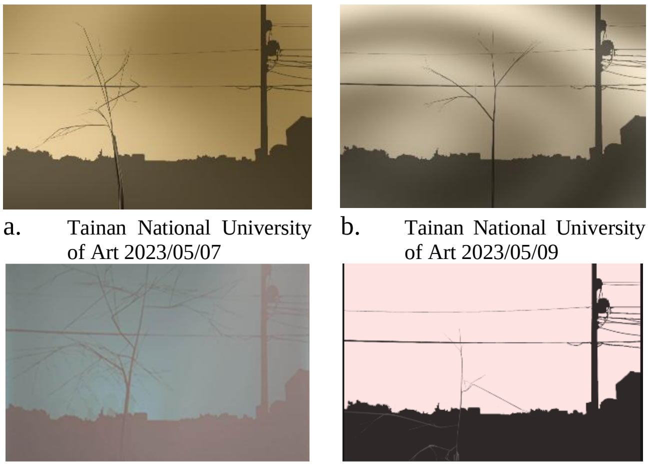


Fig.2 Streetscape filmed in front of my house

The backpack is the emotional connection between someone and his/her hometown, who left to study in college, so "PassBy2" also reflects the emotions of the hometown. The visual focus of this project was a streetscape shot in front of the first author's house (Fig. 2).

Using exhibitions in different places, the street scenes of the hometown were combined with passive interactions to record the local weather and pedestrians passing by the project. The street scenes were slowly changed by the influence of weather and people, just like a part of the hometown was affected by the local people in a strange place. Slowly wearing it with the environment and exposure to different people and weather would create different results, as shown in Fig. 3.



a. Tainan National University of Art 2023/05/07
b. Tainan National University of Art 2023/05/09
c. TransWord University 2022/11/17
d. Yongkang, Tainan 2023/10/20

Fig.3 Trees growing in different dates and places.

3. System Overview

The interaction of this project, "PassBy2," is divided into two aspects: hardware and software (Fig. 4). In terms of hardware, a Kinect camera is used as a sensor to detect the data of passing pedestrians; the received data is transmitted to the computer for calculation, and then the image calculated by the computer is instantly projected to the wall using a projector. The projector's light path will be blocked by pedestrians, causing their shadows to overlap with the projected view.

Connect the weather API on the Internet and transmit local weather information to the system by Json; the TouchDesigner software is used to integrate the data from Kinect and the weather API, and the built-in L-system algorithm and real-time composition of TouchDesigner nodes are used to generate the final compositions.

3.1. *Hardware*

Set up projectors and curtains on both sides of the crowded path so that when light is projected, people passing by will block the light and mix their shadows with the streetscape (Fig. 5). In this project, we use Kinect as the sensor. Kinect has two built-in sensors: an RGB color VGA video camera and a depth sensor [10].

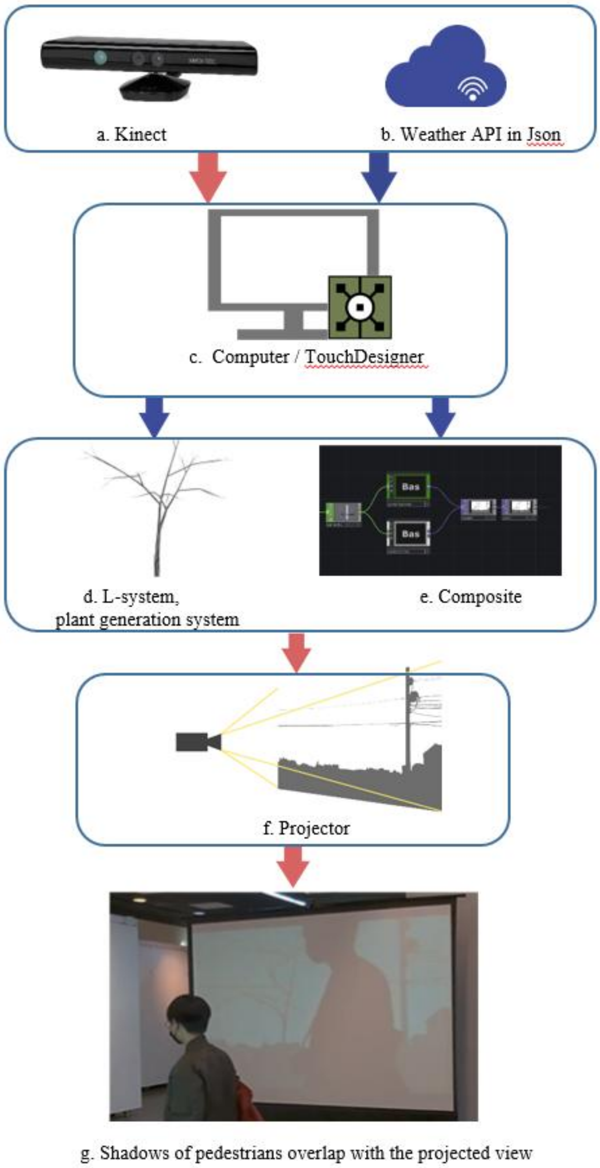


Fig.4 Overview of "PassBy2" installation: The red arrow indicates the hardware connection; The blue arrow indicates the software transfer.

Video camera helps in facial recognition and body recognition; the depth sensor helps create the 3D imagery throughout the room and provides 3D coordinates of the x, y, and z axes of the human body (Fig. 6). The Y-axis is the height z-axis is the normal of the sensor. The x-axis is the designed pedestrian walking direction (Fig. 7). This project mainly uses the functions of Kinect to perform the following three operations:

- Calculate the person's relative position and the detector using the x-axis and z-axis of the person's hip.
- Measure the displacement of the person's x-axis at intervals of 0.5 seconds and calculate the value and direction of the velocity.

- The camera detects the red, green, and blue color components, body type, and facial features. Use this feature to separate the texture of people from the background [11].

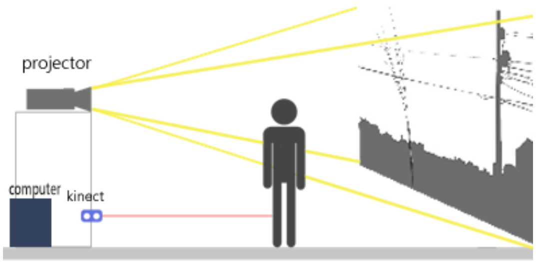


Fig.5 Hardware structure


Fig.6 Kinect can create 3D imagery throughout the room and capture human body contours.

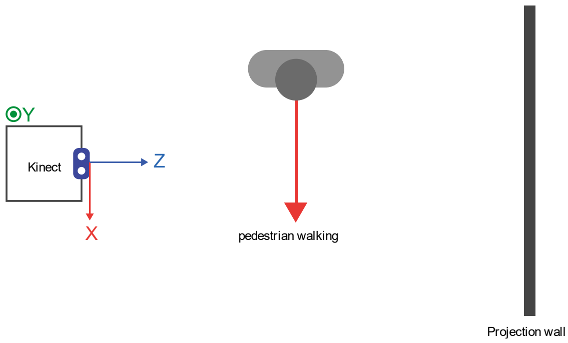


Fig.7 x, y, z axis directions in Kinect from top view.

3.2. *Software*

This project uses a software called TouchDesigner. It is a tool for real-time creativity built for artists, performers, and researchers. It allows them to work freely and efficiently with generative and recorded media [12]. This installation uses TouchDesigner to connect with Kinect.

- “Texture Operators” built into TouchDesigner read the shots from the video camera in Kinect. Select the "Player Index" mode to extract the person’s outline.
- “Channel Operators” built into TouchDesigner read the data from Kinect, which captures the 3D coordinates of the person's hip. Then, they calculate them to get distance and velocity, which transfer data to the L-system so that it can accumulate and use parameters to generate.

In addition, the following two submodules were combined with TouhDesigner's real-time computing function to complete the project: the first one is an L-system-driven tree generator, and the second one captures real-time local weather information by weather API.

- Lindenmayer systems (or L-system) is an iterative function that uses mathematics. The central concept of L-systems is that of rewriting. In general, rewriting is a technique for defining complex objects by successively replacing parts of a simple initial object using a set of rewriting rules or productions. This method is used to simulate the growth of plants and models. As seen from the Algorithm of Plant Beauty, the L system is complete in plant construction. We can create a forest scene, the growth of a vineyard over a house, and a Lily of a valley. In this project, the L-system interacts with pedestrians to generate unique tree shapes [13], [14].
- This time, use the weather API provided by "Visual Crossing." In this API, the following are specified as HTTP query parameters. If you connect to the Internet and set the longitude and latitude, we can get real-time weather information, including description, temperature, cloud cover, etc. Use weather to influence the work and strengthen the connection between the project and the location.

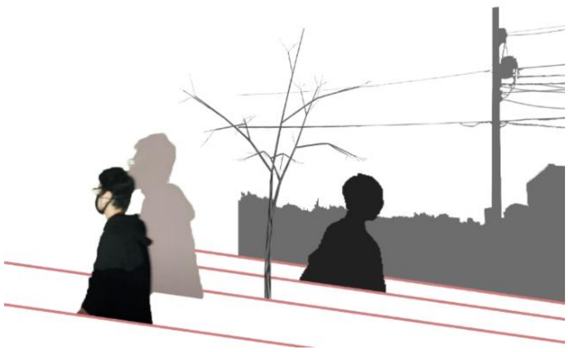


Fig.8 shows layers of visuals sorted from back to front: a streetscape with the weather, a pedestrian silhouette recorded in the past, a tree, and the shadows of current pedestrians.

4. Visual Composition

The feedback generated during interaction is mainly presented visually, and the layered overlap from back to front: streetscape with the weather, pedestrian silhouette recorded in the past, and the tree by L-system. In addition, shadows are caused by people passing by blocking the light from the projector at the front (Fig. 8).

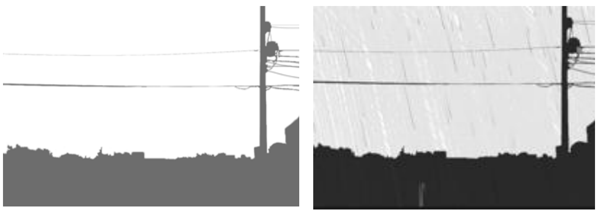
4.1. *Streetscape with Weather*

The streetscape in the project was shot in the alley in front of the station in the first author's hometown. Every time someone walks by the project, some of the colors in the street scenes will be removed, and noise will be added, gradually making the picture darker and rougher. (Fig. 9) Depending on the relative distance between the crowd and the Kinect camera, the increased points and reduced colors will be different when the crowd passes through the work. Although the value dropped by each person passing by is very small and the change cannot be seen immediately, it can produce different looks after accumulation.

Fig.9 Affected by pedestrians crossing, the streetscape becomes rough and dim.

To increase the integration with the local environment, the project introduced a weather API, which can instantly obtain local weather data affecting the project screen:

- When it rains, the number of raindrops will be adjusted based on humidity (Fig. 10).
- The frequency of harmonics that affects the noise generated every time people pass by the project based on visibility.
- Depending on the temperature at the time, the project's color will change when people pass by it.
- Obtain the day's sunrise and sunset time switch works and take screenshots of the project's street view at sunset.



a. Clear day b. Rainy day

Fig.10 Streetscape will produce raindrops when it rains.

4.2. *Catch Pedestrian Silhouette*

In addition to projecting pedestrians' shadows on the street, the motion capture function of Kinect records the outline of pedestrians, and TouchDesigner is used to edit the images into silhouettes (Fig. 11). The image is played back when the next pedestrian passes by, highlighting the traces left by people on the project.

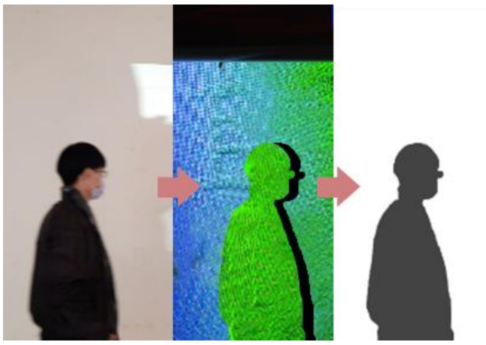


Fig.11 Using the "Player index" mode of Kinect to catch people's silhouettes.

4.3. *Growing Tree by L-system*

In this project, I use the L-system to create a tree; the generation of this tree is defined as the Kinect that calculates the sensed person's ID ordering. Furthermore, designing three models of the different branches (Fig. 12), the corresponding equation is as in [15] (Table 1). According to the current distance between the person and Kinect sensed by Kinect, it is divided into three modes: less than 50cm, between 50cm and 100cm, and greater than 100cm, which affect the growth of the three models, respectively.

Kinect measures a person's displacement within a fixed period to calculate the speed. According to the direction of the speed, it is classified as proper or left so that the model's angle changes to a positive or negative direction; the magnitude of the angle change of the model is determined based on the velocity.

Eq. (1)	Eq. (2)	Eq. (3)
$< 50\ cm$	$\geq\ 50\ cm$; $\leq\ 100\ cm$	$>\ 100\ cm$

Fig.12 Three different tree models

Table 1. L-system functions in Fig. 16

$$\begin{cases} A = !\,F[+F][-F]B \\ B = !\sim(10)"[-F][(b)F]A \end{cases} \quad (1)$$

$$\begin{cases} A = !F[-F]B \\ B = !\sim(10)"F[+(c)F]A \end{cases} \quad (2)$$

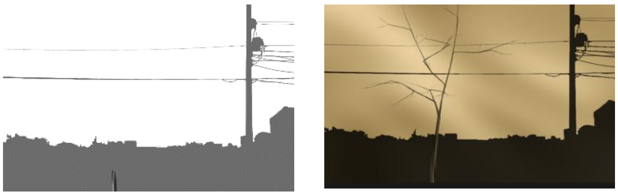
$$A = [\&FL!A]/////'[\&FL!A]/// /'[\&FL!A] \quad (3)$$

5. Conclusion

PassBy2 is a passive interactive work that expresses the accumulation of small things in life. It uses L-system and TouchDesigner to make plants and streetscapes grow and change slowly under the influence of people and record the changes over time (Fig. 13).

When PassBy2 was present at the exhibition, we found that some of the visitors stopped to inspect the trees. Some, especially children, were walking forward and backward to try to speed up the changes in the image. They waved their arms in front of the work to try to control the direction of the tree's growth.

In the future, we plan to directly depict the trajectories of people passing by (Fig. 14) and draw people's trajectories as mountains and street scenes blended instead of noise and color.



a. Beginning at sunrise b. Ending at sunset

Fig.13 Compare the beginning and ending: a is clean and has a tiny sapling; b is turbid and has a great tree that changes under people's influence.

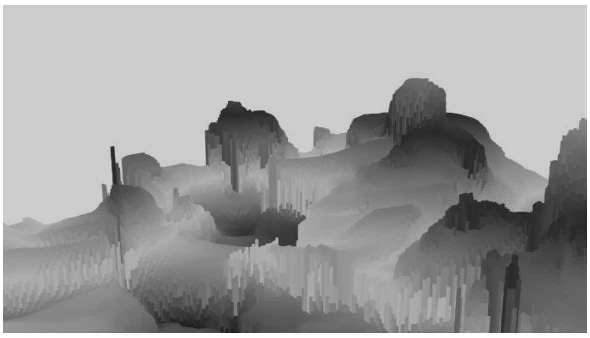


Fig.14 Try to draw the moving track into an image that fits the project style.

References

1. J. Zhong and R.P.C.J. Rajapakse, "Pass-By: Development of Pedestrian Counts-based Art Installation for Passive Interaction." Proceedings of International Conference on Artificial Life and Robotics 2023 (ICAROB2023), vol. 28, pp. 833-836, https://doi.org/10.5954/icarob.2023.os32-4
2. Jack West, "Interactive vs. Passive Art: A Comparison of Impact and Reception," https://www.linkedin.com/pulse/interactive-vs-passive-art-comparison-impact-reception-jake-west
3. Charlotte Lydia Stace, "Banksy Street Art Tour: The Best Graffiti by the Most Secretive of Artists", https://magazine.artland.com/banksy-street-art-best-graffiti-essential-city-guide/
4. R. Gschwend, "The Development of Public Art and Its Future Passive, Active and Interactive Past, Present and Future," Arts, Vol. 4, No. 3, pp. 93-100, https://doi.org/10.3390/arts4030093
5. Cooper, "The Future of Art: 8 Digital Installations and Interactive Spaces", https://architizer.com/blog/inspiration/collections/digital-art-projection-installations
6. S. U. Ahmed, "Interaction and Interactivity: In the Context of Digital Interactive Art Installation," in Human-Computer Interaction. Interaction in Context: 20th International Conference, HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings, Part II. Springer-Verlag, Berlin, Heidelberg, pp. 241–257. 2018, https://doi.org/10.1007/978-3-319-91244-8_20
7. Elburz, Interactive Media, "Interactive Art Installations: 6 Inspiring Examples", March 6, 2024, https://interactiveimmersive.io/blog/interactive-media/interactive-art-examples/
8. "TeamLab," https://www.teamlab.art/e/?type=pickup
9. Y. Ning and T. Sim, "Interactive Portrait Art.". In Proceedings of the 2008 IEEE Workshop on Applications of Computer Vision (WACV '08). IEEE Computer Society, USA, pp. 1–6, 2008, https://doi.org/10.1109/WACV.2008.4543998
10. R. Cong, R. Winters, "How Does the Xbox Kinect Work", https://www.jameco.com/Jameco/workshop/Howitworks/xboxkinect.html?srsltid=AfmBOoq_x_9enZOzMCv_FfrxsL0QxLHyxnXaOcenqZfK7QB7v3eRicqD
11. "Kinect for Windows," https://www.microsoft.com/en-us/download/details.aspx?id=40278
12. "About TouchDesigner", https://derivative.ca/
13. P. Prusinkiewicz and Aristid Lindenmayer. 1990. The algorithmic beauty of plants. Springer-Verlag, Berlin, Heidelberg
14. J. Bernard and I. McQuillan, "Techniques for Inferring Context-free Lindenmayer Systems with Genetic Algorithm," Swarm and Evolutionary Computation, Vol. 64, 2021, 100893, ISSN 2210-6502,
15. R. Sun, J. Jia, and M. Jaeger. "Intelligent tree modeling based on L-system." In 2009 IEEE 10th International Conference on Computer-Aided Industrial Design & Conceptual Design. IEEE, pp. 1096–1100, 2009, doi: 10.1109/CAIDCD.2009.5375256.

Authors Introduction

Mr. Chien-Lin Chung

He received his B.Sc. degree from the Department of Physics, National Taiwan Normal University, Taiwan, in 2021. He is a Graduate student at the Graduate Institute of Animation and Film Art, Tainan National University of the Arts, Taiwan. His research interests include interactive art, computer animation, projection mapping, human motion, and public art.

Dr. Janaka Rajapakse

R.P.C. Janaka Rajapakse is an Associate Professor at the Graduate Institute of Animation and Film Art, Tainan National University of the Arts, Taiwan. He is also a visiting scholar in the Media and Image Technology Department at the Faculty of Engineering, Tokyo Polytechnic University, Japan. He was a Postdoctoral Researcher at the Center for Hyper Media Research, Graduate School of Engineering, Tokyo Polytechnic University. He received his Ph.D. in Knowledge System Science from the Japan Advanced Institute of Science and Technology in 2008. His research interests include computer animation, motion capture, VR/AR/MR, haptic interfaces, AI, computer graphics, and Kansei Engineering. He is a member of the Motion Capture Society, The Society for Art and Science, ASIAGRAPH, and SIG-Design Creativity.

Exploring Techniques to Mitigate Interference in Drone Communication Systems

Ahmed Alsaeed Rashad[1], **Mastaneh Mokayef**[1], **M.K.A Ahamed Khan**[1], **MHD Amen Summakieh**[1], **Kim Soon Chong**[1]
[1]*Department of Electrical and Electronic Engineering, Faculty of Engineering Technology and Built Environment, UCSI University, Kuala Lumpur, 56000, Malaysia*

Abdul Qayyum[2]
[2] *National Heart and Lung Institute, Imperial College London, UK*

Moona Mazher[3]
[3] *Centre for Medical Image Computing, Department of Computer Science, University College London, UK*

Sanjoy Kumar Debnath[4]
[4] *Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India*

Chin Hong Wong[5]
[5] *Maynooth International Engineering College, Fuzhou University, Fuzhou, China*
Maynooth International Engineering College, Maynooth University, Maynooth, Co Kildare, Ireland

Chua Huang Shen[6]
[6] *Department of Electrical & Electronics Engineering, UOW Malaysia University Utropolis Glenmarie, Shah Alam, Malaysia*

Takao Ito[7]*
[7]*Graduate School of Advanced Science and Engineering, Hiroshima University, Japan*
**Email: itotakao@hiroshima-u.ac.jp*

＊Corresponding Author

Abstract

Despite the apparent advantages and strong economic efforts for cellular-connected drones, several critical challenges must be tackled for their successful implementation. Like any radio communication system interference is considered as the biggest challenge as it significantly decreases the efficiency and reliability of the drone. The Monte-Carlo simulation (MCS) strategy is based on the principle of taking samples of random variables from a given distribution. These samples are then used to assess interference in terms of the interference received signal strength compared to the desired received signal strength commonly known as C/I or SNR. The results then are derived using SEAMCAT software as probability of interference where 1 means that this system is always interfered and 0 means it's never interfered. The study has been conducted for separation distances of 2,3,4, and 6 km between the victim receiver and the interfering transmitter where a lower interference probability was achieved as far as they were separated. A frequency allocation approach was also used during this study where we achieved lower interference probability when the systems were operating at a higher frequency.

Keywords: Drones, Victim link receiver, Interfering link transmitter, Probability of Interference, SEAMCAT, iRSS

1. Introduction

The telecommunication field is greatly benefiting from the promising potential of autonomous drone systems, which are also finding extensive use in various industrial applications. In [1] and [2] to address the problem of interference in drone communication, a study was conducted. They understand that interference can hinder communication and reduce the number of drones that can be used at once. By discovering different techniques, the aim is to create a more seamless and reliable communication experience for humans. In [1] it was discovered that successive interference cancellation and space division multiple access techniques can significantly enhance the maximum supported drone density without conceding reliability. In simpler terms, these techniques allow more drones to coexist in the same airspace without causing communication problems. This means that drone operators can fly closer together, ensuring they can capture footage or perform tasks without signal interference. [2] took a different approach by focusing on beamforming and coordinated multipoint (CoMP) techniques. By implementing these methods, they reduced interference by an impressive 13-15 dBm in line-of-sight conditions. In other words, they were able to significantly decrease the unwanted signals that drones receive, resulting in clearer and more reliable communication. This is particularly important for tasks that require real-time data transmission, such as drone inspections or surveillance operations. In [3] innovative solutions to mitigate interference in cellular-connected UAVs were proposed. They recognized that cellular networks often experience interference, which can affect the communication performance of drones. Mei's approach leverages the sensing capabilities of UAVs and inactive base stations to achieve more efficient and reliable operations. By using these resources effectively, they aim to optimize the communication performance of cellular-connected UAVs, ensuring a smoother experience for both drone operators and users. These studies contribute to making drone communication more human-friendly by addressing interference challenges. Through their innovative techniques and approaches, they aim to enhance the communication quality, reliability, and efficiency of drones, ultimately improving the overall user experience.

In recent studies, [4] and [5] have proposed innovative techniques that combine active noise control and spectral subtraction to improve the sound quality of videos recorded by drones. Their work focuses on making drone-captured videos more enjoyable for human viewers by reducing unwanted noise. Additionally, [5] has introduced a fascinating radar system called the Doppler signal-to-clutter ratio (DSCR) detector. This system effectively detects and extracts radar signals from small drones, resulting in a significant reduction in missed target rates. The goal is to enhance the accuracy and effectiveness of radar systems, making them more reliable in various applications. Moreover, [6] presented significant contributions by presenting a new method for estimating signal-to-noise ratio (SNR) in UAV OFDM systems. This method, known as non-data-aided (NDA) SNR estimation, outperforms other approaches, especially in scenarios where synchronization precision is low [7]. The focus here is to improve the performance of UAV OFDM systems, particularly in challenging environments, by providing more accurate SNR estimates. Together, these studies aim to humanize the technology used in drones by reducing noise, enhancing radar capabilities, and improving the overall performance of UAV OFDM systems. By doing so, they contribute to creating a more immersive and reliable experience for human users in various drone applications.

2. Methodology

2.1. *Monte Carlo Simulation*

The objective of this research is to explore and evaluate techniques for mitigating interference in a drone communication system using SEAMCAT software. The study aims to assess the effectiveness of various mitigation strategies in enhancing the performance and reliability of the drone communication system in the presence of interference. Thus, this study follows a simulation-based design utilizing SEAMCAT software. The Monte Carlo method, which serves as the foundation of SEAMCAT, is a statistical technique that sets itself apart from traditional analytical approaches that rely on solving differential equations to comprehend physical or mathematical systems. With Monte Carlo simulations, there is no need to explicitly formulate these equations; instead, we directly simulate the physical process. To employ the Monte Carlo method, we randomly select variables from known distributions. Prior to running the simulation, we must define all parameters of our radiocommunications system, such as antenna heights, powers, frequencies, and locations. For values that remain constant (like fixed frequencies or heights), they can be predetermined. We also obtain technical specifications from industry standards, such as those established by IEEE, 3GPP, and ETSI [8]. These distributions are used by SEAMCAT to produce random events. SEAMCAT keeps track of each event's interfering signal strength and computes desired signals in separate data arrays. Lastly, you can calculate the probability of interference by comparing each event's desired and unwanted signals at the victim link receiver to the applicable interference criterion, like C/I [8].

2.2. *Scenario*

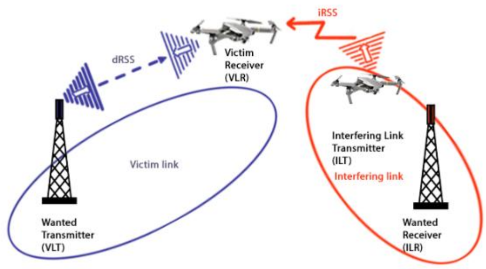


Fig. 1. System scenario.

As illustrated in Fig. 1 SEAMCAT simulates a scenario where a single victim link receiver (VLR) is connected to a victim link transmitter (VLT). This VLR operates in the presence of one or more interfering link transmitters (ILTs), which are paired with interfering link receivers (ILRs). The presence of the victim system, the interfering system, or both can lead to interference [8]. In order to replicate real-world interference situations, the interfering transmitters are strategically placed around the victim device. The user has the flexibility to define the distribution pattern, which can be either random or follow a specific relationship with respect to the victim's location. A crucial parameter in this context is the desired received signal strength (dRSS), which refers to the signal strength measured between the victim transmitter and receiver. The user has the ability to set the dRSS to a predetermined value. Additionally, the interfering received signal strength (iRSS) is measured between the interfering transmitter and the victim receiver. The iRSS serves as the primary metric for evaluating the extent of interference in the system [9].

The iRSS in our system is further divided into two main types known as $iRSS_{unwanted}$. Where the $iRSS_{unwanted}$ is the level of unwanted emissions (i.e. comprising of the out-of-band emissions and the spurious emissions of the ILT) lying within the VLR receiver bandwidth [8].

2.3. *Software Parameters*

The software parameters for SEAMCAT are detailed in this section, derived directly from ITU-R M.2233 to guarantee real-world relevance and adherence to operational standards [10].

Table 1. Victim & Interfering link parameters.

Parameter	Victim Link	Interfering Link
Operating Frequency (GHz)	5.03	5.06
Transmitting power (dBm)	40	40
Coverage radius (Km)	15	28
Propagation Model	ITU-R P.525 (Free Space)	ITU-R P.525 (Free Space)
Transmitter antenna Gain (dBi)	28	28
Receiving antenna Gain (dBi)	-10	-10
Reception Bandwidth (KHz)	37.5	37.5

According to [11] serious interference happens when the Carrier to Interference ratio (dRSS/iRSS or C/I) at the victim receiver is less than the allowable threshold. SEAMCAT produces random spatial and temporal distributions of VLT, ILT, VLR, and their parameters to find if the interference has occurred at the VLR or not. The user can adjust in the software interface the parameters of the VLT and ILT and the spatial position of VLT and ILT concerning each other. The parameters adjusted for the simulation of the victim link and Interfering link can be shown in Table 1.

When the carrier-to-interference ratio of the VLR falls below the minimum allowable protection ratio, interference occurs. The minimum protection ratio can vary depending on the modulation type, diverging from a standard value of 9 dB to higher levels [8]. Thus, interference is inevitable when the VLR's carrier-to-interference ratio fails to meet the minimum permitted protection ratio [11].

The Monte Carlo Simulation (MCS) process involves considering multiple independent events that take place at different locations and time frames. Each event is simulated by taking into account various random variables, such as the position of the interfering source relative to the victim. By incorporating these diverse scenarios, the received signal strengths (dRSS and iRSS) between the victim location receiver (VLR) and victim location transmitter (VLT), as well as between VLR and interfering location transmitter (ILT), can be computed. To obtain accurate results, a sufficient number of simulation attempts, typically around 20,000, are conducted. This ensures that the probability of interference can be reliably determined. Hence, it is recommended to use a significant number of simulation iterations to ensure consistent and dependable outcomes [11].

Four interference criteria are considered within SEAMCAT. C/I Carrier to interference ratio; C/(I+N) Carrier to interference plus noise ratio; (N+I)/N Noise plus Interference to noise ratio; I/N Interference to noise ratio. In our Scenario, we focus our study on the C/I and C/I+N, where we use the SEAMCAT built-in library to set the C/I threshold to 19dB and the C/I+N to 16dB.

3. Results and Discussion

Our methodology to mitigate interference is based on a parametric study to decrease the probability of interference the VLR undergoes to obtain more reliable communication throughout the victim link. This parametric study focuses on observing the probability of interference calculated by SEAMCAT under different distance separation scenarios between the VLR & ILT, as well as using different operating frequencies of VLR & ILT making sure they comply to the operating frequency band of drones specified by ITU. Our target here is to separate them as far as possible in order to get the best probability of interference possible complying with the interference criterion set previously, and without degrading the performance of our system

3.1. *Distance separation approach*

We use the separation distance between VLR & ILT as 2 km, and the victim link & interfering link both operate at the same frequency 5030 MHz, whereas all the rest of the parameters will be held constant throughout our observations.

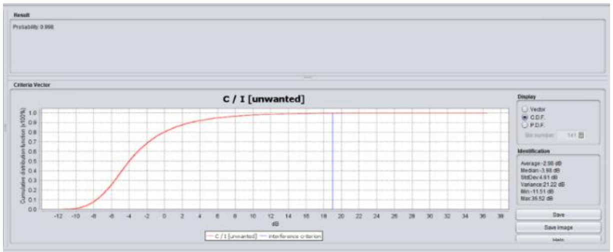


Fig. 2 iRSSunwanted at 2 Km separation distance

Fig. 2 is the result for $iRSS_{unwanted}$ at 2 Km separation distance, where we can divide the figure into 3 parts. First is the Criteria Vector where we see that the dRSS to iRSS ratio (dRss/iRSS) or C/I is plotted against the C.D.F. This plot illustrates the results obtained for dRSS/$iRSS_{unwanted}$. The second part is Identification where we focus on the Maximum power captured by the interferer in this case it's 36.52 dB. We focus on the Maximum power as it's the one above our criterion threshold. Finally, the third part, which is the result of the probability of interference gives us a 0.998 probability of interference which sets our target to decrease this number to mitigate system interference.

Table 1 Probability of interference of iRSSunwanted for distance separation between VLR & ILT.

Distance (Km)	**$iRSS_{unwanted}$**	
	Probability	**Max power (dB)**
2	0.998	36.52
3	0.995	34.96
4	0.991	40.07
6	0.98	50.89

Table 2 shows the results using the separation distance of 2,3,4,6 Km respectively, while the operating frequency is still held to 5030 MHz for both Links. From Table 2, we can observe that the probability of interference for the $iRSS_{unwanted}$ decreases as we increase the separation distance between the VLR & ILT, thus we can consider interference mitigation to our system. However, as soon as the separation distance is increased for more than 3 Km the Maximum interfering power is increased which is mainly because the ILT is randomly placed concerning the VLR while the VLT is held in a constant distance in respect to the VLR. This makes the VLR more vulnerable to greater interfering power.

3.2. *Frequency allocation approach*

For the first trial, we use the frequency for the victim link as 5045 MHz, and the interfering link as 5060 both complying with the ITU standard. When the VLR is placed 2 Km away from the ILT a better probability of interference of 0.997 is derived for both the $iRSS_{unwanted}$ that will be having the same values as depicted in Fig. 3.

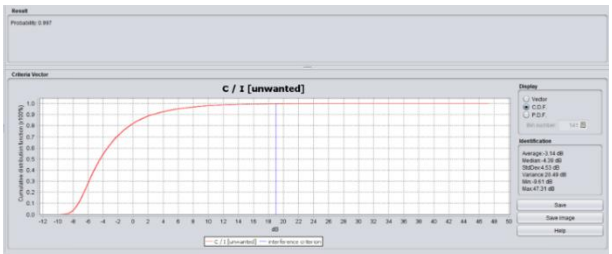


Fig.1 . iRSSunwanted when the Victim link operates at 5045 MHz & Interfering Link at 5060 MHz

Another crucial investigation we have come through is if both Links are to be operating at the same frequency, it's preferable to be a higher frequency. Fig. 3 illustrate that when both the Victim & interfering link are operating at 5060 MHz they achieve a lower interference probability of 0.987 for $iRSS_{unwanted}$ when the VLR is placed 2 Km away from the ILT. This is a better interference probability in comparison to when both the Victim Link & Interfering link were operating at 5030 MHz where they achieved a probability of 0.998.

4. Conclusion

In conclusion, the MCS method was used utilizing SEAMCAT software to assess the effect of distance separation & frequency allocation of the victim receiver in respect to the interfering transmitter on the probability of interference. The interference is divided into two types which are $iRSS_{unwanted}$. Both of the types achieved a significant lower probability of interference when the distance separation method was used. We reported a probability of interference of 0.998, 0.995, 0.991, and 0.98 for $iRSS_{unwanted}$ and 1, 0.999, 0.998, and 0.995 for $iRSS_{blocking}$ for the distance separation of 2,3,4, and 6 KM respectively. The separation distance of 3 km was prioritized as it captured the lowest interfering power. For the frequency allocation, the operation of systems on higher frequency presented lower interference probability of 0.987 for $iRSS_{unwanted}$ in comparison to operation on lower frequency when placed at the minimum separation distance.

References

1. Schalk, Lukas Marcel, Techniques for improving the cooperative traffic conflict detection among drones, 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring), IEEE, 2019.
2. Hee Fung Kim, and Mastaneh Mokayef, Interference Mitigation Techniques For The Operation Of Unmanned Aerial Vehicle (Uav), 2022 IEEE 5th International Symposium in Robotics and Manufacturing Automation (ROMA), IEEE, 2022.
3. Mei Weidong, and Rui Zhang, Aerial-ground interference mitigation for cellular-connected UAV, IEEE Wireless Communications 28.1 (2021): 167-173.
4. Ahn Hyohoon, et al., Hybrid Noise Reduction for Audio Captured by Drones, Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication, 2018.
5. Kang Byungseok, Hyohoon Ahn, and Hyunseung Choo, A software platform for noise reduction in sound sensor equipped drones, IEEE Sensors Journal 19.21 (2019)
6. Gong Jiangkun, et al., Improved Radar Detection of Small Drones Using Doppler Signal-to-Clutter Ratio (DSCR) Detector, Drones 7.5 (2023): 316.
7. Li J., Liu M., Tang N., Shang B., Non Data-Aided SNR Estimation for UAV OFDM Systems, Algorithms 2020, 13, 22.
8. SEAMCAT Handbook Edition 2 (2016).
9. Mokayef Mastaneh, et al., Spectrum sharing model for coexistence between high altitude platform system and

fixed services at 5.8 GHz, International Journal of Multimedia and Ubiquitous Engineering 8.5 (2013): 265-275.

10. Examples of technical characteristics for unmanned aircraft control and non-payload communications links M Series Mobile, radiodetermination, amateur and related satellite services, (2011).
11. Girma, Solomon T., Dominic BO Konditi, and Ciira Maina, Frequency re-use distance calculation in cellular systems based on Monte-Carlo simulation, Heliyon 5.3 (2019).

Authors Introduction

Mr. Ahmed Alsaeed Rashad

He is a UCSI graduate student with a Bachelor of Engineering (Honours) Communication and Electronic Engineering.

Dr. Mastaneh Mokayef

Her PhD from Wireless Communication Centre Faculty of Electrical Engineering in University Technology Malaysia (UTM). She has many research papers.

Dr. M. KA. Ahamed Khan

He pursued a PhD in Robotics, Power Electronics, and Controls in the United States and holds Chartered Engineer (CEng) in the UK. He is a Senior member of the IEEE in the USA and a member of MIET in the UK. He is also the past chair for IEEE RAS Malaysia chapter.

Mr. MHD Amen Summakieh

He received the B.Eng. degree (Hons.) in communication and electronics engineering from UCSI University, Malaysia, in 2016, and the B.Eng.Sci. degree from Multimedia University, Malaysia, in 2020. His research interests include heterogeneous LTE-advanced cellular networks, user association, metaheuristic algorithms, and antennas design.

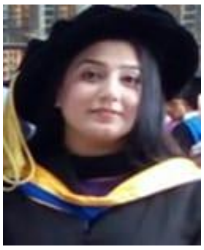
Dr. Sanjoy Kumar Debnath

He completed Ph.D. from the Faculty of Electrical & Electronic Engineering, Universiti Tun Hussein Onn Malaysia (UTHM) in 2022. He is currently working as an Assistant Professor at the Chitkara University Institute of Engineering and Technology (CUIET), Chitkara University, Punjab, India.

Dr. Abdul Qayyum

He did his PhD in Electrical and Electronics Engineering from Universiti Teknologi PETRONAS, Malaysia 2018. He developed deep learning based algorithms for depth estimation of vegetation, trees near power lines for Tenaga Nasional Berhad (TNB) and Sabah Electric Supply Berhad (SESB) under the ministry of Green , Water and Technology (KeTTHA) Malaysia. He also developed a prototype for Vital signs (heart rate, breathing rate, Sp02) estimation and assessment of stroke and Arterial fibrillation (AF) using face video analytic based on deep leaning models. Earlier, he had completed his bachelor's in computer engineering and Master in Electronic Engineering from Pakistan. Besides, he gained one-year industrial experience while working as a BSS engineer for Huawei, Pakistan. He had also taught several courses under electrical, specifically, signal processing domain for 7 years in various public and private universities in Pakistan. He was working as a research scientist in CISIR, UTP for less than one year and was developed deep learning algorithms for brain signal (EEG) classification and reconstruction, remote sensing image segmentation and biomedical image analysis. He is associated with burgundy university France as a Post Doc researcher and working on the cardiac MRI images using deep learning approach. He is also working as a consultant in various projects involving deep learning models in Big Data, Vital Sign Estimation, IoT and BCI applications. Currently he is associated with National Heart and Lung Institute, Imperial College London, UK.

Dr. Moona Mazher



She received her Ph.D. in Computer Engineering and Mathematics from the University of Rovira i Virgili, and works in UCL, UK.

Dr. Chin Hong Wong

He is a Lecturer in Maynooth International Engineering College at Fuzhou University in China. He received his PhD in Electrical and Electronic Engineering from Universiti Sains Malaysia in 2017. His research interests are Energy harvesting and control system.

Dr. Kim Soon Chong

He obtained his B.Eng., M.S., and Ph.D. degrees in electrical and electronic engineering from Universiti Kebangsaan Malaysia in 2012, 2015, and 2019, respectively. He currently serves as an Assistant Professor in the Department of Computer Engineering (Artificial Intelligence) within the Faculty of Engineering Technology and Built Environment at UCSI Kuala Lumpur. His research focuses on areas such as medical signal processing and instrumentation.

Mr. Chua Huang Shen (Mason)

He has dedicated the majority of his professional career to UOW Malaysia KDU, where he currently holds the position of Senior Lecturer in Electronics and Electrical Engineering with a keen focus on embedded systems, renewable energy, and bio-energy. He obtained his academic foundation from esteemed institutions, earning a Master of Science from the University of Belfast in 2004 and, previously, a BEng (Hons) in Electronics and Control Systems from the University of Liverpool John Moore in 2002. Currently, he is actively engaged in pursuing a Ph.D. in engineering from Universiti Tunku Abdul Rahman (UTAR), further enriching his knowledge and expertise.

Dr. Takao Ito

He is Professor of Management of Technology (MOT) in Graduate School of Engineering at Hiroshima University. His current research interests include automata theory, artificial intelligence, systems control, quantitative analysis of interfirm relationships using graph theory, and engineering approach of organizational structures using complex systems theory.

An Automated Tracking System for Locating Impact Points on a Table Tennis Surface Using Ping Pong Balls

Lee Wai Kit[1], **Mastaneh Mokayef**[1*], **MHD Amen Summakieh**[1], **M.K.A Ahamed Khan**[1], **Miad Mokayef**[1], **Sew Sun Tiang**[1], **Wei Hong Lim**[1]

[1]*Department of Electrical and Electronic Engineering, Faculty of Engineering Technology and Built Environment, UCSI University, Kuala Lumpur, 56000, Malaysia*
**Email: mastaneh@ucsiuniversity.edu.my*
https://www.ucsiuniversity.edu.my/

Abdul Qayyum[2]
[2] *National Heart and Lung Institute, Imperial College London, UK*

Moona Mazher[3]
[3] *Centre for Medical Image Computing, Department of Computer Science, University College London, UK*

Sanjoy Kumar Debnath[4]
[4] *Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India*

Takao Ito[5]
[5]*Hiroshima University, Japan*
**Email: itotakao@hiroshima-u.ac.jp*

＊*Corresponding Author*

Abstract

This project presents a novel system for real-time table tennis ball tracking and hitting point prediction, developed using OpenCV-Python. The system is designed to contribute to strategic analysis in the sport by accurately identifying the ball's trajectory and anticipating its landing point. The proposed system is comprised of four key modules: ball detection, ball tracking, hitting point prediction, and data visualization. Computer vision techniques are employed to effectively detect and monitor the ball's movement. Kalman filtering is utilized to refine the prediction of the ball's landing point. The generated data is then presented visually to facilitate analysis and comprehension of game dynamics.

Keywords: Real-time detection and tracking, Hitting point prediction, Kalman Filter

1. Introduction

Table tennis, commonly referred to as ping pong or whiff-whaff, is a sport derived from tennis, albeit adapted for play on a smaller, stationary table surface demarcated by nets rather than a traditional tennis court. Players in both sports engage in a similar style of play, aiming to return the ball over the net and into the opponent's half of the playing area using rackets. Failure to do so results in a point being rewarded to the opposing player. The game can be played either individually or in teams of two and is versatile in that it can be enjoyed both indoors and outdoors, thereby appealing to a wide range of casual participants. The compact play area and the rapid trajectory of the ball in table tennis contribute to its reputation as a fast-paced sport in comparison to other ball games. Players of table tennis are required to exhibit rapid and continuous reactions during gameplay, responding to the ball's unpredictable trajectory. This demand arises from the nature of table tennis rackets, typically constructed with rubber materials that facilitate enhanced ball spin and acceleration. The global appeal of table tennis transcends age and skill levels, with enthusiasts engaging in the sport recreationally and competitively across various settings. The sport of table tennis revolves around essential equipment, notably the table tennis ball, the table, and the racket. International regulations dictate the use of a spherical object weighing 2.7 grams (0.095 ounces) with a diameter of 40 millimeters (1.57 inches). Since 2015, these balls have transitioned from celluloid to polymer materials, offering white or orange options with a matte finish. The selection of ball color is often influenced by the playing surface and its surroundings; for instance, a white ball contrasts more effectively on green or blue tables than on grey surfaces. The table dimensions measure 2.74 meters (9.0 feet) in length, 1.525 meters (5.0 feet) in width, and stand at a height of 76 centimeters (2.5 feet). The playing surface is uniformly dark blue and matte, divided into two halves by a net positioned 15.25 centimeters (6.0 inches) high. The International Table Tennis Federation (ITTF) strictly sanctions tables crafted from wood or its derivatives. In public outdoor spaces like parks, concrete tables with steel nets or solid concrete partitions may occasionally be encountered. Modern table tennis rackets, also known as paddles or bats, play a pivotal role in shaping a player's performance. A superior-quality racket enhances ball speed and trajectory, optimizing players' techniques and strategies. Key components of a racket encompass the blade, handle, and rubber. The blade serves as the racket's primary structure, crafted from wood. The handle, the gripping section, is tailored primarily to fit the player's hand size and grip style—Penhold or Shakehand—emphasizing comfort and control. The rubber sheets, colored red and black, affixed to both sides of the blade directly interact with the ball, requiring careful selection to align with the player's playing style. In recent years, table tennis has evolved into a fiercely competitive and increasingly popular sport. Technological advancements have revolutionized the game, particularly concerning equipment. Innovations in materials and manufacturing techniques have ushered in lighter, faster, and more spin-friendly gear. Carbon fibers, advanced rubbers, and blade technologies have empowered players to unleash greater power and spin in their shots. Furthermore, components such as balls and tables have undergone rigorous standardization by the ITTF to uphold the integrity and quality of the sport [1].

Recent research has focused on developing automated systems for tracking table tennis balls and their impact points. Authors in [2] proposed a low-cost acoustic system using sensors placed under the table, achieving positioning accuracies above 94%. Study in [3] explored ball impact localization on rackets using piezo-electric sensors, with a linear regression method yielding RMSEs of 22.1mm and 19.8mm in transversal and longitudinal directions, respectively. Authors in [4] reported on a student-built system that tracks ball location and keeps score in real-time using video analysis. Prototype [5] developed a machine vision-based algorithm for tracking ball rotation and trajectory, incorporating deep neural networks, Fourier transform-based speed measurement, and CNN-based rotation direction detection. These advancements in automated tracking systems demonstrate potential applications in training, match analysis, and even robotic assistance in table tennis. The real-time automated tracking ball is a novel approach that leverages advanced imaging techniques and machine learning algorithms to enhance the precision of gameplay analysis. This system not only identifies the exact locations where the balls make contact with the surface but also analyzes the speed and angle of each shot, providing players with valuable insights to improve their performance. By integrating real-time data visualization, coaches and players can make informed decisions during training sessions, ultimately leading to a more strategic approach to the game. Furthermore, the system can be adapted for various training scenarios, allowing for customized drills that target specific skills such as spin control, shot placement, and reaction time. This adaptability ensures that players at all levels can benefit from tailored feedback, fostering a deeper understanding of their strengths and areas for improvement. In addition, the technology can track progress over time, enabling athletes to set measurable goals and monitor their development, which enhances motivation and accountability in their training regimen. This continuous feedback loop not only helps in refining techniques but also builds confidence as players see tangible results from their efforts [6]. Furthermore, the integration of video analysis allows players to visually assess their performance, making it easier to identify patterns and make necessary adjustments in real-time. This holistic approach to training empowers athletes to take ownership of their development, encouraging a proactive mindset that translates into improved performance on the field [7, 8, 9].

The sport is undoubtedly growing in the right direction, having said that, some challenges might hinder table tennis development. Many players do not understand the significance of strategic analysis in the context of table tennis gameplay, especially in light of evolving game strategies and tactics influenced by modern playing styles and equipment innovations. The problem could also be caused by individual casual players who cannot afford to invest in those expensive professional training systems or are just unfamiliar with the concept of strategic analysis in table tennis [10].

Incorporating advanced technologies like computer vision into training and match analysis tools remains a challenge due to the complexity of real-time data processing and accuracy requirements. With faster-paced rallies and heightened player athleticism, there is a growing need for tools that can analyze and adapt to these dynamic gameplay scenarios. Effective coaching and player development relies on accurate performance analysis tools that can provide actionable insights into both strategic decision-making and technical execution.

When considering the design and implementation of a computer vision system for table tennis tracking and prediction utilizing OpenCV, the primary challenge is consistently presented by the rapid movements and spins of the table tennis ball in comparison to other ball sports, thus potentially affecting the accuracy of trajectory estimation. The prediction of ball trajectory may be influenced by the diverse hitting patterns of players' rackets and their individual playing styles [11].

Implementing algorithms for real-time detection and localization of the hitting points where the ping pong ball lands on the table involve utilizing image processing techniques to identify the exact location where the ball makes contact with the table surface. Problems will likely occur during the process of providing real-time feedback on the landing position to players and coaches for performance analysis and training purposes. Not to mention that vary minor aspects including lighting conditions, player movements, and ball speed variations could possibly decrease real-time processing accuracy.

In this paper a real-time table tennis ball tracking and hitting point prediction system was designed and developed based on OpenCV-Python platform. The result of the system includes the ball trajectory and hitting point prediction was performed through a graph using Matplotlib.

2. Methodology

The system is a real-time table tennis ball tracking and hitting point prediction system. The system can be divided into four sections: ball detection, ball tracking, hitting point prediction, result visualization, and plotting. The system should be able to detect table tennis balls and track the ball's movement. Based on the ball's movement, a trajectory of the ball should be drawn. After that, the system should predict the hitting point of the ball by implementing ball trajectory data and the Kalman filter. Finally, after getting the prediction result, the system should plot out the ball trajectory and hitting point prediction in graph form. Major hardware components that are used in this project include a table tennis ball, a table tennis table, a camera, and a processing platform which is a laptop.

Fig. 1 illustrates a ping pong ball tracking and hitting point prediction system. It involves detecting the ball's location, tracking its movement, predicting where it will hit the table, and visualizing the results. OpenCV is used in this project, it is an open-source software library that developed for Artificial Intelligent, Computer Vision, object detection and Deep Learning purpose. The OpenCV library includes various functions such as image loading and resizing, real-time video processing, colour converting and detection, making this library fits perfectly to this project. There two programming languages can be used to operate the OpenCV library: C++ and Python language. Python language was chosen in this project due to its ease of use and rich yet complete libraries available.

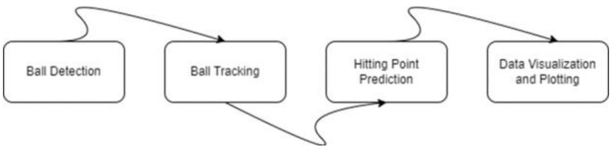


Fig.1. Ping pong ball tracking and hitting point prediction system overview.

The programming code for this system was started by importing relevant Python libraries. The first library is OpenCV library, which includes the majority of functions that will be then utilized in the program such as, real-time video capturing and processing, color filtering, target object tracking etc. The second library is the NumPy library, a Python library used for working with arrays, it includes functions for working in the domain of linear algebra, Fourier transform, and matrices. In Python, there are lists of other libraries and functions that serve the purpose of arrays, but they are slow to process. NumPy on the other hand is able to provide an array object that is up to 50x faster than traditional Python lists. The last library is the Matplotlib library, it is a low-level graph plotting library in Python that serves as a visualization utility. In this program, Matplotlib serves the purpose of plotting out the ball trajectory and hitting point prediction in easily understand graph form.

2.1. *Ball detection*

For object detection, a camera port is needed to capture real-time video. Python code 'Cv.VideoCapture ()' was used to define a camera port. The number in the bracket for this function determines the device selected for video capturing. For Windows System devices, '0' by default brings to the computer a prebuilt webcam, while other numbers indicate external connect camera devices based on USB ports.

There are various ways to achieve object detection in OpenCV such as by movement, by shape of the object and by colors, etc. Both shape and color detection methods were tried in this project to compare the performances, pros, and cons of each method. For the shape detection method, the color will not matter for the detection, therefore the camera output needs to be blurred and greyed. At the same time, lowering the clarity and image quality can lower the detection and calculation difficulty for later processes. Any circle-shaped object that occurs inside the camera angle will be detected and tracked. The second method of ball detection is the color detection method. The orange color object will be detected in this case. The color detection method was preferred over the other one. This is mainly because while the shape detection method is much easier to execute, it performs with lower accuracy when the detection range and distance is increased. Objects that are nearly circle could be defined as a ball that causes more than one object to be captured. The downside of the color detection method is the need for better lighting conditions in surroundings. After the system successfully detects a table tennis ball, a blue dot will be generated at the center of the ball image.

2.2. *Ball tracking*

The program followed by ball tracking section. In the previous part, a blue dot that represented the real-time location of the ping pong ball was generated. Based on the dots, the system was programmed to draw out a yellow line trajectory of the ball. It draws lines connecting the points in the trajectory list if drawing has started and there are at least two points. The yellow line will only start drawing once blue dots are detected.

2.3. *Hitting point prediction*

The third section of the system was hitting point prediction calculation. In this project, we utilized the Kalman Filter for prediction algorithms. The Kalman Filter is a powerful mathematical tool used for estimating the state of a system from noisy measurements. In trajectory prediction, it is used to estimate the future position of an object based on its current position and velocity, while accounting for noise and uncertainty in the measurements. This part of the program started by setting up all the parameters for the filter including State Transition Matrix (A) which defines how the state evolves over time, Measurement Matrix (H) which relates the observed measurements to the state, Process Noise Covariance (Q): which defines uncertainty in the process model, measurement Noise Covariance (R): which defines uncertainty in the measurement and Error Covariance Matrix (P): which represents the uncertainty in the initial state.

Fig. 2 shows the camera view window that will be generated while the system's program is running. The blue dot that is located at the center of the ball is achieved by the ball detection section programming code. A yellow line was drawn which indicated the ball movement trajectory. After this, the yellow dot signifies the anticipated point of impact of the ball, as calculated periodically using the Kalman Filter algorithm. This projection is established through a fusion of the ball's prior trajectory and velocity.

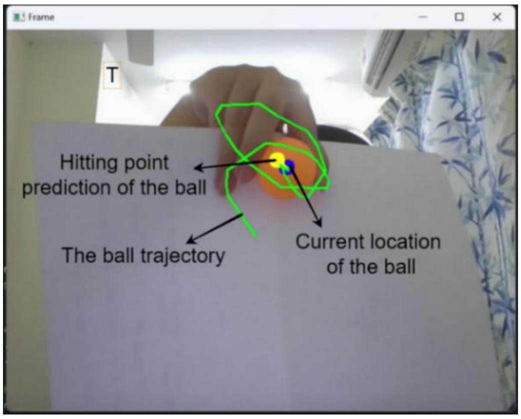


Fig. 2. Camera view while ping pong ball tracking system is running.

2.4. *Data Visualization and Plotting*

The last section after the ball detection, ball tracking, and ball-hitting point prediction was data visualization and plotting. The result of the program prediction was performed by plotting it out through a graph which will automatically pop out after the user ends the program. The library that helps to achieve the plotting is Matplotlib.

Fig. 3 shows the result plotting example of the system when the program is terminated. Information that is plotted by the program includes the complete trajectory of ball movement, start and end points, and the hitting point location prediction. The plotting function is set to only start working when a trajectory is recorded, which means no graph will be displayed when no ball was detected from the beginning. When this situation happens, the program will print out the text "No trajectory points to plot" when it ends.

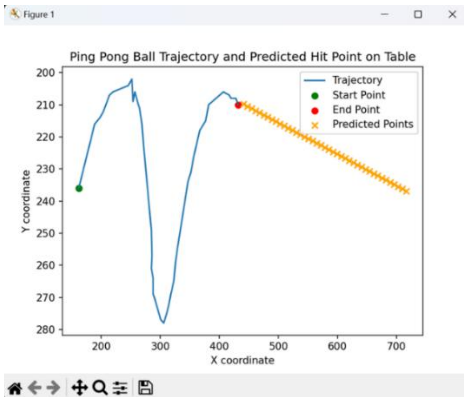


Fig. 3. The visualization of the program's outcome.

3. Results and Discussion

The project 'Ping Pong Ball Tracking System for Identifying Hitting Points on a Table Tennis Table' proved that the feasibility of a similar system and project in table tennis or other sports. The prototype made throughout this project achieves the basic required function according to the project objectives.

For the Ball Detection part, two common methods of object detection from OpenCV were both developed and tested to compare their performance, advantages, and disadvantages between them. In short, the shape detection method is simpler and easier to apply for only detection objects. When other functions were added to the program such as ball tracking and hitting point prediction, the detection result using shape detection is not accurate enough to support the following steps. Furthermore, the accuracy of the detection decreases when the detection distance and range is increased. Due to that reason, the color detection method was chosen to achieve better performance). However, it might struggle in cases where the ball is obscured or in environments with similar colors and when the lighting condition is poor since the lighting can affect the color depth of objects. This problem can be fixed by controlling surrounding lighting and adjusting the detection color range.

This color method is effective for objects with distinctive colors in controlled settings, aligning with the findings of [12], emphasizes the efficiency of color-based tracking techniques under stable lighting conditions. In the ball detection part of the program, the color detection method is used to identify the ball in each video frame. It works by converting the frame to HSV color space, creating a mask for the specified color range, and finding contours. The largest contour that matches the ratio and area criteria is considered the ball. Based on the result of this project, the ball detection is well with high accuracy when all variables are under control. The elements that affect detection the most are the detection distance and ball moving speed.

The next part of the project comes to the ball-tracking part. The expected outcome of this part is to track and draw out a trajectory of the ball's movement when the ball is detected. The ball trajectory is the indispensable data that is required for later prediction calculation. Based on the data collected from ball detection part, we visualize the current location of the ball using a blue dot. Then, by recording every location of the blue dot, the program draws a yellow line connecting the points in the trajectory list if the drawing has started and there are at least two points. The yellow line will only start drawing once blue dots are detected and will stop when no blue dot is present.

During the testing and evaluation project, the trajectory visualization is clear and accurate. The annotated video frames are displayed in real-time with the ball's position, trajectory, and predicted positions. The yellow line is always drawn close to the actual trajectory, as the ball moves, its trajectory is continuously updated and drawn on the frame. Drawing the trajectory in real-time helps in understanding the ball's movement and can be useful for analyzing the ball's behavior during gameplay.

The program continues by calculating the prediction of ball hitting point from time- to time using the ball trajectory collected. In terms of hitting point prediction, several aspects have to be counted for better, including the ball movement speed, ball location, and the direction of ball movement. For that purpose, the Kalman Filter was used in this program. The Kalman filter plays a crucial role in predicting the ball's future position and refining its trajectory estimates. By combining a predictive model with measurements, the Kalman filter smooths the trajectory and handles the measurement noise. This approach is well-documented in foundational works like those of Welch and Bishop (1995), which demonstrate the filter's robustness in dynamic state estimation [6]. The filter's performance heavily depends on the proper tuning of parameters: State Transition Matrix (A) which defines how the state evolves over time, Measurement Matrix (H) which relates the observed measurements to the state, Process Noise Covariance (Q): which defines uncertainty in the process model, measurement Noise Covariance (R): which defines uncertainty in the measurement and Error Covariance Matrix (P): which represents the uncertainty in the initial state. Adjusting these parameters can improve tracking accuracy, especially in varying lighting conditions or with different ball speeds.

The study of Kalman Filter introduces two important terms and working principles of the filter: state prediction and state correction. State Prediction, the Kalman filter uses the state transition matrix to predict the next state of the ball, including position and velocity. It predicts where the ball will be in the next frame. State Correction, the filter updates its predictions using actual measurements to refine its estimates. It draws the predicted and extended predicted positions on the frame to visualize the ball's trajectory and future path. The Kalman filter assumes linear dynamics and Gaussian noise, which may not perfectly represent real-world scenarios. For more complex motion patterns, alternative filters like the Extended Kalman Filter or Particle Filter might be considered.

In the ball-hitting point prediction part, both two-dimension and three-dimension Kalman Filter parameters were setup. Even though the three-dimension involves one more z-axis of the program, the program does not seem to get much improvement in prediction performance. Having said that, the Kalman Filter still provides accurate prediction results for the program. Based on the result from testing, various factors such as ball movement speed, ball location, and the direction of ball movement were counted into the prediction too. Similar to the trajectory result, the hitting point prediction result was also visualized through annotated video frames in real time, where the hitting point is represented by a yellow dot.

Finally, after obtaining all the results and data we need for the program, a way of data-presenting method is required so that users can understand and read the result easily in a short amount of time. The library that helps to achieve the plotting is Matplotlib. The trajectory data is converted to a NumPy array and plotted using matplotlib. The Matplotlib plot shows the trajectory, start, end points, and predicted future positions of the ball。 It provides a clear visualization of the ball's trajectory and predicted positions as shown in Fig. 4. Improving the graphical representation of predictions and trajectories could provide clearer insights into the ball's movement. The trajectory plot is useful for post-analysis and validation of the tracking system's performance and accuracy.

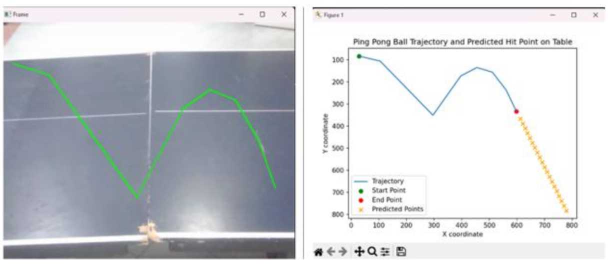


Fig. 4 Ping-pong ball's tracking and hitting point prediction.

According to the testing result conducted during this project, the ping pong ball tracking system based on the

OpenCV platform is able to perform all functionalities properly under various conditions. The system is capable of detecting a table tennis ball, drawing out the ball trajectory, and calculating the prediction of the hitting point based on it. It ends by performing program results through an easy-to-understand graph. The system effectively utilizes the Kalman filter for trajectory prediction, offering real-time tracking and visualization. However, its performance can be influenced by various factors, and further enhancements could improve robustness and accuracy.

4. Conclusion

The project concluded the process of development and design of a real-time ping pong ball tracking and hitting point system based on the OpenCV-Python platform. The system designed in this project achieved the expected outcome which is able to perform the basic functionalities of a table tennis tracking and prediction system and perform it through graph. However, the system's performance and accuracy are yet to be improved, especially in uncontrolled condition surroundings. Future development and improvement of this project could be in the direction of improving the system's performance in various surroundings such as poor lighting conditions. The accuracy of the system for detecting and capturing fast-movement balls can be improved. Another improvement that could be made in the future is to increase the data collection function so that multiple sets of data can be collected at the same time for comparison. Last but not least, a machine learning function could be added to the system so that the system can enhance its performance from time to time by studying the former data and results.

References

1. M. Mokayef, L. Q. Jian, B. Naeem, M. Mokayef, M. A. Khan, and M. A. Summakieh, Embedded Table Tennis Ball Launcher with a Trajectory Path Analyser for Junior Players, 2023.
2. H.-I. Yu, S.-C. Hong, and T.-Y. Ju, Low-cost system for real-time detection of the ball-table impact position on ping-pong table, Applied Acoustics, Vol. 195, 2022.
3. P. Blank, T. Kautz, and B. M. Eskofier, Ball impact localization on table tennis rackets using piezo-electric sensors, in Proceedings of the 2016 ACM International Symposium on Wearable Computers, 2016, pp. 72-79.
4. G. Byrd, 21st Century Pong, Computer, Vol. 48, No. 10, pp. 80-84, 2015.
5. Y. Wang and L. Wang, Retracted Machine Vision-Based Ping Pong Ball Rotation Trajectory Tracking Algorithm, Computational Intelligence and Neuroscience, Vol. 2022, No. 1, pp. 3835649, 2022.
6. Y.-F. Ji, J.-W. Zhang, Z.-h. Shi, M.-H. Liu, and J. Ren, Research on real–time tracking of table tennis ball based on machine learning with low-speed camera, Systems Science & Control Engineering, Vol. 6, No. 1, 2018.
7. H.-I. Lin and Y.-C. Huang, Ball trajectory tracking and prediction for a ping-pong robot, 2019 9th International Conference on Information Science and Technology (ICIST), 2019: IEEE, pp. 222-227.
8. M. Mokayef, A. K. MKA, A. S. MHD, L. Q. Jian, and M. Mokayef, Embedded Ball Launcher with Trajectory Path Analysis for Empowering Junior Table Tennis Players, Journal of Advances in Artificial Life Robotics, Vol. 4, No. 1, pp. 27-34, 2023.
9. K. Belal et al., Table Tennis Tournament Scores and Statistics Web Application, in 人工生命とロボットに関する国際会議予稿集, 2023, vol. 28: 株式会社 ALife Robotics, pp. 786-793.
10. L. Wang, Z. Zhou, and Q. Zou, Analysis system for table tennis techniques and tactics using data mining, Soft Computing, Vol. 27, No. 19, pp. 14269-14284, 2023.
11. P. I. Muniz Zavala, Comparison of methods for table tennis ball prediction, University of Sheffield, 2022.
12. K. Mikolajczyk, B. Leibe, and B. Schiele, Local features for object class recognition, Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, 2005, Vol. 2: IEEE, pp. 1792-1799.

Authors Introduction

Dr. Mastaneh Mokayef

Her PhD from Wireless Communication Centre Faculty of Electrical Engineering in University Technology Malaysia (UTM). She has many research papers.

Mr. MHD Amen Summakieh

He received the B.Eng. degree (Hons.) in communication and electronics engineering from UCSI University, Malaysia, in 2016, and the M.Eng.Sc. degree from Multimedia University, Malaysia, in 2020. His research interests include heterogeneous LTE-advanced cellular networks, user association, metaheuristic algorithms, and antennas design.

Dr. M. KA. Ahamed Khan

He pursued a PhD in Robotics, Power Electronics, and Controls in the United States and holds Chartered Engineer (CEng) in the UK. He is a Senior member of the IEEE in the USA and a member of MIET in the UK. He is also the past chair for IEEE RAS Malaysia chapter.

Dr. Sanjoy Kumar Debnath

He completed Ph.D. from the Faculty of Electrical & Electronic Engineering, Universiti Tun Hussein Onn Malaysia (UTHM) in 2022. He is currently working as an Assistant Professor at the Chitkara University Institute of Engineering and Technology (CUIET), Chitkara University, Punjab, India.

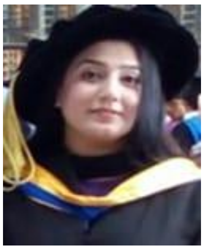
Mr. Miad Mokayef

He currently serves as the Head Coach of the Iranian Table Tennis National Team. He obtained his master's degree in Sport and Exercise Physiology from the Science and Research University in Iran.

Dr. Abdul Qayyum

He did his PhD in Electrical and Electronics Engineering from Universiti Teknologi PETRONAS, Malaysia 2018. He developed deep learning based algorithms for depth estimation of vegetation, trees near power lines for Tenaga Nasional Berhad (TNB) and Sabah Electric Supply Berhad (SESB) under the ministry of Green , Water and Technology (KeTTHA) Malaysia. He also developed a prototype for Vital signs (heart rate, breathing rate, Sp02) estimation and assessment of stroke and Arterial fibrillation (AF) using face video analytic based on deep leaning models. Earlier, he had completed his bachelor's in computer engineering and Master in Electronic Engineering from Pakistan. Besides, he gained one-year industrial experience while working as a BSS engineer for Huawei, Pakistan. He had also taught several courses under electrical, specifically, signal processing domain for 7 years in various public and private universities in Pakistan. He was working as a research scientist in CISIR, UTP for less than one year and was developed deep learning algorithms for brain signal (EEG) classification and reconstruction, remote sensing image segmentation and biomedical image analysis. He is associated with burgundy university France as a Post Doc researcher and working on the cardiac MRI images using deep learning approach. He is also working as a consultant in various projects involving deep learning models in Big Data, Vital Sign Estimation, IoT and BCI applications. Currently he is associated with National Heart and Lung Institute, Imperial College London, UK.

Dr. Moona Mazher



She received her Ph.D. in Computer Engineering and Mathematics from the University of Rovira i Virgili, and works in UCL, UK.

Dr. Sew Sun Tiang

She is an Assistant Professor in Faculty of Engineering at UCSI University in Malaysia. She received her PhD in Electrical and Electronic Engineering from Universiti Sains Malaysia in 2014. Her research interests are optimization and antenna design.

Dr. Wei Hong Lim

He is an Assistant Professor in Faculty of Engineering at UCSI University in Malaysia. He received his PhD in Computational Intelligence from Universiti Sains Malaysia in 2014. His research interests are optimization and artificial intelligence.

Dr. Takao Ito

He is Professor of Management of Technology (MOT) in Graduate School of Engineering at Hiroshima University. His current research interests include automata theory, artificial intelligence, systems control, quantitative analysis of interfirm relationships using graph theory, and engineering approach of organizational structures using complex systems theory.

An Innovative Deep Learning Technique to Identify Potato Illness

Abdul Majid Soomro[1], Muhammad Haseeb Asghar[2], Sanjoy Kumar Debnath[3], Susama Bagchi*[3], Awad Bin Naeem[2], M.K. A. Ahamed Khan[4], Mastaneh Mokayef [4], Takao Ito[5]

[1]Department of Computing, National University of Modern Languages, Multan Campus, Pakistan
**Email: abdulmajeed.soormo@numl.edu.pk*

[2]Department of Computer Science, National College of Business Administration & Economics, Multan, Pakistan

[3]Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

[4]Faculty of Engineering, UCSI University, Malaysia.

**Email: sanjoy.kumar@chitkara.edu.in, bagchi.susama@gmail.com*
https://www.chitkara.edu.in

[5]Hiroshima University, Japan
itotakao@hiroshima-u.ac.jp

∗Corresponding Author

Abstract

Potato cultivation is important for world food security as it itself is attacked by a great number of diseases like early blight and late blight, which cause a lot of damage to the yield and quality of the crop. But deep learning offers a great opportunity to address these disease detections; however, how effective this will be in the potato-growing environment in Pakistan is still not known. This research is designed to evaluate the convolutional neural network (CNN) by building custom datasets that denote the local disease description. The ultimate intention is to develop a high-accuracy, reliable disease detection model that will suit the particular needs of Pakistan. The project, therefore, tries to address data imbalance with the use of the synthetic minority oversampling technique (SMOTE) and develop a CNN architecture that is optimized to provide high diagnostic accuracy. Acquired through real-world pictures, the assessment of the model's performance shows significant progress in detecting potato diseases. This research can give innovative and productive locally useful solutions, which might transform the management of diseases for Pakistani farmers while improving food security and economic stability. These deep learning systems also need to be context-sensitive and reliable, which in turn would help preserve long-term agricultural productivity in Pakistan and beyond.

Keywords: CNN, Potato Disease, SMOTE, Deep Learning

1. Introduction

For millions of people worldwide, growing crops is more than simply an agricultural endeavor. Crop production is intricately woven with economies and cultures, from large-scale commercial agricultural operations to rural subsistence farmers [1]. This serves as a link to the inherited cultural history and customs of many people. Crop cultivation is a dynamic and resilient alternative to other sources of income because it enables people to feed themselves while contributing to their communities [2]. Potatoes are important in agriculture. They represent the tenacity and resiliency of agriculture and are more than simply a simple meal. Both smallholder farmers and major commercial producers can grow potatoes because of their well-known capacity to adapt to a range of soil types and temperatures [3]. Rich in carbohydrates, vitamins, and minerals, these tubers are nutritional powerhouses that greatly enhance dietary variety and global food security. Moreover, the efficient cultivation of potatoes in terms of land and resource utilization demonstrates their value in feeding the world's growing population [4]. On the lowest leaves of the afflicted plants, this fungal pathogen first manifests as tiny black lesions. These lesions have a distinctive concentric ring pattern and grow similarly to a target [5]. As the disease worsens and leaves wither and die, the plant's capacity to photosynthesize and grow is diminished. In addition to causing aesthetic damage, early blight may have negative financial and food security effects. Reduced photosynthesis due to leaf loss leads to lower yields because plants invest less energy in tuber growth. Moreover, damaged potatoes are usually of inferior quality, which lowers their market worth. In many regions, early blight, which grows best under warm, humid conditions, is a persistent danger to potato harvests [6]. Owing to its preference for cold, humid environments, which are prevalent in many countries that cultivate potatoes, Late Blight is more destructive than Early Blight [7]. Late blight is characterized by clear symptoms and indicators. Large, cotton-like, water-soaked sores that are often encircled by white, fuzzy growth, are the result of infection. The plant rapidly loses its leaves as the illness worsens and the lesions become black and necrotic. Moreover, tubers may rot owing to viral infiltration, making them unsuitable for human consumption. Late Blight poses a special difficulty because of its genetic variability and pathogenicity. Resistance mechanisms in potato cultivars may be swiftly overcome by the disease, thereby negating the value of previously successful management strategies. These declines in agricultural output have a significant effect worldwide. According to estimates, diseases such as early and late light reduce agricultural production worldwide by 16% annually. This is a startling statistic with important ramifications [8]. Food prices are affected by the increasing costs of food production owing to lower agricultural yields. Customers may find it more difficult to afford necessary food goods because of rising supermarket costs [9]. We would need to boost food production by approximately 70% to guarantee a consistent supply of food. However, the enduring danger of illnesses in vital commodities, such as potatoes, threatens our capacity to meet this rising need. The consequences are more noticeable in regions where potato growth is a significant industry. Potato crops are a major source of revenue for both smallholders and subsistence farmers. These delicate ecosystems are upset when illnesses occur, leading to a shortage of food, hardship on the economic front, and even forced migration. The significance of precise disease detection in agriculture cannot be overstated, especially with regard to potato diseases, such as early light and late light. Reducing the effects of these diseases on agricultural output and overall food security requires early and precise diagnosis. To address this need, several approaches and technologies have been developed, which may greatly enhance the detection of agricultural diseases [10].

The study's issue statement is that although it is a vital component of Pakistan's food security, potatoes are vulnerable to diseases such as Early Blight and Late Blight, which have a significantly lower yield. Early detection is essential for sustaining production since chronic diseases account for over 16 percent of the worldwide decline in agricultural output. Plant village datasets have been the main source of research on disease diagnosis using deep learning techniques. However, the dataset may only be applicable in some places, such as Pakistan, owing to environmental changes. Furthermore, it could be challenging to precisely detect illnesses in Pakistan's potato crop using 2155 potato photos, the bulk of which are from the United States. Current methods for detecting illnesses are time-consuming, particularly in isolated rural locations. In Pakistan's agricultural environment, this problem emphasizes the urgent need for an efficient machine-learning-based solution that can identify diseases early and provide faster, more precise diagnostics.

Because it has the potential to solve significant concerns in agriculture and food security, this study on deep-learning-based potato disease diagnostics is significant. The use of deep learning models in this study may enhance the management of potato crops, and therefore, global food security. The precise and effective disease detection capabilities of these models may reduce the monetary losses caused by illnesses and increase agricultural productivity. Furthermore, the agricultural sector, especially precision agriculture, was significantly affected by this discovery. Farmers are able to make data-driven choices because of cutting-edge technologies, such as deep learning. Thus, agricultural operations are generally more profitable, resource allocation is more effective, and the environmental impact is lessened. Furthermore, the influence of this study may extend beyond potato crops to other crops with comparable disease signs. When customized for various crops, deep learning models may provide a scalable, precise, and affordable disease detection method. Farmers worldwide may benefit from this technology by using sustainable agricultural methods to protect their crops.

The primary goals of this study were as follows: 1) To enhance Convolutional Neural Network (CNN) models for accurate and effective potato disease diagnosis in data-poor locations. 2) To examine how the amount and quality of training data impact the ability of machine learning to identify early and late blight. 3) Examine how the accuracy of diagnosing potato diseases is affected by farmers' experience and low latency. 1) In areas with little data, such as Pakistan, how can CNN-based deep learning increase the precision of potato disease detection? 2) How does the accuracy of machine learning models used to identify potato illnesses, such as early and late light, depend on the amount and caliber of the training data? 3) What effects do low latency, and a lack of domain expertise have on the precision of potato disease detection, and how can these issues be resolved?

2. Material and Methods

This section contains the methodology of the study.

2.1. *Research Methodology*

We require a dataset that includes images of potato leaves for our study. After finding this dataset on Google Drive, we had to ensure that it complied with our specifications, even though it was a helpful resource. Label management is an essential step in the data-preparation process. Our dataset had three labels for each leaf: "healthy," "late blight," and "early blight." early blight. Similar to the descriptions, these labels provide details of the state of the leaf at that moment. However, to enable our computer program to identify them more quickly, we reduced the number of labels. In place of

"Healthy," we used "Yes" to denote a leaf that is healthy and "No" to indicate a damaged leaf. Consequently, our computer model processed the data considerably easier. We divided our dataset into two parts: training and testing. During the training phase, our computer model gained knowledge from the data and gradually increased its intelligence. A CNN or a deep learning model was used. CNNs are skilled at comprehending pictures, such as these leaves. We forced our model to practice it extensively by repeatedly using this approach. Compared with running the model over the data only once, this approach is much more successful. With every "epoch," our model became more adept at identifying and assessing pictures. Finally, we wanted to determine how well our model worked. Therefore, we randomly selected a leaf photograph from the test data. The model was asked to identify whether the leaves were ill or healthy. We ascertained whether our model had been properly trained using this test. Potato illnesses were correctly identified by the model. Further tests were performed to ensure that our model was correct. Fig. 1 shows a flow chart of the research methodology.

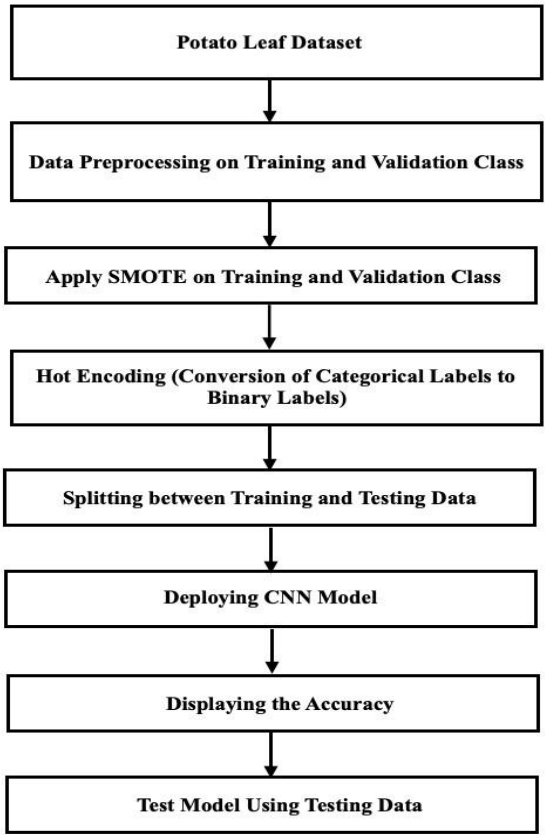


Fig. 1. Flow Chart of Research Methodology.

2.2. *Data Collection*

This study specifically focused on three distinct potato varieties: Coroda, Mozika, and Sante. These varieties are commonly cultivated in the Okara district of Central Punjab, a region known for substantial potato production. To ensure data diversity, potatoes were grown in rows and images and videos were captured under varying environmental conditions, replicating real-world scenarios. To create a valuable resource for detecting potato diseases in Pakistan's Central Punjab region, researchers developed the Potato Leaf Dataset (PLD). They collected real-time videos and images using different devices including mobile phones, digital cameras, and drones. These devices were carefully selected to capture data from various angles and distances. When capturing images with mobile phones and digital cameras, they were positioned quite close, 1-2 feet away from the potato plants. Drones, on the other hand, were flown at a higher altitude, approximately 5-10 feet, to avoid any distortion caused by the movement of the plant leaves. This ensured that the dataset contained high-quality images. To make the dataset even more useful, images and videos were taken under different environmental conditions, similar to what farmers experience in the real world. This diversity in conditions helps train models to effectively recognize diseases under various circumstances. Fig. 2 shows a sample image of the training class.

Fig. 2. Sample Images from the Training Class.

2.3. *Population of Sampling*

Table 1 shows a summary of the Potato Leaf Dataset, where the expert team of plant Pathologist finalized 4,062 potato leaf images in the potato leaf dataset (PLD). These images represent both healthy and diseased leaves, respectively. Owing to the large volume of data for various potato plant diseases, the healthy and infected leaves were annotated by the researchers using the LabelMe tool. To facilitate the development of a robust machine learning model for disease detection, plant pathologists categorized the images into three distinct classes: early blight, late blight, and normal (healthy) leaves. The resulting distribution within the PLD dataset comprised 1,628 images for early blight, 1,414 images for late blight, and 1,020 images for healthy leaves.

Table 1 Summary of Potato Leaves Dataset.

Class Labels	Samples
Blight Early	1628
Healthy	1020
Late Blight	1414
Total Samples	4062

Table 2 and Table 3 show the test class potato leaves dataset before and after smote, the SMOTE technique has been widely used to handle a class imbalance in various machine learning applications, including image classification, text classification, and disease detection.

After applying the SMOTE technique, our balanced dataset improved the performance of our model and increased its accuracy in detecting potato disease. These annotated and labeled images were used to train a deep learning model for potato disease detection.

Table 2 Testing Class Potato Leaves Dataset.
(Before SMOTE)

Class Labels	Quantity
Blight Early	1132
Healthy	1303
Late Blight	816
Total Quantity	3251

Table 3 Validation Class Potato Leaves Dataset.
(Befoe SMOTE)

Class Name	Quantity
Early Blight	163
Late Blight	151
Healthy	102
Total Quantity	416

Table 4 shows the training and validation class potato leaf dataset after smoothing, which ensured the availability of high-quality annotated images for training the model, resulting in improved accuracy and reliability of the detection system.

Table 4 Training and Validation Class Potato Leaves Dataset.
(After SMOTE)

Class Name	Labels	Quantity
Training	Blight Early Blight, Healthy Late	1303,1303,1303
Validation	Blight Early Blight, Healthy Late	1303,1303,1303
	Total Quantity	7818

A CNN is a deep learning model that uses interconnected neurons to detect and learn hierarchical features from images, making it effective in tasks such as image classification, object detection, and facial recognition. SMOTE addresses class imbalance in datasets by generating synthetic examples for minority classes, thereby preventing biases in models such as disease detection and fraud detection. One-hot encoding is a data preprocessing technique that converts categorical data into a numerical format for machine-learning algorithms, particularly in tasks such as natural language processing and image classification.

3. Results and Discussion

We aimed to evaluate the performance of our deep learning model for potato disease detection under various conditions. We conducted a series of experiments to assess the accuracy and reliability of our model. Fig. 3 shows model accuracy. These experiments included evaluating the model with and without the synthetic minority oversampling technique (SMOTE) and varying the dropout rates. The results of these tests shed light on the behavior of the model in different scenarios. Before SMOTE and setting the dropout value to 0.25 accuracy the training fluctuated between 0.94 and 0.99. This means that during training, the model performed well, with accuracies ranging from 94% to 99%. However, during validation, the accuracy of the model was slightly lower, fluctuating between 0.80 and 0.85. This indicates that when the model was evaluated on new unseen data (validation), its performance was slightly lower than that of the training data.

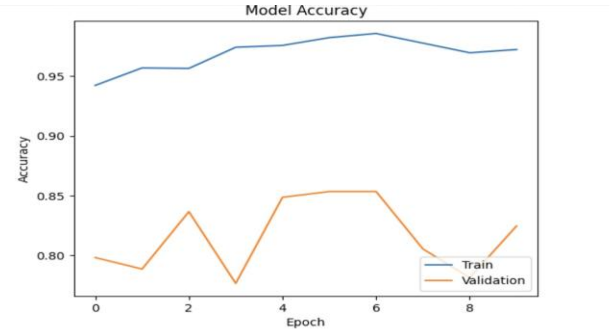


Fig. 3. Model Accuracy Graph.

Fig. 4 shows the model loss graph. During training, it varied between 0.1 and 0.2. A lower loss value indicated that the model fits the training data well. However, during validation, the model loss was higher, ranging from 0.6 0.9. This suggests that the performance of the model is not as good when evaluated on new data.

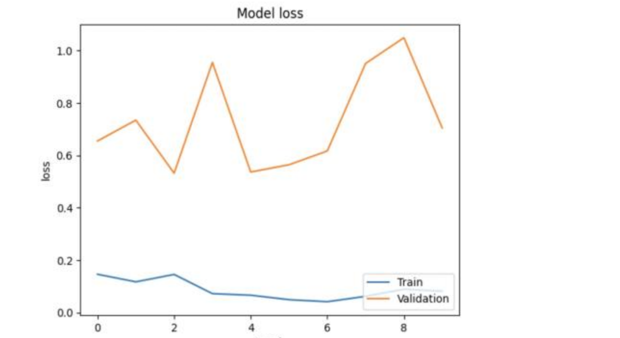


Fig. 4. Model Loss Graph

Fig. 5 shows the model accuracy after SMOTE, and with a dropout value of 0.25, it is evident that there was a notable difference between the training loss and accuracy in the deep learning model. In terms of the model accuracy, during training, it fluctuated between 0.94 and 0.99. This means that during training, the model performed well, with accuracies ranging from 94% to 99%.

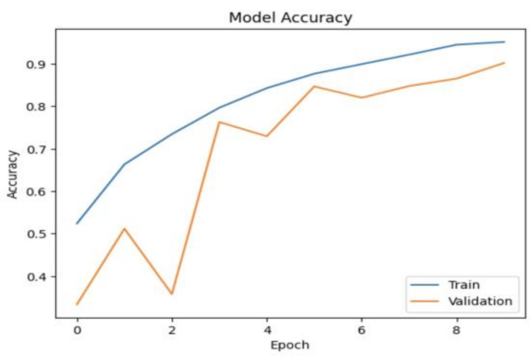


Fig. 5. Model Accuracy Graph after SMOTE.

Fig. 6 shows the model loss graph after SMOTE. However, during validation, the model's accuracy was slightly lower, fluctuating between 0.80 and 0.85. This indicates that when the model was evaluated on new unseen data (validation), its performance was slightly lower than that of the training data.

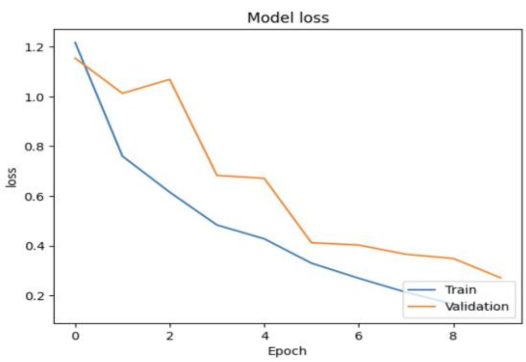


Fig. 6. Model Loss Graph after SMOTE.

Overall, applying SMOTE in combination with a dropout rate of 0.25 played a crucial role in enhancing the model's accuracy and reducing its loss, making it a promising approach for improving disease detection in potato leaves. The increased capacity of the model to classify healthy and diseased leaves can have a positive impact on potato crop management and yield. Although the above parameters demonstrated great results, fine-tuning of the parameters may result in an optimal model. Fig. 7 shows the model accuracy with SMOTE applied to balance the dataset, and a dropout rate of 0.35, which exhibited substantial improvements in accuracy and loss. The accuracy for validation increased from 0.55 to 0.95, indicating that the model became more proficient at correctly identifying healthy and diseased potato leaves. Training accuracy improved from 0.1 to 0.99, highlighting the model's effectiveness in learning from the data.

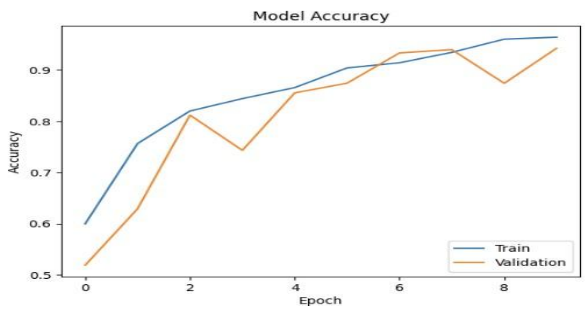


Fig. 7. Model Accuracy Graph.

Fig. 8 shows the model loss graph with SMOTE. In terms of loss, which measures the extent of errors, the validation loss decreased significantly from 1 to 0.1. The training loss also experienced a remarkable reduction, dropping from 0.99 0.05. These lower loss values indicate that the model is highly efficient in making precise predictions and minimizing errors.

The combination of SMOTE and a dropout rate of 0.35 resulted in a model with optimal performance. This not only increased the accuracy but also significantly reduced the loss, making it a highly effective tool for potato disease detection and classification. This optimized model makes a substantial contribution to potato crop management and disease control. Table 5 shows the results for each "epoch," which is a complete cycle through the training data. Loss" represents how well the model performs. When training started, the loss was relatively high (1.04 in the first epoch).

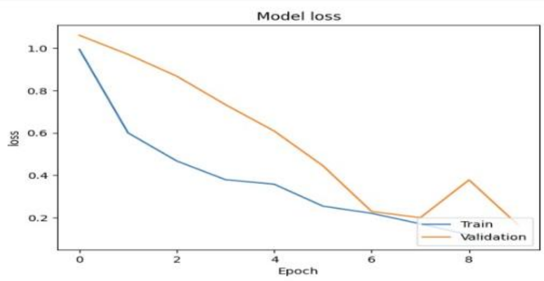


Fig. 8. Model Loss Graph.

Table 5 Validation Class Potato Leaves Dataset.

Epoch	Loss	Improvement
1	1.04	Improved
2	0.72	Improved
3	0.60	Improved
4	0.44	Improved
5	0.35	Improved
6	0.30	Improved
7	0.24	Improved
8	0.19	Improved
9	0.14	Improved
10	0.10	Improved

This means that the model was not very good at identifying diseases at the beginning. As the training continued (from epochs 1 to 10), it was observed that the validation loss consistently decreased. This is beneficial because the model is better at its job. By the time we reached the 10th epoch, the loss was only 0.10, which is much lower than the initial value. This suggests that the model is highly accurate in identifying potato leaf diseases. It is crucial to select the correct number of epochs during training to obtain the best performance from the model without making it overly complex.

The decision to limit the training of the deep learning model to only 10 epochs was made with careful consideration. Although training for more epochs can lead to further improvements in the model's performance, it is important to strike a balance. The use of a higher number of epochs can have drawbacks, particularly the risk of overfitting. In our case, this would mean that the model could become too specific to the examples it was trained on, potentially making it less effective in identifying new instances of potato leaf diseases. By observing the validation loss as the training progressed, it became clear that after 10 epochs, the loss reached a plateau and did not improve further. This suggests that the model learned as much as possible from available data. Pushing the model to train for more epochs might not have yielded significant benefits but could have increased the risk of overfitting. These insights collectively demonstrate the nuanced relationship among the number of epochs, model progression, and overfitting mitigation. By carefully selecting an appropriate number of epochs and incorporating techniques, such as dropout, we were able to utilize the power of deep learning to achieve accurate potato leaf disease classification while avoiding common pitfalls. This study aimed to identify better ways to identify diseases in potato plants, such as early and late blight. You see that potatoes are important for both food and life. Many people rely on them, but sometimes diseases can harm these potato plants, which means there is less food to eat, and farmers can lose money. The problem with these diseases is that they can damage not only the leaves but also the stems, roots, and potatoes themselves. When that happens, it is not good because we need potatoes to feed people worldwide. In addition, when diseases hurt potato crops, it can be expensive to grow. Looking into the future, it is important to know that an increasing number of people will live on Earth. To ensure that there is enough food for everyone, we need to produce a lot more food. Therefore, finding better ways to protect potato plants from disease is important. Current methods for identifying these diseases in potato plants are slow.

4. Conclusion and Future Work

We embarked on a journey to utilize the power of emerging deep-learning approaches to revolutionize the detection of potato plant diseases, particularly in early and late light. Our research journey can be divided into several key phases, starting with dataset collection, preprocessing, and ensuring balanced data using the SMOTE technique. We then delved into model development, where we defined a Convolutional Neural Network (CNN) architecture tailored to our task. We achieved impressive accuracy by training our model, thereby demonstrating the potential of deep learning in agriculture. Our study culminated in the successful implementation of our model for practical disease detection, which is evident in the high accuracy achieved on a sample image. This marks a significant step towards early disease identification and, subsequently, better crop management and higher yields for potato farmers. Although our study has made significant strides in the field of potato disease detection using deep learning, some areas warrant further exploration. The limitations of available datasets, especially in the context of specific regions such as Pakistan, point to the need for region-specific data-collection efforts. Additionally, research can focus on optimizing the model architectures and training processes to achieve even higher accuracy levels. Future research in the field of potato disease detection using deep learning should prioritize the collection of diverse region-specific datasets. This involves capturing images and data from various potato-growing regions in Pakistan that encompass different potato varieties and environmental conditions. The creation of a more extensive dataset will enable deep learning models to generalize and improve their accuracy in real-world settings. Combining these data with deep learning models can lead to a holistic approach disease management, allowing for proactive measures to prevent outbreaks. As deep learning models evolve, efforts should be made to make these solutions more accessible to farmers, especially in the remote areas of Pakistan. User-friendly interfaces and mobile applications can be developed to enable farmers to capture images, diagnose diseases, and receive recommendations directly from their smartphones. This approach can democratize advanced technology and empower farmers to make informed decisions regarding their potato crops.

References

1 Arshad M. S., Rehman U. A., & Fraz M. M. (2021, May) Plant disease identification using transfer learning, 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2) pp. 1-5, IEEE.

2 Ghosal S., & Sarkar K. (2020, February) Rice leaf disease classification using CNN with transfer learning, 2020 IEEE Calcutta Conference (CALCON) pp. 230236, IEEE.

3 Gupta S., Garg G., Mishra P., & Joshi R. C., (2020, December) Cmd: an efficient crop disease detection and pesticide recommendation system using mobile vision and deep learning, Proceedings of International Conference on Big Data, Machine Learning and their Applications: ICBMA 2019, pp. 295-305, Singapore: Springer Singapore.

4 Hou C., Zhuang J., Tang Y., He Y., Miao A., Huang H., & Luo S. (2021) Recognition of early blight and late blight diseases on potato leaves based on graph cut segmentation, Journal of Agriculture and Food Research, 5, 100154.

5 Islam M., Dinh A., Wahid K., & Bhowmik P. (2017, April) Detection of potato diseases using image segmentation and multiclass support vector machine, 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), pp. 1-4, IEEE.

6 Kamble J. K. (2018, February) Plant disease detector, 2018 International Conference On Advances in Communication and Computing Technology (ICACCT), pp. 97-101, IEEE.

7 Kosamkar P. K., Kulkarni V. Y., Mantri, K., Rudrawar S., Salmpuria S., & Gadekar N. (2018, August) Leaf disease detection and recommendation of pesticides using convolution neural network, 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), pp. 1-4, IEEE.

8 Kumar S., & Shukla A. (2022) Automatic Grading of Potato Leaf using Machine Learning & Computer Vision.

9 Kumar S. (2021) Plant disease detection using CNN, Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(12), 2106-2112.

10 Militante S. V., Gerardo B. D., & Dionisio N. V. (2019, October) Plant leaf detection and disease recognition using deep learning, 2019 IEEE Eurasia conference on IOT, communication and engineering (ECICE), pp. 579-582, IEEE.

Authors Introduction

Dr. ABDUL MAJID SOOMRO

He has over 20 years of teaching experience and is now a day HOD at NUML University. He possesses a robust enthusiasm for research and innovation in the fields of information science, information technologies, and IT management, with a particular emphasis on IT/IS research domains such as machine learning, data sciences, AI, deep learning, and network computing applications, all aimed at addressing practical network issues. He was awarded Ph.D. in Information Technology from Universiti Tun Hussein Onn Malaysia.

Muhammad Haseeb Asghar

He received his Mphil. degree in Computer science in 2023 from NCBA&E Multan in Pakistan.

DR. SANJOY KUMAR DEBNATH

He completed Ph.D. from the Faculty of Electrical & Electronic Engineering, Universiti Tun Hussein Onn Malaysia (UTHM) in 2022. He is a “Graduate Engineer” from the Board of Engineers Malaysia. He is currently working as an Assistant Professor at the Chitkara University Institute of Engineering and Technology (CUIET), Chitkara University, Punjab, India.

DR. SUSAMA BAGCHI

She received a Ph.D. in Electrical Engineering from Universiti Tun Hussein Onn Malaysia in 2022. She completed an M.Sc. in Modern Communication Technologies with Business Management from the University of Sussex, UK, in 2007 and a bachelor’s degree (B.E.) in Electronics and Communication Engineering (ECE) from the University of Burdwan, India, in 2002. She is currently working as an Associate Professor at the Chitkara University Institute of Engineering and Technology (CUIET), Chitkara University, Punjab, India. She has 15 years of experience in engineering management, academia, and research.

AWAD BIN NAEEM

He completed his Master's degree in Computer Science from NCBA&E Multan in September 2022, focusing on smart transportation systems, machine learning, and deep learning, which sparked his interest in the subject. Awad's Master's studies focused on Vehicle-to-Everything (V2X) Communication and the development of an intelligent transportation system for different types of vehicles. He also proposed an Intelligent Transportation System framework for Autonomous, Non-Autonomous, and prioritized vehicles.

Dr. M. K. A. Ahamed Khan

He pursued a PhD in Robotics, Power Electronics, and Controls in the United States and holds Chartered Engineer (CEng) in the UK. He is a Senior member of the IEEE in the USA and a member of MIET in the UK. He is also the past chair for IEEE RAS.

Dr. Mastaneh Mokayef

Her PhD from Wireless Communication Centre Faculty of Electrical Engineering in University Technology Malaysia (UTM) in 2014. She has many research papers.

Dr. Takao Ito

He is Professor of Management of Technology (MOT) in Graduate School of Engineering at Hiroshima University. His current research interests include automata theory, artificial intelligence, systems control, quantitative analysis of interfirm relationships using graph theory, and engineering approach of organizational structures using complex systems theory

A Wearable Walking Support System Design and Simulation

Omar Ayaman Yehiya[1], M. K. A. Ahamed Khan*[1], Mastaneh Mokayef [1], Amar Ridzuan Bin Abd Hamid[1]
[1]UCSI University, Faculty of Engineering, Taman Connaught, 56000 Malaysia
**Email: Mohamedkhan@ucsiuniversity.edu.my*

Abdul Qayyum[2], Moona Mazher[3]
[2]Imperial College, London, United Kingdom, Email: qayyum@imperial.ac.uk
[3]Centre for Medical Image Computing, Department of Computer Science, University College London, UK

Susama Bagchi[4], Sanjoy Kumar Debnath[4]
[4]Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

Takao Ito[5]
[5]Hiroshima University, Japan
itotakao@hiroshima-u.ac.jp
**Corresponding Author*

Abstract

This research article is focused on the development of a Robotic lower limb exoskeleton model using MATLAB Simulink. The primary aim is to design a dynamic and flexible exoskeleton capable of assisting individuals with lower limb impairments, thus enhancing their mobility and overall quality of life. The model incorporates realistic representations of the lower limb anatomy, encompassing thigh, shank, and foot segments, with carefully integrated joints, constraints, and actuators to emulate natural human motion. A closed-loop control strategy optimizes the exoskeleton's performance, ensuring safe and stable operation during walking and other activities. Extensive simulations are conducted to evaluate the exoskeleton's efficacy, analyzing key parameters such as joint angles, joint torques, and power consumption.

Keywords—component; Lower Limb Exoskeleton; Wearable Walking Support System; PID controller.

1 Introduction

Walking is a fundamental human movement, yet elderly individuals often encounter significant challenges in performing this basic activity due to the deterioration of their lower extremities. Muscle strength loss, often associated with aging, and weaker knee joints are common health problems faced by the elderly, impacting their ability to walk comfortably. Malaysia, with its growing aging population, underscores the urgency of developing walking support systems to meet the increasing demand for assistive devices. However, the country's diverse geography, featuring mountains and uneven terrain, poses additional difficulties for traditional walking support systems such as wheelchairs, canes, and crutches. To address these challenges, engineers, scientists, and doctors have developed various walking support devices, each with its advantages and limitations. Recently, wearable walking support systems have emerged as a promising solution. These robotic devices attach to the user's leg and employ sensors, actuators, and potentiometers to enhance lower extremity strength, protect knee joints, and reduce cardiac system stress during walking or other lower limb movements. This paper presents a novel wearable walking support system model equipped with a PID controller, intended to aid individuals with weakened lower extremities, whether due to aging or injuries. The proposed system delivers torque to the joints through actuators, providing support to the user during movement. Unlike existing linear actuator-based systems, this model offers improved mobility and comfort without the added inconvenience of carrying additional weight or restrictions while sitting. The simulation of the support system aims to showcase its potential as an effective and user-friendly solution.

2 Literature review

Gait analysis is a fundamental aspect of developing a wearable walking support system. It involves a comprehensive examination of the patterns of movement exhibited by the lower extremities during walking. A systematic review conducted by Benson et al. focused on wearable devices for gait analysis, specifically targeting three categories of individuals: healthy younger adults, older adults, and those with various pathologies. The review covered 61 articles, with 47 addressing gait analysis during walking, 13 during running, and one covering both walking and running. Notably, gait analysis revealed distinct differences in variability, regularity, and symmetry between older adults and healthy young adults during the gait cycle. Instrumented measurements of joint kinematics and kinetics lie at the core of gait analysis.

The information derived from gait analysis serves as a basis for developing the algorithm and structure of the support system. Gait analysis can be categorized based on its intended purpose. It is predominantly employed for clinical rehabilitation, helping individuals with lower extremity injuries or posture-related issues. Moreover, gait analysis finds extensive use in clinical research, where it serves to study specific conditions affecting groups of individuals or assess the impact of interventions. While the criteria for clinical testing and clinical research differ, both are essential in advancing our understanding of gait patterns and the development of effective walking support systems.

For the development of a wearable walking support system, gait analysis offers critical insights. It aids in understanding the user's unique walking pattern, including foot angles, joint range of motion, and center of gravity placement. This valuable information allows for the customization of support systems to meet individual needs. Furthermore, gait analysis helps identify areas of weakness or instability in the user's gait, providing a targeted approach for support. Additionally, by analyzing the user's gait improvements can be made to enhance system efficiency. Gait analysis is an indispensable tool in developing wearable walking support systems. Its role in understanding walking patterns, identifying areas of weakness, and measuring the effectiveness of support systems makes it invaluable in creating personalized, efficient, and effective solutions for individuals with lower extremity challenges. The systematic and comprehensive examination of gait patterns helps in advancing the field of assistive technologies and rehabilitation, ultimately improving the quality of life for those in need.

2.1 Motion monitoring system

Human motion tracking systems are the systems used to conduct gait analysis, due to their preciseness and capabilities. Human motion tracking systems can produce real-time data using modern sensor technologies. Human motion tracking systems are usually used either as a substitute rehabilitation environment that could be installed in the homes of patients or used to conduct gait analysis.

In the UK, 135,000 people experienced stroke in the year 2001and 2002, and required admission to hospital [1]. Therefore, the demand for technologies like the Human motion tracking systems increased heavily. Generally, the study classifies the human motion tracking system into four categories: Visual based systems, non-visual based systems, a combination of both, and Robot-aided tracking. The inertial accelerometer sensor would be suitable for a portable device due to its physical compatibility and lightweight. To collect human movement patterns and detailed information, sensors are used in non-visual based systems. Mostly they are mechanical based. Exoskeletons have a left-right pair, it is effective to wear based on one's comfort.

In general, it is important to have a lightweight material and an actuator, for the driving section of the robotic exoskeleton fixed on the knee part [2].

To support antigravity muscles on the lower extremities a model-based control algorithm without using biological signals, can be fixed to the knee joint for better support. Based on a human model one can calculate all the interaction forces acting on the left thigh moment [3].

2.2 Wearable walking support systems

Exoskeleton robots, also known as wearable walking support systems, have been the subject of development and research for several decades. One of the earliest exoskeletons, the Hardiman, was developed by General Electric in 1965, primarily for strength augmentation of the arms and legs [4]. Over the years, exoskeleton technology has advanced significantly, leading to the creation of various exoskeletons with diverse applications. One notable example is BLEEX, developed by Zoss, Kazerooni, and Chu at the University College of Berkeley in 2005, BLEEX is a lower limb exoskeleton that operates as the first energetically autonomous robot of its kind. Designed to be worn by users, BLEEX enables individuals to carry heavy payloads while maintaining walking endurance, making it valuable for load-carrying applications. Subsequently, Kazerooni, in collaboration with Ekso Bionics, created the HULC exoskeleton in 2009, which utilizes hydraulic power to allow soldiers to walk longer distances while carrying heavy loads [5]. Kazerooni's contributions have earned him the moniker "Father of modern exoskeleton." Exoskeletons have also proven to be beneficial beyond augmentation, with applications in support and assistance. For instance, Ikeuchi et al. from Honda developed a bodyweight support exoskeleton in 2009, reducing the user's perceived weight during walking by redistributing weight to the support structure [6]. In a later advancement, Honda created the Honda Stride Management Assist in 2014, which boosts user hip motion during walking and has demonstrated efficacy in assisting patients with Parkinson's disease by increasing step length [7]. Rehabilitation exoskeletons form another category, aiding patients undergoing therapy under specialized supervision. Notable examples include

BLEEX, developed by Kazerooni, Zoss, and Chu, and Kawamoto et al.'s HAL-3 (Hybrid Assistive Leg) [8]. To support the hip joint by providing flexion and extension torques, and to walk for extended periods carrying heavy loads, these devices are used. To understand well the exoskeletons, researchers used human gait analysis, which involves understanding various human walking movements and their mechanics behind and their involvement in to various joints [9], [10]. Such studies are pivotal in designing effective exoskeletons. For example, Miao et al. explored four human movements involving the lower limbs studying walking, running, jumping, and squatting, to determine suitable action states for exoskeleton therapy [11]. However, developing prototypes of exoskeletons can be prohibitively expensive due to the cost of fabrication and the uncertainty surrounding the prototype's efficacy.

To address this challenge, simulation using computer-aided design (CAD) software has emerged as a cost-effective alternative. Software like Solidworks, AutoCAD, and Fusion360 allows designers to create and simulate exoskeleton structures before actual fabrication, ensuring efficiency and effectiveness. Researchers have utilized SimMechanics in MATLAB Simulink [12], SolidWorks [13], and ADAMS in MATLAB for simulation purposes. The primary goal is to develop a simulation of a wearable walking support device (exoskeleton) design using Fusion360 software. The two- legged exoskeleton design aims to provide knee joint support using BLDC motors without relying on biological signals. By simulating the design in Fusion360, stress analysis, support efficacy, and other critical results will be obtained, reducing the need for costly physical prototypes. This approach contributes to the advancement of exoskeleton technology and facilitates the development of efficient and affordable walking support systems.

3 METHODOLOGY

3.1 PROCESS

In the upcoming chapter the main procedures to develop the wearable walking support system will be covered with explanation and illustrations. The chapter will include all the materials used, as well as the explanation of the wearable walking support device.

3.2 METHODS and MATERIALS

The system design is modelled as a CAD design. A control system was designed for the lower limb exoskeleton model, Specifically, a PID controller. The Model is then further simulated using MATLAB's add-on Simscape Multibody (formally SimMechanics), that provides a multibody simulation environment for 3D mechanical systems, such as Robotic systems. The PID controller was designed using the Newton-Euler equations to calculate the inverse dynamics of the Lower part.

3.3 WEARABLE WALKING SUPPORT SYSTEM

The process of designing the Wearable walking support system started with designing a 3D CAD design using Fusion360. The components of the exoskeleton included feet, thigh rods, knee rods, back, and holders for the control components. The model of the exoskeleton designed is illustrated in Fig. 1. The model is a two-legged exoskeleton, which is connected at waist. The model has 6 Degrees of Freedom (DOF) and joints with each leg containing three: hip, knee, and ankle joints. The CAD design does not include the actuator nor gear; however, the simulation model includes an actuator block.

Fig.1 CAD design of the lower part.

Following the Design of the Lower Limb Exoskeleton in Fusion 360, the file is imported into Autodesk Inventor. Autodesk was chosen due to a unique feature that is not present in many computer-aided design (CAD) applications/software, which is to allow the installation and implementation of the Simscape Multibody add-on. This feature provided a less complicated procedure to help convert the files from their original (step or stl) format to an (XML) format included with their design specifications, which can be imported into MATLAB's Simscape Multibody add-on to allow for simulation.

3.4 DESIGN AND SIMULATION PROCESS FLOW

The process of developing the Lower Limb Exoskeleton is explained in a block diagram as shown in Fig. 2. The process begins with designing a 3D CAD Model for the exoskeleton using suitable CAD software. Preferably, either Autodesk Inventor, SolidWorks, or Creo, as the mentioned software is the only software that supports the Simscape Multibody add- on. The next step is to import the design into MATLAB using the import function "smimport ('xml file name');". Once the xml file of the exoskeleton model is imported, a block model of the exoskeleton is created in Simulink. The block model is then rearranged with the addition of the actuators, the joints, and the control systems for each joint. There are 6 DOF/joints in the lower limb exoskeleton, 3 DOF/joints in each leg with one degree of freedom in each joint of the three joints hips, knee, and ankle. The joints revolve as the joints should move only in the sagittal plane. Once the block diagram as Fig. 3 is rearranged, the simulation is run, and the results can be obtained. There are 6 DOF/joints in the lower limb exoskeleton, 3 DOF/joints in each leg with one degree of freedom in each joint of the three joints hips, knee, and ankle. The joints revolve as the joints should move only in the sagittal plane.

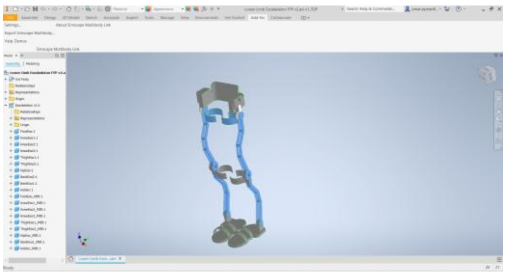


Fig.2 The CAD design of the Lower Limb Exoskeleton in Autodesk Inventor.

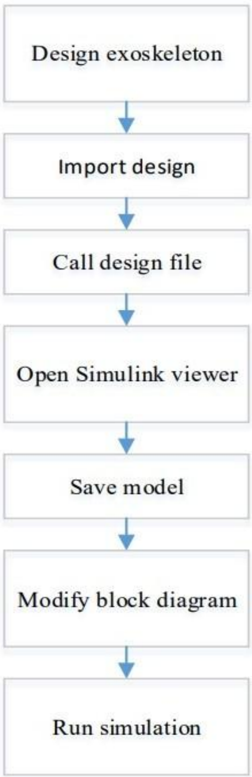


Fig. 3 Process flow of the Design and Simulation.

4 PID Controller

A PID controller has been selected to control the trajectory of the joints. PID controllers are very common due to their simplicity, also they consist of three basic control actions (proportional, integral, and derivative), each of which contributes to the overall control effort, other than simplicity, PID controllers provide stability. The error of the system is the difference between the desired and actual angular trajectory of each joint.

$$\theta_e = \theta_{Actual} - \theta_{Desired} \qquad (1)$$

The difference between the actual and desired of the PID are the error θ_e. The equation of PID controller is as follows:

$$\tau = K_p\theta_e + K_i\int_0^\tau \theta_e dt + K_d\frac{d\theta_e}{dt} \qquad (2)$$

The K_p, K_i, and K_d are the gains of the PID. The value of the gains determines the position of the joint; each joint requires a different gains value. The gains value can be obtained from the Ziegler-Nichols (Z-N) method for tuning the PID controllers. The method determines the PID parameters through identifying the critical gain and critical period of the system. The Pesudo code of the Z-N is explained in algorithm 1.

Algorithm 1. Pesudo code of the Z-N

1. Start
2. Set $K_i = 0$;
3. Set $K_d = 0$;
4. Set $K_p = K_u$;
5. Set step value for K_u as K_s;
6. **While** observe oscillation **do**;
   set value for K_s;
   Calculate $K_u = K_u + K_s$;
7. End **While**
8. Measure frequency of oscillation as T_u;
9. Determine PID parameters;
10. End

PID parameters are determined as follows,

$K_p = 0.7\times K_u$ (3)

$K_i = 1.75 \times (K_u/T_u)$ (4)

$K_d = (21 \times K_u \times T_u)/200$ (5)

where K_u is the value of K_p when the value of both K_i and K_d is equal to zero, while T_u is the frequency of the of the oscillation. Table 1 shows the gain values of each PID and the PID parameters values for each joint.

Table 1 The gain values of each PID and the PID parameters values for each joint.

Joints	Kp	Ki	Kd
Hip	46	142.213	4.28
Knee	35	103.578	3.35
Ankle	5.14	15.532	0.5125

5 RESULT and DISCUSSION

5.1 JOINT TRAJECTORY

Fig. 4, Fig. 5, and Fig. 6 show the desired and actual joint trajectories. Because the movement is symmetrical, only the findings for the joints of the left leg are shown. We can see that the joint positions are similar to the desired ones, although they are not identical. The collision of the foot exoskeleton with the ground is one explanation for such a disparity. This effect may be mitigated by increasing the derivative gain. A higher derivative gain, on the other hand, produces in excessive torque, making the bracing's trajectory unstable. Despite minor variations in the required location, the bracing stays stable.

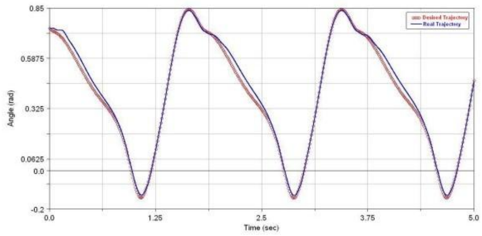


Fig.4 Angle(rad) x time(sec) for joint 1, left hip joint.

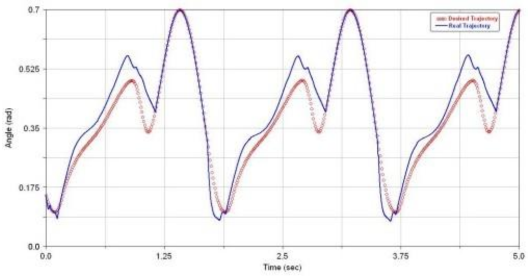


Fig.5 Angle(rad) x time(sec) for joint 3, left knee joint.

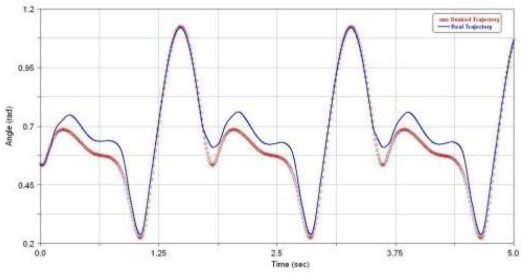


Fig.6 Angle(rad) x time(sec) for joint 5, left ankle joint.

5.2 INTERACTION FORCES

The interaction forces can be calculated, the behavior of forces is shown in Fig 7, Fig 8. The impact with the ground results in fast changes. According to the graphs, the majority of these variances occur during the same period.

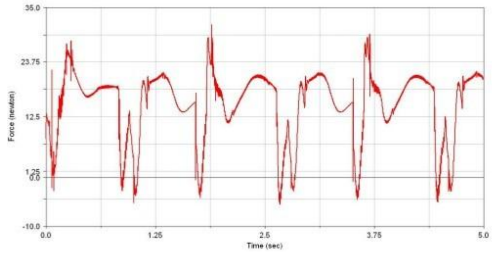


Fig.7 Interaction Torques acting at the torso.

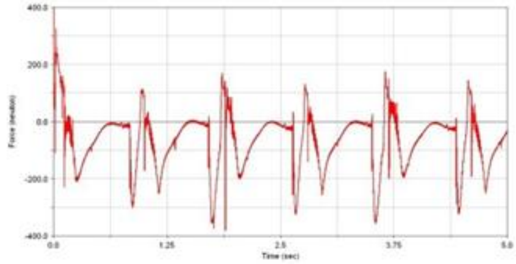


Fig.8 The interaction forces acting on the left thigh.

5.3 DISCUSSION

The results obtained show that the structure of the exoskeleton modelled using Simscape Multibody as a Simulink block diagram can be actuated at each joint. The process is done by adding an actuator block to actuate the joints, a sensor block, that is added as feedback from the output to form a control system. In addition to this, further kinematic analysis could be conducted by differentiating the position angle output into velocity which is then further differentiated to come out with the acceleration. The joint trajectories were simulated, in addition to the applied torques inside the joints being illustrated. The exoskeleton has shown reliability in the majority of the results. The results obtained show that the proposed exoskeleton model has the capabilities to support the walking movement of a human. Therefore, based on the results, the proposed wearable walking support system should be of assistance to individuals with weakened lower extremities, or individuals with lower limb impairments whether it is due to age or injuries. More research can be done to improve its output response.

6 CONCLUSION

In conclusion, this paper successfully developed and simulated a lower limb exoskeleton with PID control using MATLAB Simulink. The implementation of PID control allowed precise and responsive joint angle regulation, ensuring smooth and natural gait patterns. The simulation results demonstrated the exoskeleton's efficacy in tracking desired trajectories, responding to external disturbances, and minimizing tracking errors. The use of MATLAB Simulink provided a robust platform for modeling and tuning the PID controllers, enabling iterative improvements before physical prototyping and testing. The successful simulation of the exoskeleton signifies a significant advancement in wearable walking support systems, offering great potential for assisting individuals with mobility impairments. As future work, advanced control algorithms and real-world testing will further enhance the exoskeleton's performance and usability, revolutionizing the field of assistive technologies and rehabilitation.

References

1 Huiyu zhou, Huosheng Hu, Human motion tracking for rehabilitation— A survey, 2007

2 Dae-Hoon Moon, Donghan Kim, Young-Dae Hong, Development of a Single Leg Knee Exoskeleton and Sensing Knee Center of Rotation Change for Intention Detection, 2019.

3 Takahiko Nakamura, Kazuhiro Kosuge, Model-based walking support system with Wearable Walking Helper, Japan, 2018.

4 J. L. Pons, Wearable Robots: Biomechatronic Exoskeletons, Wiley, 2008.

5 H. Kazerooni et al., On the Control of the Berkely Lower Extremity Exoskeleton (BLEEX), Proceedings of IMECE2005 2005 ASME International Mechanical Engineering Congress and Exposition, 2005, Vol. 25, No. April, pp. 561–573.

6 Y. Ikeuchi, J. Ashihara, Y. Hiki, H. Kudoh, and T. Noda, Walking assist device with bodyweight support system, 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009, 2009, pp. 4073–4079.

7 T. N et al., Effect of honda stride management assist device (SMAD) on gait in patients with Parkinson's disease, Journal of Parkinson's Disease, Vol. 28, No. 2, 2016.

8 H. Kawamoto and Y. Sankai, Power Assist System HAL-3 for Gait Disorder Person, in Computers helping people with special needs, 2002, pp. 196–203.

9 J. Chen, X. Mui, F. Du, Z. Zhu, and D. Margenstern, Human lower limb kinematics and muscle biomechanics for exoskeleton design, Journal of Mechanical Engineering Research and Developments, 2017.

10 J. Chen, X. Mu, and F. Du, Biomechanics analysis of human lower limb during walking for exoskeleton design, Journal of Vibroengineering, Vol. 19, No. 7, pp. 5527–5539, 2017.

11 Y. Miao, F. Gao, and D. Pan, State classification and motion description for the lower extremity exoskeleton SJTU-EX, Journal of Bionic Engineering, Vol. 11, No. 2, pp. 249–258, 2014.

12 Y. Li et al., SolidWorks / SimMechanics-Based Lower Extremity Exoskeleton Modeling Procedure For Rehabilitation, World Congress on Medical Physics and Biomedical Engineering, IFMBE Proceedings 39, 2013, pp. 2058–2061.

13 N. A. Shaari, I. S. Isa, and T. C. Jun, Torque Analysis of The Lower Limb Exoskeleton Robot Design By Using Solidwork Software, ARPN Journal of Engineering and Applied Sciences, Vol. 10, No. 19, pp. 1–10, 2015.

Authors Introduction

Mr. Omar Ayaman Yehiya

He completed his Degree in Mechatronics Engineering at UCSI University and currently working as an engineer Egypt. He has working experience of one year.

Dr. M. K. A. Ahamed Khan

He pursued a PhD in Robotics, Power Electronics, and Controls in the United States and holds Chartered Engineer (CEng) in the UK. He is a Senior member of the IEEE in the USA and a member of MIET in the UK. He is also the past chair for IEEE RAS Malaysia chapter.

Dr. Mastaneh Mokayef

She has received her PhD from University Technology Malaysia (UTM) in 2014. Her research interests include: Wireless communications, spectrum sharing method, spectrum management, cellular communication systems and Antenna design.

Ts Amar Ridzuan Bin Abd Hamid

He has completed his master's degree from Universiti Putra Malaysia (UPM), Postgraduate Diploma in Tertiary Teaching (PGDTT) from UCSI University Malaysia. Currently he is the Head of MM Department in UCSI

Dr. Abdul Qayyum

He is currently working at National Heart and Lung Institute Imperial College London, UK. He received his Ph. D in electrical & electronics from Universiti Teknologi Petronas Malaysia.

Dr. Moona Mazher

She received her Ph.D. in Computer Engineering and Mathematics from the University of Rovira i Virgili, Spain and works in UCL, UK

Dr. Susama Bagchi

She received her Ph.D. in Electrical Engineering from Universiti Tun Hussein Onn Malaysia in 2022. She is currently working as an Associate Professor at the Chitkara University Institute of Engineering and Technology (CUIET), Chitkara University, Punjab, India. She has 15 years of experience in engineering management, academia, and research.

Dr. Sanjoy Kumar Debnath

He completed Ph.D. from the Faculty of Electrical & Electronic Engineering, Universiti Tun Hussein Onn Malaysia (UTHM) in 2022. He is currently working as an Assistant Professor at the Chitkara University Institute of Engineering and Technology (CUIET), Chitkara University, Punjab, India.

Dr. Takao Ito

He is Professor of Management of Technology (MOT) in Graduate School of Engineering at Hiroshima University. His current research interests include automata theory, artificial intelligence, systems control, quantitative analysis of interfirm relationships using graph theory, and engineering approach of organizational structures using complex systems theory.

A Floor Tiling Robotic System

Hue Chau Jieng, M. K. A. Ahamed Khan*, Mastaneh Mokayef, Amar Ridzuan Bin Abd Hamid
UCSI University, Faculty of Engineering, Taman Connaught, 56000 Malaysia

Abdul Qayyum
Imperial College, London, United Kingdom

Moona Mazher
Centre for Medical Image Computing, Department of Computer Science, University College London, UK

Susama Bagchi, Sanjoy Kumar Debnath
Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

Ito Takao
Graduate School of Advanced Science and Technology, Hiroshima University, Japan
Email: Mohamedkhan@ucsiuniversity.edu.my, a.qayyum@imperial.ac.uk, itotakao@hiroshima-u.ac.jp

** Corresponding Author*

Abstract

With the accelerated advancement of robot technology and sensor technology, construction challenges have become less difficult. The construction industry has been revolutionized by innovations in materials, equipment, and procedures, making it more efficient and safer. In recent years, however, an accelerated ageing society is compensating for the dearth of youthful and middle-aged workers. This paper proposes a Floor Tiling Robot robotic system that uses a vision-based solution to minimize labour-intensive, improve productivity, and increase the precision of the floor tiling process in order to reduce the material cost. The Floor Tiling Robot has implemented several systems, including a pneumatic vacuum suction system as a method for grasping floor tiles, a finite state machine as a method for robotic arm movement control algorithm and Canny Edge Detector algorithm as a method for floor tile positioning.

Keywords: Robotics, Automation, Computer Vision, IoT, Construction industry, Floor tiling robot

1. INTRODUCTION

The construction industry relies on floor tiling to provide functional and aesthetically appealing surfaces for a variety of spaces. Floor tiles contribute to the durability, sanitation, and overall atmosphere of residential and commercial buildings alike. Nevertheless, the manual process of floor tiling presents a number of obstacles and limitations.

Traditional methods of floor tiling rely largely on skilled labourers who lay each tile by hand. This procedure requires precise alignment, levelling, and uniform grout lines, which can be physically demanding and time-consuming. Moreover, manual tile installation is prone to human error, resulting in misalignments, irregular surfaces, and costly rework. There are a number of technologies and solutions able to implement robotic systems, however the author would like to choose the most suitable technology and solution to come out with the robotic systems for the user.

2. LITERATURE REVIEW

In 2014, Region-based Convolutional Neural Networks (R-CNN) models was published, which are used in computer vision and image processing. The R-CNN family recognizes the objects through their classification. The primary objective of R-CNN was to correctly locate the elements in the image [1]. Joseph Redmon and Ali Farhardi published YOLO v2, also known as YOLO 9000, in the same year as YOLO v1. This new YOLO model can detect over nine thousand object categories and can operate at 67 frames per second with 76.8% mAP on the Pascal VOC 2007 dataset [2]. Floor tile paving robot is a fully automated solution for paving large-format floor tiles at high speed and with consistently high quality. The compact, wireless system features all-wheel drive with an omni-directional mobile chassis. The visual measurement and positioning system with four cameras enable the tiles to be precisely aligned. To do this, the robot arm with vacuum grippers first vacuums the floor tile and rotates it 180 degrees. Then the robot moves back to apply the cement to the substrate. Finally, the tile is laid with millimetre precision on the fresh bed of adhesive [3]. This section discussed the technologies and solutions to implement the robotic system, Floor Tiling Robotic System. Fast R-CNN employs several innovations to improve training and testing speed while also increasing detection accuracy. Fast R-CNN trains the very deep VGG16 network 9x faster than R-CNN, is 213x faster at test-time, and achieves a higher mAP on PASCAL VOC 2012. Compared to SPPnet, Fast R-CNN trains VGG16 3x faster, tests 10x faster, and is more accurate [4].

Fast R-CNN employs several innovations to improve training and testing speed while also increasing detection accuracy. A computer interprets digital images as 2D or 3D matrices, where each value or pixel in the matrix represents the amplitude or "intensity" of the pixel. People are typically accustomed to interacting with 8-bit images, in which the amplitude value ranges from 0 to [5], [6]. Inverse kinematics can be accomplished, according to robot kinematics, when the rank of the system is less than the rank of the Jacobian matrix of each wheel of the robot, which reduces the degree of freedom of the robot's joints [7]. In this paper, the author has analysed several methods of manual floor tiling, overall [8], it has three manual floor tiling methods which are normally used in different applications. The manual floor tiling methods are thick-bed tiling (mortar bed tiling), thin set tilling (dry bond tiling), dry system. The authors have proposed a floor tiling robotic system called Floor Tiling Robot I System (FTR-I). Worldwide, over 10 billion m^2 of ceramics are consumed once a year. Gramazio Kohler Architects in partnership with ROB Technologies noticed that automatic tiling work can be of great value, they aimed to design a robot that can entirely replace human labour [9]. There are numerous varieties of industrial robots on the market, and each has its own benefits and drawbacks. Cartesian robots, SCARA robots, articulated robots, cylindrical robots, delta robots, polar robots, and collaborative robots are the most common forms of industrial robots [10], [11].

Control panel unit, main controller unit, mobile platform unit, collaborative robotic arm, end effector unit, and sensor unit comprise the FTR-1's mechanical structure as shown in Fig. 1. The control panel unit is a robotic arm control panel that serves as the FTR-1's interface for human interaction.

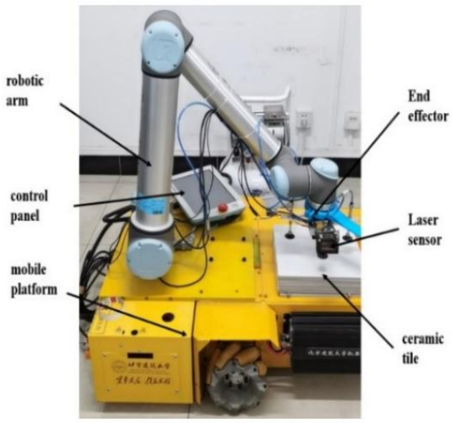


Fig. 1. FTR-I.

In this floor tiling robot, the authors have implemented a hierarchy robotic control system. As depicted in Fig. 2, the robotic system is composed of several layers: the user layer, the system control layer, the intermediate driving layer, and the hardware execution layer [12].

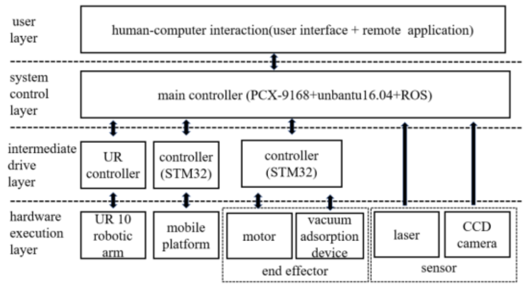


Fig. 2. Hierarchy Robotic System of FTR-I.

The RPN and the Fast R-CNN share convolutional computations, which drastically reduces the processing time [13]. In 2014, an improve edge detection Canny algorithm published. Particle swarm optimization (PSO) has been applied as a control algorithm for a number of selected mathematical models [14]. Apart from performing well basic functions like moving forward and backward, turning left and right, the robot is able to detect preceding obstacle, stop movement, and then identify suitable clear way for avoiding the obstacle [15].

3. RESEARCH METHODOLOGY

This section describes the design of a floor tiling robotic system that uses robotics to automate the floor tiling operation. The process of tiling a floor can require a lot of effort and time; this invention was created to increase the building industry's productivity. The author illustrated the potential of automation technologies in the construction industry in this part.

3.1 System Architecture of Floor Tiling Robotic System

The hardware design plays a critical role in laying the foundation for the robotic floor tiling system prototype. This segment demonstrated the considerations involved in selecting components and additionally furnished an in-depth examination of every facet of the hardware design procedure in order to guarantee accurate perception and precise motion of the robotic system.

The floor tiling robotic system's system architecture comprises various components, including a main controller, a microcontroller, sensors, and a pneumatic system. As illustrated in Fig. 3 below, the Raspberry Pi functions as the main controller of the entire system and the central processing unit responsible for monitoring and controlling critical operations such as image processing, communication, and decision-making. Raspberry Pi establishes communication connection with Arduino as a publisher. Arduino will operate as the microcontroller of the robotic system, facilitating control operations at a low level and enabling communication with sensors such as servo motors. Arduino facilitates synchronised operation of the pneumatic system, which includes a negative pressure vacuum pump, a solenoid valve, and a suction cup with spring actuator, in addition to the servo motors.

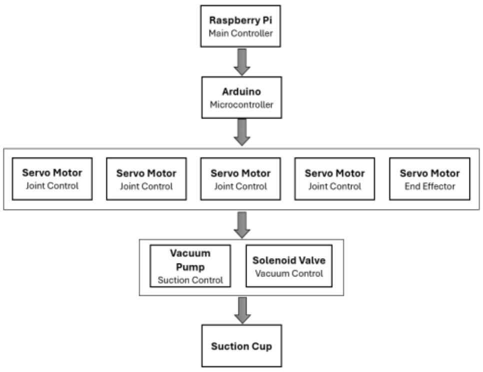


Fig. 3. System Architecture of Floor Tiling Robot.

3.2 Vacuum System and Camera Module

End effectors are affixed to the extremity of a robot's arm to facilitate its interaction with the immediate surroundings. Robotic systems rely on end effectors to effectively handle, manipulate, and sense objects. The robotic arm of the floor tiling system incorporates a pneumatic system as its end effector. This system comprises a negative pressure vacuum pump, a suction cup equipped with a spring plunger, and a solenoid valve.

The suction cup, vacuum suction pump, and 3-way 2-position solenoid valve work in concert to enable the suction-based manipulation of objects, including tiles, within a robotic system. The airflow is regulated by the solenoid valve, which transitions between phases in order to generate or discharge suction at the suction cup, as illustrated in Fig. 4 below. In the absence of activation of the solenoid valve, the airflow from the negative pressure vacuum compressor is redirected to the suction cup, thereby generating a vacuum, and ensuring the cup's secure attachment to the tile's surface. As the source of negative pressure, the vacuum suction compressor produces suction force that secures the object to the suction cup. By maintaining the necessary vacuum level, it guarantees a steady and dependable hold on the object during the entirety of the manipulation procedure.

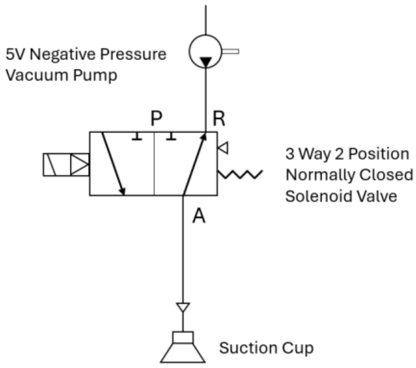


Fig. 4. Pneumatic Circuit Diagram of Vacuum System.

When used in conjunction with the suction system, the camera module serves as the visual monitor of the autonomous system, granting it perceptive abilities. By utilising the camera module's lens, the system acquires instantaneous visibility of its environment, which empowers it to precisely identify tile perimeters, corners, and possible impediments. The process of choosing an appropriate camera module is determined by factors including field of view, resolution, and compatibility with image processing algorithms. This guarantees a smooth integration of the module into the control architecture of the system. The IMX219 camera module will be designated as the visual sensor of the end effector.

3.3 Prototype and Wiring Diagram

The construction and connecting of physical parts necessary for the functioning of the robotic system constitute the hardware integration process. As shown in Fig. 5, the main parts of the system are the rigidly mounted robotic arm, servo motors, vacuum pump, solenoid valve, suction cup, and camera module.

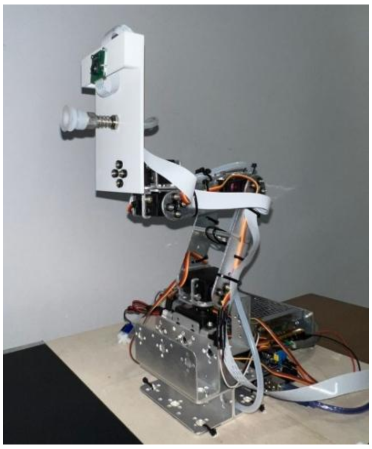


Fig. 5. Floor Tiling Robot Prototype.

The floor tiling robot's wiring diagram is depicted in Fig. 6. The Arduino Uno was affixed to the Arduino Sensor Shield, to which each MG996R servo motor was joined. Furthermore, a 5V 20A AC-DC power supply was linked to the Arduino. In the meantime, the IMX 219 camera module was connected to the Raspberry Pi, which maintained serial communication with the Arduino via a USB A-B cable.

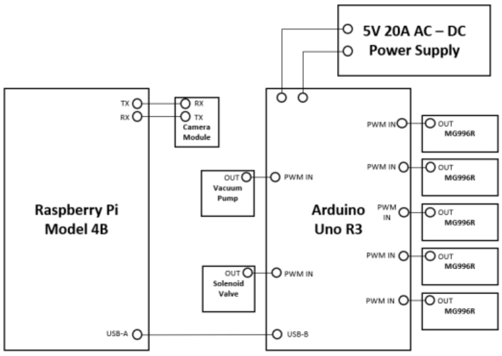


Fig. 6. Wiring Diagram of Floor Tiling Robot.

3.4 Floor Tiling Robot Control Algorithm Design

In conjunction with the control algorithm, a finite state machine (FSM) regulates the motion of the robotic arm. As illustrated in Fig. 7, the FSM establishes a number of states that correspond to different phases of the tile installation procedure, such as idle, initialization, tile collection, and tile placement. Transitions between states are initiated in response to user commands, while user commands will be determined based on the image processing result, thereby ensuring that the movements of the robotic arm are systematically controlled and coordinated. The FSM can be sectioned into several states:

I. Initialization State: The robotic arm enters an initialization state during system launch, during which it calibrates servo motors, initializes sensor readings, and makes operational preparations.
II. Idle State: During periods of inactivity, the robotic arm awaits control system instructions before initiating the installation of tiles. It maintains a stationary state until a subsequent assignment is made.
III. Tile Pickup State: When a tile pickup command is given, the robotic arm enters the state designated for tile pickup. In this instance, the object performs accurate motions in order to approach the specified tile, activate the suction cup mechanism, and firmly grasp the tile.
IV. Tile Placement State: After a successful collection operation, the robotic arm enters the state of tile placement. The object precisely locates the desired location on the floor, aligns it with adjacent tiles, and then releases it with deliberate and controlled motions.
V. Error Handling State: Error handling states are integrated into the FSM in order to accommodate unforeseen issues, such as sensor failures and communication errors. These states facilitate the robotic arm in ceasing operations in a secure manner, informing the control system, and awaiting additional instructions.

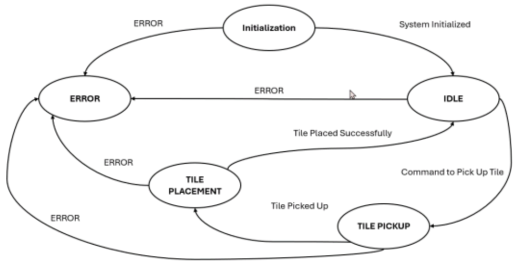


Fig. 7. Floor Tiling Robotic System Finite State Machine Diagram.

Although hardware limitations present certain difficulties, the control algorithm has been intentionally developed to be versatile and adjustable. To allow for the inclusion of uncertainties introduced during calibration and variations in tile positions, tolerance thresholds are incorporated. Furthermore, the implementation of adaptive strategies, including dynamic path planning and obstacle avoidance, serves to augment the system's capacity to adjust to environmental fluctuations and unanticipated impediments.

In summary, the control algorithm development for the robotic system for floor tiling recognises and resolves the hardware limitations that were imposed. By employing an open-loop control approach and a manual calibration procedure, the system adeptly manipulates and guides tiles to their assigned locations. Although functioning within these limitations, the control algorithm maintains its flexibility and adaptability, guaranteeing resilient performance in diverse circumstances.

3.5 Image Processing and Computer Vision

Within the framework of the floor tiling robotic system, the accurate perception and interpretation of the robot's surroundings are at the mercy of image processing and computer vision methodologies. This section provides further details regarding the robust image processing pipeline that has been developed to identify and evaluate the alignment of tiles within the robot's field of view.

The image pre-processing pipeline Fig. 8 commences by applying calibration parameters to undistorting raw images captured by the IMX219 camera module. After the image has been corrected for distortion, colour filtration is performed to distinguish white pixels that represent tiles. In this stage, the image is thresholder according to the colour of the tiles in order to preserve exclusively the white areas, thereby streamlining subsequent processing operations.

After filtration, the image undergoes a grayscale conversion in order to decrease computational complexity and improve the extraction of features. The grayscale image then undergoes a median blur operation in order to eliminate noise and level out irregularities, thereby enhancing the precision of subsequent feature extraction. By sharpening the blurred image, edges and details are accentuated, which facilitates the identification of tile boundaries and contours. In order to identify edges within the sharpened image, which highlights significant gradients and facilitates the localization of tile boundaries, canny edge detection is utilised. By defining and masking a region of interest (ROI) that corresponds to the floor area from the edge-detected image, subsequent processing is limited to pertinent areas that contain tiles.

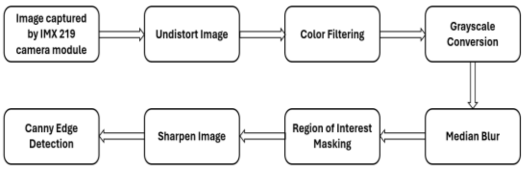


Fig. 8. Image Preprocessing Pipeline.

In order to identify and evaluate the alignment of floor tiles, tile detection and localization algorithms are implemented on the pre-processed image. The OpenCV 'find Contours' function is employed to identify contours that symbolise prospective tile boundaries, thereby delineating the precise locations of individual tiles within the ROI. Squares are selected from the detected contours according to their shape characteristics; contours are filtered by area, perimeter, and aspect ratio in order to identify square-shaped regions that correspond to tiles. The coordinates of the four vertices of each detected square are obtained, which symbolise the corners of specific tiles and function as benchmarks for further analysis.

4. RESULTS AND DISCUSSION

4.1 Quantitative Metrics

To assess the Floor Tiling Robot's performance, an experiment was carried out. Every tile's paving time is recorded during the experiment, beginning when the Floor Tiling Robot is in the finite state machine's "IDLE" state. In order to assess if the Floor Tiling Robot had accurately positioned each tile, as shown in Fig. 9 and Fig. 10, six squares with a gap of 50 mm were marked on the floor so that they could be checked for alignment.

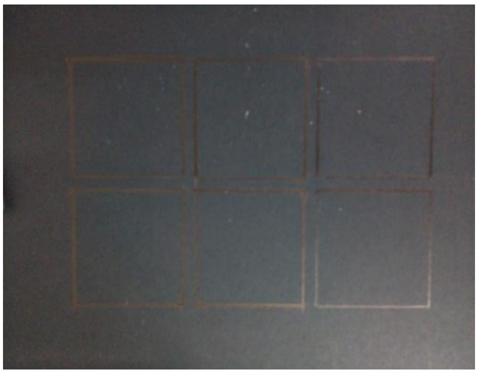


Fig. 9. Six Squares Marked on the Floor.

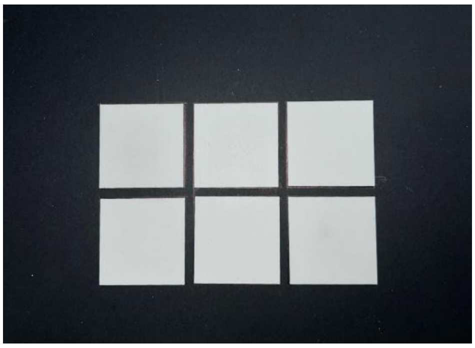


Fig. 10. Well Installed Floor Tiles.

The complete tiling process is shown in Fig. 11.The experiment was repeated 20 times, and the following quantitative results were obtained, as illustrated in Table 1 and Table 2:

- Total Installation Time: The total time for all tile installations across the 20 experiments was approximately 233.2 seconds.
- Average Installation Time per Tile: The average installation time per tile for tiles No. 1 to 6 was recorded as follows: 38.83 seconds, 38.75 seconds, 38.6 seconds, 39.1 seconds, 38.9 seconds, and 39 seconds, respectively. The total average installation time for all six tiles is approximately 38.87 seconds.
- Accuracy of Tile Placement: Overall, 70% of tiles were accurately placed in their correct position, with 84 tiles successfully positioned correctly. However, 36 tiles required adjustment or were inaccurately placed.
- Defect Angles: Tiles with imperfect angles were measured using an angle ruler ranging from 0° to 180°. Six tiles exhibited defective angles ranging from -5° to 7°.

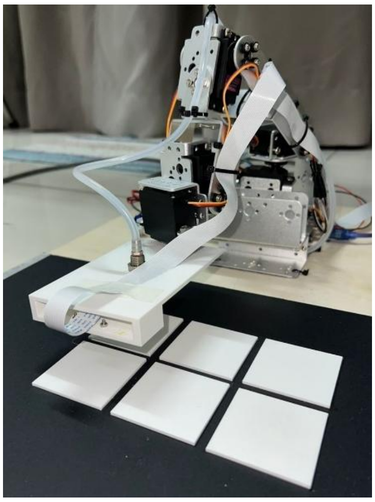


Fig. 11. Floor Tiling Robot's Tiling Process.

Table 1. Tiling Time Recorded During Experiments.

Tile No.	Tiling Time (s)																				Total Time (s)	Average Time (s)
1	38	38	40	37	35	39	41	37	40	41	36	41	39	38	40	39	40	36	40	42	777	38.85
2	39	37	41	38	37	36	38	40	42	39	37	38	40	35	38	39	41	39	38	40	775	38.75
3	37	40	35	38	36	37	40	41	38	40	40	39	42	36	40	38	40	36	38	41	772	38.6
4	40	39	41	37	40	40	38	41	36	39	40	39	35	41	40	41	39	38	37	41	782	39.1
5	36	39	41	39	37	40	46	38	43	35	37	39	40	41	37	39	40	38	36	37	778	38.9
6	35	41	38	49	35	39	40	37	36	40	40	37	39	38	39	40	38	41	37	41	780	39
																					Total Average Time (s)	233.2

Table 2. Accuracy of Tile Installation Recorded during Experiments.

Tile No.	Tile in Correct Position (1 = Yes, 0 = No)																				Total	Accuracy (%)
1	1	1	1	0	1	0	1	1	0	0	1	1	0	1	0	1	1	0	1	1	13	65
2	0	1	1	1	0	0	1	0	1	1	0	1	1	1	1	0	1	1	1	1	14	70
3	1	1	1	1	0	1	0	1	1	0	1	1	1	0	1	1	1	1	1	0	15	75
4	1	0	1	1	0	1	1	0	1	1	1	0	0	1	1	1	0	1	1	1	14	70
5	1	0	1	1	1	0	1	1	1	1	0	0	1	1	0	1	1	1	1	1	15	75
6	0	1	1	0	1	0	1	0	1	1	1	1	0	0	1	1	0	1	1	1	13	65
																					Average Accuracy (%)	70

4.2 Qualitative Analysis on System Reliability

Based on the inaccurate results obtained from the experiment, it can be observed that servo motors may not always reach the exact same position despite receiving the same input signals. MG996R is typically used in small robotics applications. On the other hand, standard MG996R servo motors do not have any feedback mechanisms like potentiometers or encoders. There is no direct method for the servo motor to determine its current position if there is no feedback mechanism in place. As a result, compared to servo motors with feedback systems, the motor's position precision may be limited, and it is more difficult to adopt advanced control techniques like hysteresis.

Furthermore, the MG996R servo motor's datasheet states that its dead band width is 5 microseconds. The range of input signals that the servo motor does not react to or move inside is referred to as the dead band width. Because the servo motor might not react to slight changes in the input signal until the signal exceeds the dead band range, dead band width might cause differences in placement. Because of this, the motor might not change its position if the input signal is within the dead band range, which could result in tiny positioning errors, particularly if the input signal varies near the dead band's edge.

On the other hand, the threshold values utilized by the Canny algorithm for edge detection have a substantial effect on the final results, as shown in Fig. 12, the red colour region still contains noise. These thresholds determine which edges are deemed strong and which are deemed weak, and they play a crucial role in determining the sensitivity and precision of edge detection.

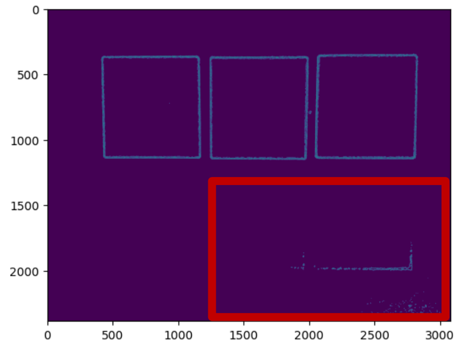


Fig. 12. The image noise increased due to light condition.

The qualitative analysis concludes by highlighting how crucial it is to solve issues and improve the functionality of crucial parts, like servo motors and image processing algorithms, in order to increase the floor tiling robotic system's dependability and longevity. To reduce positioning mistakes, strategies for enhancing servo motor precision should be investigated. These may include the use of feedback systems or the improvement of control approaches. Similar to this, threshold settings in image processing algorithms can be carefully adjusted to provide consistent performance in a variety of environmental situations and increase the accuracy of edge identification. Through the implementation of these principles, the system can become more dependability and robust in practical tiling applications, hence augmenting its overall efficacy and user contentment.

5. CONCLUSION

The aim of this study was to conduct a thorough assessment of the floor tiling robotic system to ascertain its efficiency, reliability, and ease of use for the automation of floor tiling process. Combining objective evaluation, quantitative data, and qualitative analysis led to several important findings. These results provided useful details regarding the advantages and disadvantages of the system. Although the system demonstrated potential for automating floor tiling process, further development and optimisation are needed to address identified issues and enhance general usability and dependability.

References

1. Donahue J., Girshick R., Darrell T., Malik J., & Berkeley U. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, 2014 IEEE Conference on Computer Vision and Pattern Recognition, doi: 10.1109/CVPR.2014.81, https://arxiv.org/abs/1311.2524
2. P. E. Nikravesh, Planar Multibody Dynamics, Formulation, Programming with MATLAB, and Applications, 2nd edn., CRC Press, Boca Raton, 2018, https://www.routledge.com/Planar-Multibody-Dynamics-Formulation-Programming-with-MATLABr-and-Applications-Second-Edition/Nikravesh/p/book/9781138096127?srsltid=AfmBOorYm4L6Ciz7aFVW2KOyD64lYsigmejpyj9K1mevCzsMicjHvlRw
3. Floor Tile Paving Robot R-19, https://www.red-dot.org/project/floor-tile-paving-robot-r-19-58278
4. Girshick R. (2015) Fast R-CNN, 2015 IEEE International Conference on Computer Vision (ICCV), doi:10.1109/ICCV.2015.169, How Are Industrial Robots Built? A Guide on the Components and the Movement of Robot Arms. (2018, May 22). Retrieved from KawasakiRobotics: https://robotics.kawasaki.com/ja1/xyz/en/1804-03/
5. Image Processing: Techniques, Types, & Applications [2024], https://www.v7labs.com/blog/image-processing-guide
6. Kundu R., What is YOLO architecture and how does it work? Learn about different YOLO algorithm versions and start training your own YOLO object detection models, YOLO: Algorithm for Object Detection Explained [+Examples], https://www.v7labs.com/blog/yolo-object-detection
7. Li W., Yang C., Jiang Y., Liu X., & Su C. Y., Motion Planning for Omnidirectional Wheeled Mobile Robot by Potential Field Method, Journal of Advanced Transportation, 2017, doi:10.1155/2017/4961383, https://onlinelibrary.wiley.com/doi/10.1155/2017/4961383
8. Liu T., Zhou H., Du Y., Zhang J., Jianping Z., & Li Y. (2018) A Brief Review on Robotic Floor-Tiling, IECON 2018, 44th Annual Conference of the IEEE Industrial Electronics Society, 5583-5588, doi: 10.1109/IECON.2018.8591123, https://researchr.org/publication/LiuZDZZL18
9. Mobile Robotic Tiling, SEC Singapore-ETH Centre, 2013-2016, https://gramaziokohler.arch.ethz.ch/web/e/forschung/257.html
10. Rao R., 7 Types of Industrial Robots: Advantages, Disadvantages, Applications, and More, Retrieved from Wevolver: https://www.wevolver.com/article/7-types-of-industrial-robots-advantages-disadvantages-applications-and-more
11. Rao R., What are End Effectors in Robotics? Types of End Effectors, Applications, Future. Retrieved fromWevolver: https://www.wevolver.com/article/what-are-end-effectors-in-robotics-types-of-end-effectors-applications-future
12. Redmon J., Santosh D., Girshick R., & Farhardi A., (2016) You Only Look Once: Unified, Real-Time Object Detection, *Computing Research Repository,* https://arxiv.org/abs/1506.02640
13. Ren S., He K., Girshick R., & Sun J. (2016) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, https://arxiv.org/abs/1506.01497
14. Ku Nurhanim, I. Elamvazuthi, P. Vasant, T. Ganesan, S. Parasuraman, M.K.A. Ahamed Khan, Joint Torque Estimation Model of Surface Electromyography (sEMG) Based on Swarm Intelligence Algorithm for Robotic Assistive Device, https://scholar.google.com/citations?view_op=view_citation&hl=ja&user=s5qOVpcAAAAJ&citation_for_view=s5qOVpcAAAAJ:u-x6o8ySG0sC
15. Phung Quang Anh, Tran duc Chung, Tran Tuan, M.K.A. Ahamed Khan, Design and Development of an Obstacle Avoidance Mobile-controlled Robot, https://www.researchgate.net/publication/337503620_Design_and_Development_of_an_Obstacle_Avoidance_Mobile-controlled_Robot

Authors Introduction

Mr. Hue Chau Jieng

He completed his Degeree in Mechatronics Engineering at UCSI University. He has working experience of one year.

Dr. M. K. A. Ahamed Khan

He pursued a PhD in Robotics, Power Electronics, and Controls in the United States and holds certifications as a Professional Engineer (PEng) in the USA and a Chartered Engineer (CEng) in the UK. He is a Senior member of the IEEE in the USA and a member of MIET in the UK. He is also the past chair for IEEE RAS Malaysia chapter.

Dr. Mastaneh Mokayef

She has received her PhD from University Technology Malaysia (UTM) in 2014. Her research interests include: Wireless communications, spectrum sharing method, spectrum management, cellular communication systems and Antenna design.

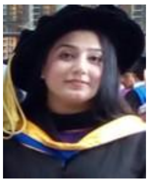
Ts Amar Ridzuan Bin Abd Hamid

He has completed his master's degree from Universiti Putra Malaysia (UPM), Postgraduate Diploma in Tertiary Teaching (PGDTT) from UCSI University, Malaysia. He works at UCSI University.

Dr. Abdul Qayyum

He received his Ph.D from Universiti Teknologi Petronas Malaysia. His area of interest is machine learning, deep learning and quantum machine learning for signal processing and bioemdical imaging.

Dr. Moona Mazher



She received her Ph.D. in Computer Engineering and Mathematics from the University of Rovira i Virgili, and works in UCL, UK.

Dr. Susama Bagchi

She received her Ph.D. in Electrical Engineering from Universiti Tun Hussein Onn Malaysia in 2022. She is currently working as an Associate Professor at the Chitkara University Institute of Engineering and Technology (CUIET), Chitkara University, Punjab, India. She has 15 years of experience in engineering management, academia, and research.

Dr. Sanjoy Kumar Debnath

He completed Ph.D. from the Faculty of Electrical & Electronic Engineering, Universiti Tun Hussein Onn Malaysia (UTHM) in 2022. He is currently working as an Assistant Professor at the Chitkara University Institute of Engineering and Technology (CUIET), Chitkara University, Punjab, India.

Dr. Takao Ito

He is Professor of Management of Technology (MOT) in Graduate School of Engineering at Hiroshima University. His current research interests include automata theory, artificial intelligence, systems control, quantitative analysis of interfirm relationships using graph theory, and engineering approach of organizational structures using complex systems theory.

Evaluation of Heart Disease Risk Using Deep Learning Technique with Image Enhancement

Abdul Majid Soomro[1], **Asad Abbas**[2], **Susama Bagchi***[3], **Sanjoy Kumar Debnath**[3], **Awad Bin Naeem**[2],
M. K. A. Ahamed Khan[4], **Mastaneh Mokayef**[4], **Takao Ito**[5]

[1]*Department of Computing, National University of Modern Languages, Multan Campus, Pakistan*
**Email: abdulmajeed.soormo@numl.edu.pk*

[2]*Department of Computer Science, National College of Business Administration & Economics, Multan, Pakistan*

[3]*Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India*

[4]*Faculty of Engineering, UCSI University, Malaysia*
**Email: bagchi.susama@gmail.com, sanjoy.kumar@chitkara.edu.in*
https://www.chitkara.edu.in

[5]*Hiroshima University, Japan*
itotakao@hiroshima-u.ac.jp

**Corresponding Author*

Abstract

This study emphasizes the significance of the heart in the human body. Numerous serious vascular conditions exist in the heart and the blood. The dataset, study goals, methodology, approach, and efficient algorithms for identifying and classifying electrocardiogram (ECG) data are all covered in this paper. Picture scaling, grayscale conversion, and training/testing dataset segmentation are part of the cardiovascular ECG image-processing process. To assess ECG images, researchers used a convolutional neural network. Iterations in model training increase the accuracy. We examined generalization and model recall using an additional dataset. The accuracy, F1 score, confusion matrices, and ECG pattern identification must all be evaluated to diagnose heart disease. We used a Cardiovascular ECG Image Collection that was openly accessible. The system was constructed in Python using Matplotlib, NumPy, and Keras. The GPU-based machine learning platform was Google Colab. Photos were analyzed, categorized, and processed using MobileNet-V2. With a remarkable accuracy rate of 99.3 %, the developed model offers a viable basis for further hyperparameter investigation. Additionally, to demonstrate the efficacy of the selected method, we present a graphic depiction of ECG data. Overall, this study combines advanced machine learning algorithms, strict assessment criteria, and ECG image analysis to enhance the diagnosis of heart-related disorders.

Keywords: Electrocardiogram (ECG), convolutional neural network (CNN), heart disease, MobileNet-V2

1. Introduction

The human heart is an important and complex organ that is necessary for life because it pumps blood throughout the body. Blood is continuously pumped through the chest by the heart, which is a strong organ of the size of a fist. The body's hormones, nutrients, and oxygen are supplied by the four chambered hearts, which consist of two atria and two ventricles. Muscle contractions and relaxations, which occur 60–100 times per minute, are perfectly timed by an internal electrical system to assist this essential function. In addition to its medical significance, the heart has a significant cultural and emotional value as a representation of love and strong human emotions. Throughout history, this has inspired many authors, painters, and artists [1]. As an organ that sustains life and represents our deepest emotions, the heart is a remarkable and significant part of who we are. The intricate structure of blood circulation and carefully controlled functions are essential for maintaining human health. Symbolic values are deeply ingrained in human emotions, culture, and artistic expression. One of the most important and aesthetically pleasing parts of the human body is the organ [1], [2], [3]. This general phrase covers a wide range of conditions that affect the functioning of the heart and blood vessels. A person's general health and well-being may be significantly affected by disease-related cardiac dysfunction [4]. Numerous modifiable and non-modifiable risk factors are associated with heart disease.

Diet, physical inactivity, high alcohol use, and smoking significantly increased the risk of heart disease. Additionally, the likelihood of developing heart disease may be increased by age, genetics, and certain medical conditions [5], [6]. The effects of heart disease are extensive and often include major health problems, a reduced quality of life, and early mortality. For patients to make informed decisions about their cardiovascular health and for healthcare professionals to deliver high-quality treatment, patients must be informed about the causes, symptoms, prevention measures, and therapies of heart disease [7]. On Electrocardiogram (ECG), these impulses are clearly seen as waves. Electrocardiograms (ECGs) are used by cardiologists to diagnose heart disease, track the health of known cardiac patients, and evaluate symptoms, including palpitations and chest discomfort [8]. Many elements of the ECG waveform, including the QRS complex, P-wave, T-wave, and ST segment, provide vital details regarding the rhythm and operation of the heart. Every type of ECG is intended for a specific clinical situation. While resting ECGs are used for everyday monitoring, Holter and event monitors are crucial for recognizing uncommon arrhythmias because they collect continuous data over extended periods. While exercise stress tests examine how the heart reacts to physical effort, ambulatory electrocardiograms (ECGs) record a patient's whole cardiovascular activity over the course of several weeks or months. With its noninvasive analysis of the electrical activity of the heart, ECG is an essential tool in cardiology that helps doctors diagnose, monitor, and treat cardiac conditions [9]. A technique called Deep Learning allows a computer to recognize visual inputs without explicitly removing any aspect [10]. In terms of accuracy, deep-learning models outperform human-level abilities. Labeled massive data was used to train deep learning models. The architecture of a traditional neural network consists of several layers. Levels with many hidden levels are referred to as deep layers. Similar to machine learning, deep learning comprises several layers with nonlinear processing units. The characteristics of the tiers above provide the input to the lowest layer. A Convolutional Neural Network (CNN) is an algorithm used for deep learning. These networks consisted of completely connected layers and several convolutional layers. In addition, the layers were clustered together. The use of Boltzmann machines is restricted to Deep Belief Networks (RMBs). Deep networks can be trained using (RMB)/RMB/auto-encoders [11], [12], [13]. The prize for the best object recognition performance is given by the CNN. Deep neural networks outperform all other machine learning methods in terms of efficiency. Convolutional neural networks, particularly deep neural networks, have shown excellent performance in image recognition. Picture categorization has undergone a paradigm shift owing to recent developments in computer vision and machine learning [14], [15]. By thoroughly examining the dataset's real-world variability and using appropriate preparation approaches, a CNN-based system can be trained to accurately recognize ECG data and make better generalizations to other scenarios [16], [17]. To demonstrate the superiority of the current ECG classification approaches, including deep learning-based and classical methods, a comparative study using a CNN-based methodology was conducted. The results of this study have broad implications for ECG classification and computer vision applications. The proposed method will enhance the medical sector by empowering healthcare providers to automate ECG grading tasks. Finally, our research will help create automated ECG identification systems for the medical industry by addressing the problems of ECG categorization and using state-of-the-art machine learning algorithms. The methodology, experimental data, and comments are covered in more detail in later chapters, which will provide a comprehensive assessment of the ECG classification system. The use of deep learning techniques to categorize ECG images is the main advantage of the proposed method [18].

This study has initiated the development of the previously discussed objectives that were outlined and implemented regarding the requirements of the research question and expected results. One of these was to develop a deep learning-centric ECG recognition and classification model and an exhaustive review of the literature. An important contribution of our study is: 1) assessment of the effectiveness of the machine learning model, 2) use of image datasets as image dataset suits modern research better, and 3) identifying which machine learning algorithm performed best when classifying heart disease based on performance.

2. Methodology

It is important to select an effective and optimal methodology for identifying and classifying ECG. We can then accurately process our data and train the machine using this limited dataset. Here, we choose the type of methodology and algorithm workflow that we use in this study, which gives us a good result in the output. This chapter describes the research objective, the algorithm we used, and our data with some analysis, along with the proposed workflow. We analyzed the ECG classification dataset of Cardiovascular ECG Images using appropriate data preprocessing. In the preprocessing step, the effects of resizing, grayscale conversion, and labeling were examined. Subsequently, a deep learning model called MobileNet-V2 [18] is trained and implemented. Predictions were performed using the aforementioned algorithms. Each step is described in detail below: The proposed approach for ECG classification consists of five main processes and supporting steps: Fig. 1 shows a flowchart of our methodology.

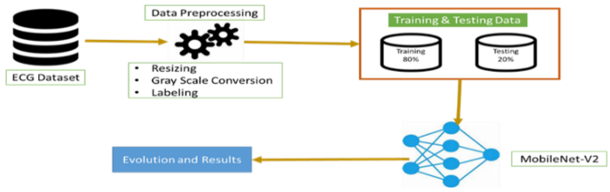


Fig. 1. Flow Chart.

2.1. *Dataset Collection*

The Cardiovascular ECG Image dataset chosen for the proposed study is freely accessible at Kaggle [14]. Because the dataset used was not proprietary or self-compiled, it was taken from an online source, so there was no relevant population of the study from which the dataset was collected. The cardiovascular ECG image collection included four ECG combinations. ECG images of patients with abnormal heartbeats and ECG images of a normal person are the only two categories of ECG that were used to support the theory. By selecting these two categories, I narrowed down the dataset to what was relevant for their research or analysis. This filtering process yielded a dataset comprising 741 ECG images. The primary objective was to create a dataset suitable for training and testing the machine learning models.

2.2. *Data Preprocessing*

This is a crucial phase in the preprocessing workflow that prepares the ECG images for training. Because preprocessing ensures that the input is in a format suitable for training models, it is crucial to machine learning and computer vision. The initial step was to resize the ECG images to standard resolution. This is important because machine-learning algorithms often demand constant dimensions in the input data. Resizing makes processing the model easier by keeping the width and height of each photograph consistent. By removing the color information, grayscale conversion also simplifies the images, which may help reduce the complexity of the model and improve its generalizability. Finally, splitting the dataset into training and test subsets is a crucial machine-learning stage for precisely assessing the model's performance. While the training subset was used to train the model, the test subset was kept separate for evaluation to ensure that the model could generalize well to new data. These steps maximize the training process and eventually enhance the performance of the model in ECG interpretation tasks by standardizing, cleaning, and organizing the data. Fig. 2 shows the processed image.

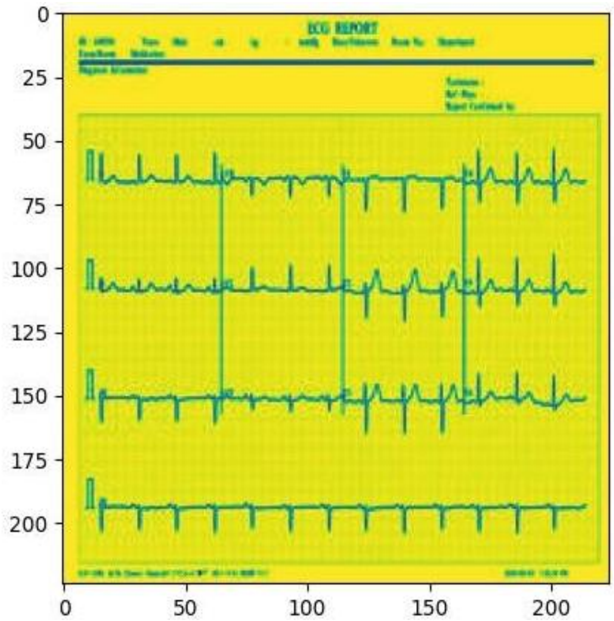


Fig. 2. Processed Image.

2.3 *Model Architecture Design*

Building an appropriate architecture for a convolutional neural network for ECG image processing is one of the critical tasks in medical imaging analysis. Intelligent architecture is important in the acquisition of critical information as well as avoiding problems like overfitting and underfitting. The first set of layers in a CNN model for processing ECG images is comprised of a series of convolutional layers. Filters are used in these layers to explore the input images and capture low-level features such as edges and curves. For ECG analysis, detection of waveform patterns and structures can be included in the process.

2.4 *Model Training with Training Dataset*

Training a CNN model using preprocessed fruit images. This involves feeding the training dataset through a CNN model, computing loss (a measure of the model's prediction error), and optimizing the model's weights using an optimizer and a predetermined learning rate. Repeated updates were made until the model achieved an acceptable degree of accuracy on the training dataset. To train the model using the training dataset, the system used 593 ECG images. For training purposes, we selected the MobileNet-V2 model.

2.5 *Model Testing*

Testing the performance of a trained Convolutional Neural Network model with a separate test dataset is a critical step in the machine learning pipeline. The model has never seen a portion of the data, such as the test dataset, which serves as an independent and hidden subset. This replicates real-world situations in which the model discovers previously undiscovered data. The main goal of this testing step is to evaluate the model's ability. to generalize or how well it can make accurate predictions on data that it has not been explicitly trained on. During this evaluation, key performance metrics such as accuracy were computed. These metrics provide a quantitative evaluation of the model's performance by analyzing how well it categorizes objects or patterns in the test data. The model is a reliable tool for a variety of tasks because it can effectively apply the significant qualities and patterns learned from the training data to new, uncovered data, as indicated by the high recall and accuracy levels in the test dataset.

2.6 *Model Evaluation*

Understanding the operational efficiency of deep learning models requires understanding their accuracy and efficacy. These models are similar to highly clever computers that have access to large volumes of data. To measure how well people perform at certain jobs, such as identifying cardiac patterns in ECG readings, we require specific technology. We used a device called "accuracy." This provides details regarding the accuracy rate of the model. However, sometimes, particularly when one type of material dominates over another, the accuracy is not sufficient to capture the whole picture. Therefore, we also utilized a statistic called the "F1 score." This score allowed us to evaluate the accuracy of the model in identifying relevant data while reducing the number of errors produced. The next kind of graphic called “confusion matrix," lists all of the advantages and disadvantages of the model's operation. It is simpler to identify areas on a map where the model is doing well and those that require improvement. By using these techniques, we will be able to assess the extent to which our deep learning models help physicians identify heart anomalies, particularly when interpreting electrocardiograms. Patient care has improved, and everyone has more access to healthcare.

OpenCV Python was chosen for use in the software part for user detection. Processing a video implies executing it.

3. Results and Discussion

In computer vision and machine learning, preprocessing is essential to ensure that the data are suitable for model training. The greatest effort is to resize the ECG pictures to a standard resolution. This technique is essential because uniform-dimensional input data are required for machine-learning algorithms. Scaling helps the model to handle data by ensuring that the picture dimensions are the same. The research also shows how the grayscale conversion of photographs expanded the collection. The removal of color information from photos allows the model to handle larger datasets with less complexity owing to this conversion. Dividing the dataset into training and test subsets is another essential machine-learning step necessary to correctly examine the model's performance. While the training subset is used to train the model, the test subset is set aside to assess how well the model generalizes to data that has never been seen before. Google created the MobileNet-V2 modular convolutional neural network (CNN) architecture in response to the need for rapid and portable computer vision models. MobileNet-V2 is an ideal choice for low-processing-power applications because it is faster and requires less effort than its predecessor. This efficiency might be the result of cutting-edge technology, such as depth-wise separable convolutions, which use a single filter for each input channel to significantly reduce the processing burden. Deeper network topologies may be built using the paradigm's inverted residual blocks without significantly increasing processing complexity. MobileNet-V2 also includes global average pooling for spatial compression, and squeeze-and-excitation blocks for channel-wise feature correction. The model is often retrained using huge picture datasets, such as ImageNet, after which it is adjusted to suit specific demands. This widely used alternative is often used for mobile apps and edge device inference in real-time. It can be adjusted to fit specific applications and take a variety of input sizes. The novel CNN architecture, known as MobileNet-V2, combines accuracy and efficiency. Its unique design choices, such as depth wise separable convolutions and inverted residual blocks, may allow it to attain competitive accuracy with fewer parameters. Therefore, it can compete well in deep learning applications, even in systems with constrained resources. The fact that it is widely used in industry attests to its efficacy in satisfying the requirements of real-time inference in embedded and mobile applications, while producing excellent quality results. Matplotlib is one of these libraries that is used for picture plotting and presentation. NumPy is used to represent matrices and perform the operations required to build convolutional neural networks. The keras. layers, keras. preprocessing and keras.models modules of the TensorFlow Keras API were imported. These were used by the automatic learning processes of the system. In addition, I have included some specialized modules for minor system tasks, such as glob, which allows me to expand Unix pathname patterns and operating systems (OS), giving me access to some fundamental system operations. The efficiency or composition of the CNN was not affected by these modules. A systematic approach was adopted to comprehensively execute the experiment aimed at classifying various ECG patterns. Initially, a baseline model was trained to establish a fundamental performance benchmark. The effectiveness of the model as a whole is then evaluated through a series of controlled experiments to determine how certain hyperparameters affect it. The preliminary results of our training indicated a remarkable achievement, with our model achieving an impressive accuracy rate of 97.6 %. This exceptional baseline performance not only serves as a solid starting point but also paves the way for insightful investigations into how individual hyperparameters can further enhance the model's classification capabilities. Through these rigorous experiments, we aimed to refine our model and develop a robust ECG classification system. We displayed each class's annotated ECG individually using the matplotlib plot function after importing all our modules to verify that the photos would display and test the annotations. Fig. 3 displays the script used to display the normal and abnormal ECG images of the dataset.

```
#image of abnormal heartbeat
img = mpimg.imread('E://train//HB(1).jpg')
imgplt = plt.imshow(img)
plt.show()

#image of normal heartbeat
img = mpimg.imread('E://train//Normal(1).jpg')
imgplt = plt.imshow(img)
plt.show()
```

Fig. 3. The Script Used to Display the Dataset's Normal and Abnormal ECG Pictures.

Subsequently, the data were scaled by dividing each feature value by 255, which is a common practice when working with image data, to rescale the values from a range of 0 to 255 to a normalized range of 0 to 1. This scaling helps ensure that the features are on a similar numerical scale, facilitating the training of the machine learning models. This preprocessing is crucial for building and evaluating predictive models that aim to classify data as "normal" or "abnormal" based on the provided features. Fig. 4 shows that the dataset was split into training and testing groups at an 80:20 ratio.

Fig. 5 and Fig. 6 shows that our model successfully learns the ECG images with the highest validation accuracy of 99.3 %.

A confusion matrix was used to verify the MobileNet-V2 model. The prediction rate of the proposed image processing model was 99.3 %. Fig. 7 shows the confusion matrix for the MobileNet-V2 model.

```
#Train Test Split
X = normal_abnormal
Y = np.asarray(labels)
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=2)
print(X.shape, X_train.shape, X_test.shape)

# scaling the data
X_train_scaled = X_train/255

X_test_scaled = X_test/255
print(X_train_scaled)
```

Fig. 4. Data Training and Testing.

```
Final Train Accuracy: 0.9763513803482056
Final Validation Accuracy: 0.9932885766029358
Final Train Loss: 0.0958276018500328
Final Validation Loss: 0.07850576937198639
Precision: 1.0
Recall: 0.9764705882352941
F1 Score: 0.988095238095238
```

Fig. 5. Model Training Result.

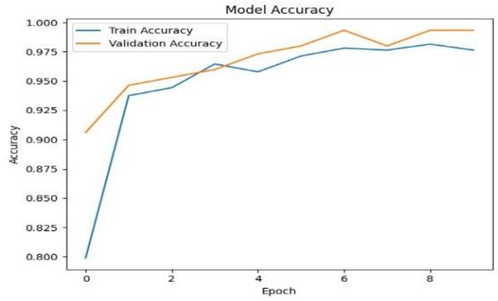


Fig. 6. Model Training and Testing Accuracy Result.

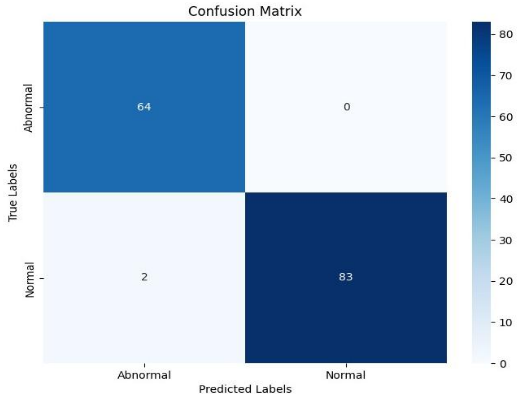


Fig. 7. Confusion Matrix Derived from the Outcome of the Developed MobileNet-V2 Model.

In this study, we presented a systematic approach to ECG pattern categorization using the CNN architecture of MobileNet-V2. We begin by examining an essential method for ECG estimation from the system viewpoint. Subsequently, we thoroughly analyzed the data in our dataset using conceptual and statistical methods to identify and compile patterns. Effective photo preprocessing, such as scaling to a consistent resolution, grayscale conversion, and dataset splitting for training and testing, ensured the best possible data preparation for our machine-learning model.

Our classification model is built on top of the widely known and flexible MobileNet-V2, which has innovative features, such as inverted residual blocks and depth-wise separable convolutions. We draw attention to the critical role that basic libraries and modules play in our methodology, which streamlines the process of creating models and conducting tests. Thorough model training, which involves establishing a high-accuracy baseline, improved our approach. As a result, testing could identify ECG patterns with an amazing 99.3 % accuracy. Furthermore, we used a confusion matrix to evaluate the accuracy, precision, and recall of the proposed model. These findings demonstrate the potential of this method for use in medicine by providing better clinical ECG classification accuracy and indicating future research possibilities for larger and more diverse datasets.

4. Conclusion

In this study, we built and evaluated an ECG classification system using MobileNet-V2 architecture. The importance of evaluating the ECG pattern at the beginning of the inquiry was highlighted through a thorough methodical evaluation. The following step was data analysis when the data were carefully examined and analyzed to identify any noteworthy trends. This method was supported by statistical data and conceptual tools. The research's main discovery was the benefits of picture preprocessing for computer vision and machine learning. Equitable sizing of ECG images made it possible to provide machine-learning models with comparable input sizes and processing speeds. Grayscale picture conversion was made easier by removing color information from the dataset, which reduced the model complexity and improved generalization. More specifically, the training and testing portions of the dataset were separated, which is a critical step in machine learning that allows for the consistent evaluation of model performance. The main goal of this study was to make use of Google's MobileNet-V2, a thin CNN architecture that offers significant processing reductions. MobileNet-V2 is considered the best option for applications with low resources because of its special characteristics, which include depth-wise separable convolutions and inverted residual blocks. The baseline model's amazing 99.3 % accuracy rate was reached after extensive testing, providing a solid basis for further research on how hyperparameters affect the model's classification performance. ECG imaging data were visually evaluated and confirmed using matplotlib. Data scaling was used to normalize the features during the training phase once the dataset was completely divided into the training and testing sets. During the model assessment phase, a system's remarkable 99.3 % prediction performance is emphasized via the use of a confusion matrix. The findings of this study highlight the potential of the MobileNet-V2-based ECG classification system as a useful tool for medical diagnosis and treatment, even if additional development and improvement are required to increase its accuracy and efficiency.

In the future, the ECG categorization methods may change. Currently, 741 images from ECGs are used to train the algorithm to understand them. However, we still have potential. Initially, we had more ECG pictures to speed up the learning process of the system. Modifications to computer program 31 might result in even greater accuracy. As an alternative, we may ask the machine to explain why it determines whether a certain ECG is good or terrible. Physicians may be asked to test a computer to ensure that it functions properly in actual hospitals. We may also make the required adjustments if the equipment interprets certain ECGs incorrectly. We also guarantee that computers will maintain and protect the privacy of people's medical records. We may also consider using more medical data to enhance computer performance. Finally, we may try to utilize a computer to track people's heart rates in real time in an attempt to identify issues early on. All these factors will impact the usefulness of the ECG computer system for physicians and patients.

References

1. Ayon S. I., Islam M. M., & Hossain M. R. (2022) Coronary artery heart disease prediction: a comparative study of computational intelligence techniques, IETE Journal of Research, 68(4), 2488-2507.
2. Bharti R., Khamparia A., Shabaz M., Dhiman G., Pande S., & Singh P. (2021) Prediction of heart disease using a combination of machine learning and deep learning, Computational intelligence and neuroscience, 2021.
3. Gan S. E. S., Debnath S. K., Alshebly Y. S., Nugroho H., Bagchi S., & Nafea M. (2021, October) Fetal ECG extraction from abdominal ECG using Chebyshev and Butterworth filters, 2021 IEEE Symposium on Computers & Informatics (ISCI), pp. 25-30, IEEE.
4. Farzana S., & Veeraiah D. (2020, October) Dynamic heart disease prediction using multimachine learning techniques, 2020 5th International Conference on Computing, Communication and Security (ICCCS), pp. 1-5, IEEE.
5. Gavhane A., Kokkula G., Pandya I., & Devadkar K. (2018, March) Prediction of heart disease using machine learning, 2018, the second International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 1275-1278, IEEE.
6. Rani P., Kumar R., Jain A., Lamba R., Sachdeva R. K., Kumar K., & Kumar M. (2024) An Extensive Review of Machine Learning and Deep Learning Techniques on Heart Disease Classification and Prediction, Archives of Computational Methods in Engineering, 1-19.
7. Guo L., Sim G., & Matuszewski B. (2019) Inter-patient ECG classification with convolutional and recurrent neural networks, Biocybernetics and Biomedical Engineering, 39(3), pp. 868-879.
8. Kachuee M., Fazeli S., & Sarrafzadeh M. (2018, June) Ecg heartbeat classification: A deep transferable representation, 2018 IEEE International Conference on Healthcare Informatics (ICHI), pp. 443-444, IEEE.
9. Kavitha M., Gnaneswar G., Dinesh R., Sai Y. R., & Suraj R. S. (2021, January) heart disease prediction using hybrid machine learning model, 2021 6th International Conference on Inventive Computation Technologies (ICICT), pp. 1329-1333. IEEE.
10. Khan F. S., Mohd M. N. H., Khan M. D., & Bagchi S. (2020, September) Breast cancer histological images nuclei segmentation using mask regional convolutional neural network, 2020 IEEE Student Conference on Research and Development (SCOReD), pp. 1-6, IEEE.
11. Kim E., Kim J., Park J., Ko H., & Kyung Y. (2023) TinyML-Based Classification in an ECG Monitoring Embedded System, Computers, Materials and Continua, 75(1), 1751-1764.
12. Katal N., Gupta S., Verma P., & Sharma B. (2023) Deep-Learning-Based Arrhythmia Detection Using ECG Signals: A Comparative Study and Performance Evaluation, Diagnostics, 13(24), 3605.
13. Yuen B., Dong X., & Lu T. (2019) Inter-patient CNN-LSTM for QRS complex detection in noisy ECG signals, IEEE Access, 7, 169359-169370.
14. https://www.kaggle.com/datasets/jayaprakashpondy/ecgimages
15. Naeem A. B., Senapati B., Chauhan A. S., Makhija M., Singh A., Gupta M., Tiwari P. K., & Abdel-Rehim W. M. F. (2023) Hypothyroidism Disease Diagnosis by Using Machine Learning Algorithms, International Journal of Intelligent Systems and Applications in Engineering, 11(3), 368–373.
16. Bagchi S., Huong A., & Tay K. G. (2017) Investigation of different spatial filters' performance toward mammogram de-noising, Int. J. Integr. Eng, 9(3), 49-53.
17. Naeem A. B., Senapati B., Chauhan A. S., Kumar S., Gavilan J. C. O., & Abdel-Rehim W. M. (2023) Deep

Learning Models for Cotton Leaf Disease Detection with VGG-16, International Journal of Intelligent Systems and Applications in Engineering, 11(2), 550-556.
18. Sandler M., Howard A., Zhu M., Zhmoginov A., & Chen L. C. (2018) Mobilenetv2: Inverted residuals and linear bottlenecks, Proceedings of the IEEE conference on computer vision and pattern recognition, (pp. 4510-4520).

Authors Introduction

Dr. Abdul Majd Soomro

He has over 20 years of teaching experience and is now a day HOD at NUML University. He possesses a robust enthusiasm for research and innovation in the fields of information science, information technologies, and IT management, with a particular emphasis on IT/IS research domains such as machine learning, data sciences, AI, deep learning, and network computing applications.

Dr. Asad Abbas

He received his Mphil. degree in Computer science in 2023 from NCBA&E Multan in Pakistan. all aimed at addressing practical network issues. He was awarded Ph.D. in Information Technology from Universiti Tun Hussein Onn Malaysia.

DR. Susama Bagchi received a Ph.D. in Electrical Engineering from Universiti Tun Hussein Onn Malaysia in 2022. She completed an M.Sc. in Modern Communication Technologies with Business Management from the University of Sussex, UK, in 2007 and a bachelor's degree (B.E.) in Electronics and Communication Engineering (ECE) from the University of Burdwan, India, in 2002. She is currently working as an Associate Professor at the Chitkara University Institute of Engineering and Technology (CUIET), Chitkara University, Punjab, India. She has 15 years of experience in engineering management, academia, and research.

Dr. Sanjoy Kumar Debnath

He completed Ph.D. from the Faculty of Electrical & Electronic Engineering, Universiti Tun Hussein Onn Malaysia (UTHM) in 2022. He is a "Graduate Engineer" from the Board of Engineers Malaysia and a "Graduate Member" from the Institute of Engineers Malaysia. He is also a member of the IEEE Delhi-India section. He is currently working as an Assistant Professor at the Chitkara University Institute of Engineering and Technology (CUIET), Chitkara University, Punjab, India.

Mr. Awad Bin Naeem

He completed his Master's degree in Computer Science from NCBA&E Multan in September 2022, focusing on smart transportation systems, machine learning, and deep learning, which sparked his interest in the subject. Awad's Master's studies focused on Vehicle-to-Everything (V2X) Communication and the development of an intelligent transportation system for different types of vehicles. He also proposed an Intelligent Transportation System framework for Autonomous, Non-Autonomous, and prioritized vehicles. He is passionate about intelligent transportation systems and data science problems. Awad has an impressive research background, having submitted 17 research publications, eight of which have been published. His research papers cover various topics, including disease detection, gender identification, population forecasting, deep learning algorithm.

Dr. M. K. A. Ahamed Khan

He pursued a PhD in Robotics, Power Electronics, and Controls in the United States and holds Chartered Engineer (CEng) in the UK. He is a Senior member of the IEEE in the USA and a member of MIET in the UK. He is also the past chair for IEEE RAS Malaysia chapter.

Dr. Mastaneh Mokayef

Her PhD from Wireless Communication Centre Faculty of Electrical Engineering in University Technology Malaysia (UTM) in 2014.She has many research papers.

Dr. Takao Ito

He is Professor of Management of Technology (MOT) in Graduate School of Engineering at Hiroshima University. His current research interests include automata theory, artificial intelligence, systems control, quantitative analysis of interfirm relationships using graph theory, and engineering approach of organizational structures using complex systems theory.

Driver State Monitoring Using Pose Estimation: Detecting Fatigue, Stress, and Emotional States for Safer Roads

Hao Feng Chan, Dexter Sing Fong Leong, Shakir Hussain Naushad Mohamed, Wui Chung Alton Chau
School of Engineering, Deakin University, Australia

Andi Prademon Yunus
Telkom University, Indonesia

Takao Ito
Hiroshima University, Japan

Zheng Cai, Xinjie Deng, Yit Hong Choo*
Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Australia
**Email: s221210401@deakin.edu.au, y.choo@deakin.edu.au*

Abstract

Driving under fatigue, stress, or emotional impairment poses significant risks to road safety. This paper proposes a custom pose estimation framework designed to detect driver states, such as fatigue and stress, by analyzing body posture, head pose, and gesture dynamics. Using a novel deep learning approach trained on diverse driving scenarios, the model identifies physiological and behavioral markers associated with impaired states. Unlike existing methods, this system integrates pose estimation with emotional and movement analysis, enabling robust performance in challenging conditions, including poor lighting and occlusions.

Keywords: Driving, Pose estimation, Fatigue, Deep learning, Computer vision

1. Introduction

Driver fatigue is a critical factor contributing to road accidents worldwide, impairing driver attention, decision-making capabilities, and reaction time. According to studies [1], drivers who are deprived of more than four hours of sleep are 10.2 times more likely to cause accidents. In Australia, fatigue is classified as one of the Fatal Five contributory causes of road crashes, alongside speeding, alcohol, distractions, and seatbelt non-compliance. The Transport Accident Commission (TAC) suggests that 20% of serious road crashes and 30% of fatal crashes in Australia are related to driver fatigue [2]. Fatigue-related crashes are significantly more common on rural roadways and during extended driving periods, especially at night [3]. Real-time monitoring and timely warnings can significantly reduce fatigue-related incidents and improve road safety.

Current driver fatigue detection methods can be categorized into traditional machine learning and deep learning approaches. Traditional methods, such as the Histogram Oriented Gradient (HOG) [4] feature descriptor combined with Support Vector Machines (SVM) [5], rely on visual cues like the Eye Aspect Ratio (EAR) [6] and Mouth Aspect Ratio (MAR) [7] to identify drowsiness through eye closure and yawning. While effective in controlled environments, these methods lack adaptability to varying lighting conditions, diverse facial features, and head postures. Additionally, they are computationally less efficient and unsuitable for real-time applications in dynamic driving environments [8]. Deep learning approaches leverage advanced neural network architectures to enhance accuracy and robustness. For instance, the Residual Channel Attention Network (RCAN) [9] combines facial landmark detection with attention mechanisms to classify the states of the eyes and mouth more accurately. However, these methods require heavy computational resources and large datasets to generalize well across different driving conditions and demographics [10]. Moreover, non-intrusive fatigue detection systems remain sensitive to challenges such as poor lighting, facial occlusions (e.g., sunglasses and masks), and varying driver behaviors, which can reduce detection accuracy in real-world scenarios [11].

Recent advancements, such as lightweight driver fatigue detection methods based on facial analysis, have been proposed to address these challenges. By directly extracting the eye and mouth regions using facial key points, these methods eliminate the need for manual parameter adjustment and adapt to changes in head posture [12]. For example, a facial ROI state recognition network (SRNet-FR) with the integration of Ghost module and SimAM achieves an accuracy of 99.03% with a parameter size of only 0.61M, meeting practical real-time requirements even under conditions of partial occlusion and low lighting [13].

In addition to visual analysis, emerging studies have explored physiological signals for fatigue detection. For example, a novel study utilized electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS) to monitor driver fatigue during seven-hour simulated driving experiments [14]. Results showed a significant link between behavioral data and hemodynamic changes in the prefrontal lobe after four hours, indicating key periods of performance decline. Although limited by a small sample size, the findings align with established fatigue standards and suggest physiological methods can complement visual fatigue detection systems.

To address the challenges of driver fatigue detection, computer vision models utilizing deep learning have emerged as efficient and reliable solutions. Among these models, You Only Look Once (YOLO) [15] stands out due to its speed and accuracy. YOLO is a real-time object detection system that processes images in a single pass, making it significantly faster than other detection models that require multiple passes. YOLO has gained popularity for its versatility and robustness in various computer vision applications, from autonomous driving to surveillance and beyond [16][17]. For driver fatigue detection, YOLO can handle varying lighting, partial occlusions, and diverse behaviors, making it a practical choice. Its lightweight design, optimized for real-time applications, excels at object detection tasks while being open-sourced, enabling broad accessibility and adaptability for diverse implementations, such as detecting the fatigue level of drivers. [18].

In addition to YOLO, PyTorch-based models such as ResNet and EfficientNet have emerged as strong benchmarks for fatigue detection. ResNet, with its skip connections, mitigates the vanishing gradient problem, allowing deep networks to capture intricate patterns like eye closure and head nodding [19]. EfficientNet, using a compound scaling approach, optimizes performance and computational efficiency, making it suitable for resource-constrained environments [20]. Both models have been successfully applied in detecting fatigue indicators like yawning and reduced alertness, providing a different perspective to YOLO [21][22].

This study evaluates YOLO for real-time driver fatigue detection, benchmarking it against ResNet [23] and EfficientNet [20]. ResNet and EfficientNet were included for comparison due to their known classification accuracy and scalability, as they effectively utilize residual connections and compound scaling approaches [23][24].

2. Methodology

This study evaluates driver fatigue detection using three lightweight models: YOLO, ResNet, and EfficientNet.

2.1. *You Only Look Once (YOLO)*

YOLO is a lightweight, widely used object detection framework known for its high-speed inference and accuracy [15]. Its single-stage architecture integrates feature extraction and classification, optimizing both computational efficiency and model performance [25]. This design eliminates the need for multiple passes through the data, significantly reducing latency compared to multi-stage models, making it particularly suitable for fast decision-making in safety-critical applications like driver monitoring. YOLO excels in challenging conditions, such as low-light environments or partial occlusions, its grid-based detection approach ensures effective feature extraction even from limited or obscured visual data [26]. YOLOv11 is chosen for its latest version and varied configurations: YOLOv11n, YOLOv11s, YOLOv11m, YOLOv11, and YOLOv11x [25]. For this study, YOLOv11 was selected for its balance of computational efficiency and detection accuracy.

2.2. *ResNet*

ResNet50 was chosen for this study because its 50-layer structure with residual connections effectively captures detailed patterns in facial expressions and poses, while mitigating the vanishing gradient problem [23]. It offers better feature representation than ResNet18 and ResNet34 without the high computational demands of deeper models like ResNet101 and ResNet152. This makes it suitable for detecting fatigue-induced behaviors in driver monitoring systems.

2.3. *EfficientNet*

EfficientNet_B0 was selected for this study due to its lightweight architecture and computational efficiency. As the baseline model in the EfficientNet family, it uses fewer layers, narrower channels, and processes images at a resolution of 224×224 pixels, making it ideal for resource-constrained environments like embedded systems or edge devices [20]. Its efficiency is suitable for driver fatigue detection applications requiring real-time monitoring and quick decisions [27], enabling continuous processing of cues like yawning or eye closure without straining system resources.

2.4. *Data Processing and Training*

To compare and verify performance between the models, a standardized dataset was utilized throughout the study. A publicly accessible dataset from Roboflow [28] was used and reorganized into two categories: safe driving and dangerous driving, as shown in Fig. 1 and Fig. 2, respectively.

The dataset consists of 3544 images and is divided into three subsets: 70% (2496 images) for training, 20% (713 images) for validation, and 10% (335 images) for testing. This method in machine learning defines dataset partitions, ensuring balanced and fair distribution for models. It provides ample training data while maintaining enough validation and test samples to consistently assess model performance. Mathematically, this is represented as:

$$N = N_{train} + N_{val} + N_{test}$$

where N is the total number of images, N_{train}, N_{val}, N_{test} are the number of training, validation, and testing samples, respectively. The dataset split evenly represents both classes to prevent bias during training and evaluation [29].

Fig. 1 Safe Driving

Fig. 2 Dangerous Driving

YOLOv11, ResNet, and EfficientNet were trained on the same dataset for 25 epochs to ensure fair comparison. The models classified images into two categories using pose estimation and visual features. YOLOv11s used a lightweight setup with a batch size of 32, image size of 640 pixels, and two data loading workers for efficiency. Validation measured precision, recall, and F1-score, with optional prediction saving for further analysis.

ResNet and EfficientNet models were fine-tuned on the dataset, with final layers adjusted for two output classes. Training used the Adam optimizer (learning rate 0.001) and cross-entropy loss. Images were resized to 224x224 pixels, with a batch size of 32.

The dataset preparation and split aimed to capture and minimize overfitting in the dataset, while the validation and test sets provided unbiased metrics for evaluating generalization abilities of the model. Performance metrics such as precision, recall, and F1-score were planned for evaluation post-training to measure the effectiveness of each model in detecting driver states defined mathematically as:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

where TP, FP, and FN denote true positives, false positives, and false negatives, respectively. These formulations are standard in evaluating classification models, providing clear criteria for assessing performance. In the context of driver state detection, such metrics are crucial for determining the effectiveness of models in identifying various driver behaviors [30].

The primary evaluation metric for this study is mean Average Precision (mAP), complemented by classification accuracy, precision, recall, and F1-score. These metrics provide an assessment of the model performance in detecting driver states. All models underwent training and evaluation under identical conditions to ensure a fair comparison. This setup is an assessment of the capability of different models in classifying driver states, with a focus on their computational efficiency, accuracy, and feasibility for real-time applications.

3. Results and Discussion

3.1. *Simulation and Performance Analysis*

In this section, a comparative analysis of the YOLOv11, ResNet, and EfficientNet models is presented for the task of driver fatigue detection. The experiments were executed on a system equipped with an NVIDIA GeForce RTX 4060 GPU, an Intel Core i7-12560H CPU operating at 2.50 GHz, and 16 GB of RAM.

The comparative performance of YOLOv11, ResNet, and EfficientNet is summarized in Table 1. Performance metrics such as precision, recall, and F1-score were evaluated post-training to assess the effectiveness of each model in detecting driver states.

Table 1. Comparative performance Metrics of YOLOv11, Resnet and EfficientNet

Metrics	YOLOv11	Resnet	EfficientNet
Loss	0.28	0.68	0.54
Accuracy	98.9%	96.85%	97.48%
Precision	98.7%	97.0%	97.4%
Recall	98.8%	96.7%	97.5%
F1-Score	98.8%	96.8%	97.5%
mAP (Final)	98.09%	97.74%	96.85%

YOLOv11 outperformed ResNet and EfficientNet in terms of accuracy and F1-score. This may be attributed to its single-stage detection framework, which integrates feature extraction and classification which results in its precise predictions with minimal latency. In contrast, ResNet and EfficientNet, despite their advanced architectures, showed lower recall and F1-scores. This is likely due to their reliance on sequential processing, smaller input image sizes (224×224 vs. YOLOv11 640×640), and higher sensitivity to input resizing. These factors reduced their ability to preserve spatial details crucial for distinguishing fine-grained features, particularly in complex scenarios such as detecting the Dangerous Driving class.

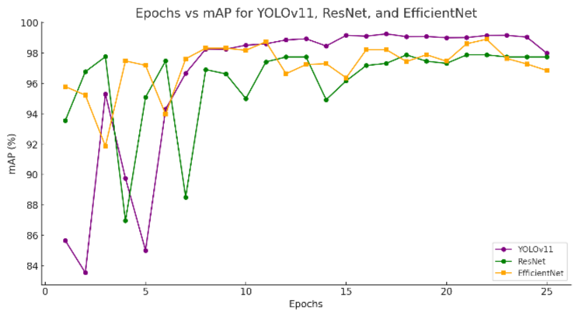


Fig. 3 Epochs vs mAP for YOLOv11, ResNet and EfficientNet

Fig. 3 illustrates the training and validation accuracy curves for all three models. YOLOv11 accuracy stabilized earlier, indicating faster convergence compared to ResNet and EfficientNet. Comparing Fig. 4, Fig. 5 and Fig. 6 their confusion matrix reveal that YOLOv11 achieved the highest true positive rates for both driving classes and exhibited fewer false negatives in the Dangerous Driving class, which is critical for safety applications. ResNet and EfficientNet, on the other hand, had slightly higher false negatives, potentially due to architectural limitations in their ability to process detailed spatial features.

Interestingly, as seen in Table 1, the mean Average Precision (mAP) scores for ResNet and EfficientNet were lower than those for YOLOv11. While the high accuracy and F1-scores of ResNet and EfficientNet suggest strong classification ability, their mAP scores reflect a limitation in balancing precision and recall across varying confidence thresholds. This can be attributed to a combination of factors: the smaller input sizes used for ResNet and EfficientNet led to a loss of critical spatial features, and their sequential processing architectures may struggle with optimizing precision and recall across all confidence thresholds. In contrast, YOLOv11 grid-based detection mechanism and end-to-end design allow it to preserve spatial details effectively and maintain high performance across varying thresholds.

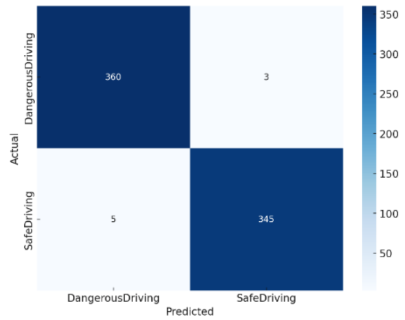


Fig. 4 Confusion Matrix for YOLOv11

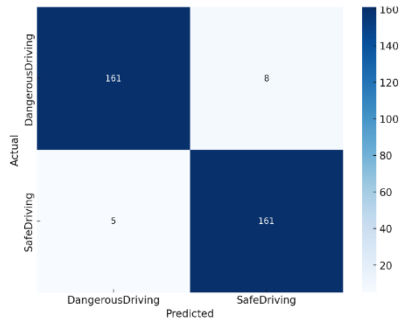


Fig. 5 Confusion Matrix for ResNet

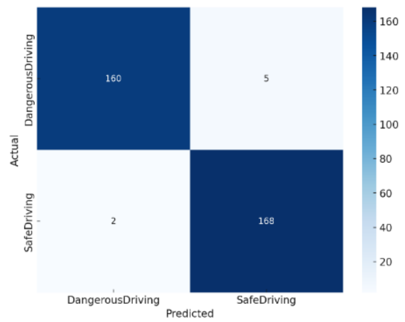


Fig. 6 Confusion Matrix for EfficientNet

4. Conclusion

This study compared the performance of YOLOv11, ResNet, and EfficientNet for driver fatigue detection using metrics such as accuracy, precision, recall, and F1-score. YOLOv11 outperformed the other models, achieving the highest accuracy (98.9%) and F1-score (98.8%), attributed to its single-stage detection framework and larger input resolution (640×640). ResNet and EfficientNet, while achieving competitive results with F1-scores of 96.8% and 97.5%, respectively, were limited by smaller input sizes (224×224) and higher sensitivity to image resizing, reducing their effectiveness in distinguishing complex features. These findings highlight YOLOv11 suitability for real-time driver monitoring systems due to its notable classification performance and computational efficiency. Future work shall focus on testing these models in more diverse and real-world driving scenarios to further validate their applicability.

References

1. Sagberg, F., *Road accidents caused by drivers falling asleep.* Accident analysis & prevention, 1999. 31(6): p. 639-649.
2. Siskind, V., et al., *Risk factors for fatal crashes in rural Australia.* Accident Analysis & Prevention, 2011. 43(3): p. 1082-1088.
3. Casey, G.J., T. Miles-Johnson, and G.J. Stevens, *Heavy vehicle driver fatigue: Observing work and rest behaviours of truck drivers in Australia.* Transportation Research Part F: Traffic Psychology and Behaviour, 2024. 104: p. 136-153.
4. Tamba, M., et al., *Classification of Autism Histogram of Oriented Gradient (HOG) Feature Extraction with Support Vector Machine (SVM) Method.* 2023 International Conference on Modeling & E-Information Research, Artificial Learning and Digital Applications (ICMERALDA), 2023: p. 144-148.
5. Agrawal, A., R. Gupta, and N.R.S. Jebaraj, *Advancing Support Vector Machines for Automated Medical Image Diagnosis.* 2024 International Conference on Optimization Computing and Wireless Communication (ICOCWC), 2024: p. 1-7.
6. Ram, D., D.J.V.S. Koushik, and H. Pavan, *Drowsiness Detection using EAR (Eye Aspect Ratio) by Machine Learning.* INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT, 2024. 08: p. 1-13.
7. Sri Mounika, T.V.N.S.R., et al. *Driver Drowsiness Detection Using Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR), and Driver Distraction Using Head Pose Estimation*. in *ICT Systems and Sustainability*. 2022. Singapore: Springer Nature Singapore.
8. Florez, R., et al., *A Real-Time Embedded System for Driver Drowsiness Detection Based on Visual Analysis of the Eyes and Mouth Using Convolutional Neural Network and Mouth Aspect Ratio.* Sensors, 2024. 24(19): p. 6261.
9. Ye, M., et al., *Driver fatigue detection based on residual channel attention network and head pose estimation.* Applied Sciences, 2021. 11(19): p. 9195.
10. Gowda, M.S., et al. *A Multimodal Approach to Detect Driver Drowsiness*. in *2024 5th International Conference on Circuits, Control, Communication and Computing (I4C)*. 2024.
11. Valuthottiyil Shajahan, T., B. Srinivasan, and R. Srinivasan, *Real-Time Fatigue Monitoring System in Diverse Driving Scenarios*. 2024. 124-127.
12. Jiang, H. and W. Xu. *A lightweight method for fatigue driving detection based on facial analysis*. in *2024 36th Chinese Control and Decision Conference (CCDC)*. 2024.
13. Alansari, M., et al., *GhostFaceNets: Lightweight Face Recognition Model From Cheap Operations.* IEEE Access, 2023. 11: p. 35429-35446.
14. Li, T., et al., *Advancements in Fatigue Detection: Integrating fNIRS and Non-Voluntary Attention Brain Function Experiments.* Sensors, 2024. 24(10): p. 3175.
15. Khanam, R. and M. Hussain, *YOLOv11: An Overview of the Key Architectural Enhancements.* ArXiv, 2024. abs/2410.17725.
16. Redmon, J., et al., *You Only Look Once: Unified, Real-Time Object Detection*. 2016. 779-788.
17. Jiang, P., et al., *A Review of Yolo Algorithm Developments.* Procedia Computer Science, 2022. 199: p. 1066-1073.
18. Jones, A., *Cutting-Edge Detection of Fatigue in Drivers: A Comparative Study of Object Detection Models*. 2024.
19. Guan, Z., *Research Progress on the Development of Fatigue Driving Detection Based on Deep Learning.* Theoretical and Natural Science, 2024. 52: p. 128-136.
20. Careem, R., M.G. Md Johar, and A. Khatibi, *Deep neural networks optimization for resource-constrained environments: techniques and models.* Indonesian Journal of Electrical Engineering and Computer Science, 2024. 33: p. 1843.
21. Alghanim, M., et al., *A Hybrid Deep Neural Network Approach to Recognize Driving Fatigue Based on EEG Signals.* International Journal of Intelligent Systems, 2024. 2024.
22. Khan, T., G.A.-O. Choi, and S.A.-O. Lee, *EFFNet-CA: An Efficient Driver Distraction Detection Based on Multiscale Features Extractions and Channel Attention Mechanism. LID - 10.3390/s23083835 [doi] LID - 3835.* (1424-8220 (Electronic)).
23. Prajwal, T.S. and I.A. K, *A Comparative Study Of RESNET-Pretrained Models For Computer Vision*, in *Proceedings of the 2023 Fifteenth International Conference on Contemporary Computing*. 2023, Association for Computing Machinery: Noida, India. p. 419–425.
24. Tajane, K., et al., *EffiConvRes: An Efficient Convolutional Neural Network with Residual Connections and Depthwise Convolutions*. 2023. 1-5.
25. Alif, M.A.R., *YOLOv11 for Vehicle Detection: Advancements, Performance, and Applications in Intelligent Transportation Systems.* Cornell University, 2024.
26. Shovo, S., et al., *Advancing low-light object detection with you only look once models: An empirical study and performance evaluation.* Cognitive Computation and Systems, 2024. 6: p. 119-134.
27. Lin, C., et al., *Efficient and accurate compound scaling for convolutional neural networks.* Neural Networks, 2023. 167: p. 787-797.
28. Model, D. *Driver Detection Dataset*. 2024 [cited 2025 January 16]; Open Source Dataset]. Available from: https://universe.roboflow.com/drowsy-model/driver-detection-ajfzd.
29. Zhuo, S., et al. *on YOLOv5 and Dlib*. 2023. Singapore: Springer Nature Singapore.
30. Wang, J. and Z. Wu, *Driver distraction detection via multi-scale domain adaptation network.* IET Intelligent Transport Systems, 2023. 17(9): p. 1742-1751.

Authors Introduction

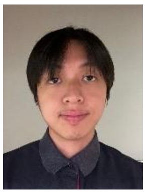
Mr. Hao Feng Chan

He is currently an undergraduate student majoring in Bachelor of Mechatronics Engineering (Honours) at Deakin University, Australia.

Mr. Shakir Hussain Naushad Mohamed

He is currently an undergraduate student majoring in Bachelor of Electrical and Electronics Engineering (Honours) at Deakin University, Australia.

Mr. Dexter Sing Fong Leong



He is currectly an undergrate pursuing Bachelors Of Mechatronic Engineering (Honours) in Deakin University, Australia.

Mr. Chau Wui Chung Alton

He is currently an undergraduate student in Bachelor of Mechanical Engineering (Honours) at Deakin University, Australia.

Mr. Zheng Cai

He is currently a PhD candidate at Deakin University's Institute for Intelligent Systems Research and Innovation (IISRI). His research interests include multiobjective optimisation algorithms such as metaheuristic algorithms and evolutionary algorithms for scheduling problems. He is also exploring the integration of machine learning with optimisation algorithms.

Ms Deng Xinjie

She is pursuing a PhD in Information Technology at the Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, with her research centered on creating lightweight deep learning algorithms for computer vision applications.

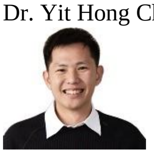
Andi Prademon Yunus, Ph.D

He is an Assistant Professor at Telkom University, and he received his PhD in Engineering from Mie University, Japan. His research focuses on applied and fundamental machine learning for motion and behavior computing. He also collaborates with industry partners to develop AI-based tools for language modeling and image analytics.

Dr. Takao Ito

He is Professor of Management of Technology (MOT) in Graduate School of Advanced Science and Engineering at Hiroshima University. His current research interests include automata theory, artificial intelligence, systems control, quantitative analysis of interfirm relationships using graph theory, and engineering approach of organizational structures using complex systems theory.

Dr. Yit Hong Choo



He has completed his PhD and is now a Research Fellow in Operations Analytics at Deakin University's Institute for Intelligent Systems Research and Innovation (IISRI), supported by the Rail Manufacturing Cooperative Research Centre (RMCRC). His research focuses on advanced multi-objective optimisation algorithms for complex maintenance scheduling in rolling stock. He collaborates with transportation industry partners to develop AI-based tools for video and image analytics.

Optimizing Face Embedding Sizes and Accuracy in Facial Recognition Systems

Wui Chung Alton Chau, Hao Feng Chan, Dexter Sing Fong Leong, Shakir Hussain Naushad Mohamed
School of Engineering, Deakin University, Australia

Andi Prademon Yunus
Telkom University, Indonesia

Takao Ito
Hiroshima University, Japan

Zheng Cai, Xinjie Deng, Yit Hong Choo*
Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Australia
Email: y.choo@deakin.edu.au

Abstract

Face recognition technology is integral to security, access control and identity verification in finance, healthcare and transportation. It protects personal data, secures online transactions, controls access to areas, and helps prevent identity theft. This paper proposes a novel hybrid optimization algorithm, Moss Particle Swarm Optimization (MPSO) to perform hyperparameter tuning, aiming to identify the neural network, alongside Triple Loss metrics for efficient recognition. The proposed method is tested with the Labeled Faces in the Wild (LFW) dataset, demonstrating its effectiveness in improving facial recognition systems.

Keywords: Face Recognition, Optimization, Hyperparameter Tuning, Embedding Sizes, Deep Learning

1. Introduction

Face recognition technology has emerged as a cornerstone of the computer vision industry, with diverse applications spanning mobile phone security, identity verification, and access control [1]. The performance of these systems relies heavily on the optimal representation of facial features, which involves complex trade-offs between computational efficiency and recognition accuracy [2]. This balance is achieved through the careful tuning of model configurations, particularly embedding sizes and hyperparameters, which determine how well a model generalizes to unseen data. By refining these aspects, face recognition systems can achieve robust performance, safeguarding sensitive personal data and ensuring reliable operations across industries such as finance, healthcare, and transportation [3].

Achieving state-of-the-art performance in face recognition systems often involves leveraging deep learning architectures like FaceNet [4], ResNet [5], and VGGFace [6]. These models employ sophisticated loss functions such as Center Loss [7], Cross-Entropy Loss [8], and Triplet Loss [9] to learn discriminative facial feature embeddings. However, beyond architectural design, hyperparameter tuning plays a pivotal role in optimizing these systems. Hyperparameters define the configuration before training and influence both the training process and the final predictive performance [10]. Effective hyperparameter optimization ensures that models not only perform well on training data but also generalize to challenging real-world scenarios. In practical applications, these models must be capable of accurately recognizing faces under a wide variety of conditions such as low lighting, partial occlusions, and changes in facial expression without requiring significant retraining. This level of robustness can only be achieved through the optimization of both the model architecture and the hyperparameters. Hyperparameter Tuning has also increasingly been tackled using metaheuristic algorithms due to their speed, flexibility, and ability to explore complex and high dimensional search spaces efficiently. Metaheuristic algorithms are particularly suited for such tasks due to their flexibility and ability to navigate large, complex search spaces that traditional methods struggle with [11]. By applying techniques like Genetic Algorithms (GA) [12], Whale Optimization Algorithm (WOA) [13] and Harris' Hawks Optimization (HHO) [14], researchers can fine-tune deep learning models, ensuring that they maintain high performance across a range of real-world conditions. These metaheuristic algorithms are particularly useful in machine learning for neural architecture search [15], hyperparameter tuning, and feature selection. Thanathamathee et al. [16] used grid search and nested cross-validation to enhance facial and masked facial recognition, while Ozcan et al. [17] applied Particle Swarm Optimization (PSO) [18] in conjunction with transfer learning to improve expression recognition systems. The emergence of advanced algorithms like Moss Growth Optimization (MGO) [19] further highlights the importance of exploring innovative approaches for hyperparameter tuning, as these techniques have demonstrated success in solving complex optimization problems across various domains. However, choosing the right algorithm is challenging due to problem-specific characteristics and performance metrics. The complexity of modern machine learning models makes manual tuning impractical and increases the risk of suboptimal solutions. Thus, efficient and robust hyperparameter tuning methods are essential to improve model performance.

This study introduces a new approach, Hybrid Moss Particle Swarm Optimization (HMPSO). In Section 2, a review on metaheuristic algorithms, hyperparameter tuning on facial recognition systems are provided. In Section 3, the proposed hybridization of PSO and MGO, into HMPSO, is explained in detail. In Section 4, the effectiveness of HMPSO in enhancing the accuracy of the ResNet-18 by hyperparameter tuning is evaluated, compared, discussed. Concluding remarks for future work are presented in Section 5.

2. Deep Learning CNN model

Deep Convolutional Neural Networks (DCNNs) are widely used in computer vision tasks to identify patterns in images and videos. In facial recognition and image processing, DCNNs like ResNet-18 [5] shown in Figure 1 can effectively identify and extract features from facial images, making them a robust solution for tasks such as facial recognition.

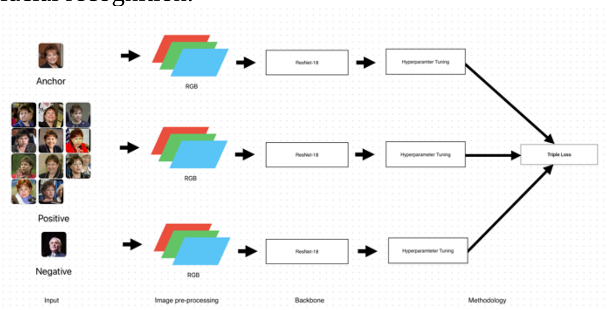


Figure 1 The Structure of the ResNet-18 with Hyperparameter Tuning and Triple Loss Metrics

The Labelled Faces in the Wild (LFW) dataset is divided into Anchor, Positive, and Negative triplets. Each image is resized to a new resolution in RGB colors. The ResNet-18 model generates embedding vectors for each image in the triplet, and hyperparameter tuning is used to optimize the embedding size and other parameters. These embeddings are then passed to the Triplet Loss function for results.

2.1 Facial Image Pre-processing

Image pre-processing involves transforming raw image data into a usable and meaningful format, eliminating unwanted distortions and enhancing specific features critical for computer vision applications. It serves as an essential step in preparing image data before inputting it into machine learning models. Instead of using grayscale images, RGB (Red, Green, Blue) color images are typically employed, as they provide a more detailed and comprehensive representation of facial features. In RGB images, each pixel is represented by three intensity values corresponding to the red, green, and blue color channels, enabling the capture of a broader range of colors and subtle variations in facial features.

2.2 ResNet-18

The ResNet-18 architecture is a lightweight deep neural network ideal for face recognition tasks. It balances high accuracy with computational efficiency, making it suitable for resource-constrained environments. ResNet-18 extracts face embeddings through optimized convolutional layers, which are then used in a triplet loss function to ensure embeddings of the same identity are closer in feature space. Its compact design and strong performance make it well-suited for real-world face recognition applications.

2.3 Embeddings and Triple Loss Function

Face embeddings represent facial features as vector arrays for comparison via distance, similarity, or search. This study examines embedding sizes between 64 and 256. Smaller embeddings reduce latency, while larger ones improve accuracy but require more resources. Using ResNet-18 and the triplet loss function, the model optimizes the embedding space for facial recognition. Triplet loss ensures embeddings of the same identity are closed while keeping different identities apart, is expressed as:

$$L = \max\left(d(a,p) - d(a,n) + m, 0\right) \qquad (1)$$

where a, p, and n are the anchor, positive, and negative samples respectively; d(a,p) and d(a,n) are the distance functions; and m is the margin ensuring sufficient separation. This setup enhances identity discrimination for verification and identification tasks. The lightweight ResNet-18 architecture ensures efficient training and embedding optimization. Experiments focus on minimizing triplet loss while maintaining high accuracy, optimizing performance across embedding sizes for scalable and effective facial recognition systems.

2.4 LFW Dataset

The Labelled Faces in the Wild dataset (LFW) [20] is designed for studies in unconstrained face recognition. It comprises over 13,000 labelled face images, including 1,680 individuals with multiple distinct photos, all collected from the web. For this study, 2,950 selected images were selected to evaluate the face recognition performance of the ResNet-18. The original image resolution of these images 250x250 are resized into a smaller resolution for the computational efficiency of the embedding size chosen by the ResNet-18 model.

3. Methodology

The proposed Hybrid Moss Growth Particle Swarm Optimization (HMPSO) algorithm enhances the optimization performance of the ResNet-18 by combining key elements of Particle Swarm Optimization (PSO) and Moss Growth Optimization (MGO). In Figure 2, it demonstrates the structure of the HMPSO. The PSO component provides global search capabilities by guiding particles using personal and global best positions, while MGO introduces adaptive growth dynamics to explore diverse regions and prevent premature convergence. By combining PSO's convergence ability with MGO's adaptability, HMPSO balances exploration and exploitation, improving search efficiency and solution quality. This hybrid approach enhances robustness and flexibility in navigating complex optimization landscapes, making it effective for finding optimal configurations for the ResNet-18 model.

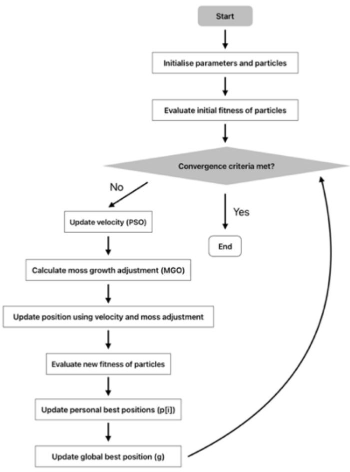


Figure 2 Flowchart of the Hybrid Moss Particle Swarm Optimization (HMPSO)

3.1 *Velocity Update*

Particle Swarm Optimization (PSO) is used to control the particle velocities and ensure effective exploration of the search space. PSO employs the concept of inertia weights, cognitive factors, and social factors to guide each particle based on its personal best position and the global best position found by the swarm. The velocity update equation in PSO helps particles move through the search space, allowing them to converge toward an optimal solution over time. This is given by the following equation:

$$v_i(t+1) = w \cdot v_i(t) + c_1 \cdot r_1 \cdot (p_i(t) - x_i(t)) + c_2 \cdot r_2 \cdot (g(t) - x_i(t)) \quad (2)$$

Where $v_i(t)$ is the velocity of particle i at time t, $x_i(t)$ is the position of particle i at time t, w is the inertia weight, c_1, c_2 are the cognitive and social scaling factors, r_1, r_2 are the random numbers between 0 and 1, $p_i(t)$ is the personal best position of particle I, g(t) is the global best position across all particles at time t. This equation drives the particles to explore the search space effectively and converge toward the best-found solutions, maintaining a balance between exploration and exploitation.

3.2 *Moss Growth Update*

The addition of moss growth dynamics introduces an additional element of growth adjustment in the position update. This adjustment is based on a dynamic model that simulates the growth of moss over time. The moss growth model enables particles to experience an additional growth force that alters their position, enhancing both the exploration and convergence processes. The moss growth adjustment is given by the following formula:

$$\Delta x_i = \frac{K}{1 + (\frac{K - x_i}{x_i})e^{rt}} - x_i \quad (3)$$

Where Δx_i is the moss growth adjustment for the particle I, K is the constant related to the moss growth model, r is the rate constant associated with moss growth dynamics, t is the time or iteration index, e^{rt} is the exponential growth term, simulating moss growth over time, x_i is the current position of particle i. This equation introduces an exponential growth effect that alters the particle's position, enabling enhanced exploration of the search space by simulating the natural expansion of moss.

3.3 *Position Update Rule*

The overall position update in Hybrid MPSO combines the classical PSO position update with the moss growth adjustment. This results in the following combined position update equation:

$$x_i(t+1) = x_i(t) + v_i(t+1) + \Delta x_i \quad (4)$$

Where $x_i(t+1)$ is the updated position of particle i at time t+1, $v_i(t+1)$ is the updated velocity based on PSO equation, Δx_i is the moss growth adjustment based on the moss growth model. This combined position update rule ensures that particles are guided both by their velocities, based on their personal and global best positions, and by the moss growth dynamics, which provide an additional layer of exploration.

4. Experimental Results and Discussion

The result demonstrates the effectiveness of the proposed novel approach, Hybrid Moss Particle Swarm Optimization (HMPSO) by comparing with the Particle Swarm Optimization (PSO) and the Moss Growth Optimization (MGO). As shown in Table 1, 2 and 3, each optimization runs 3 times to get 3 different sets of best hyperparameters. And Figure 3, 4 and 5 shows the training history of loss and accuracy of PSO, MGO and HMPSO respectively. In addition, comparing the average accuracies between HMPSO, PSO and MGO shown in Table 4, we conducted that HMPSO has better sets of hyperparameters chosen and achieve the highest average accuracy with a 97.23%, PSO the second with 94.82 % and MGO the last with 85.71%. According to Figure 5, the first run of HMPSO has the reached 100% accuracy during training in the last epoch a few times, implies that the model is performing well.

According to Figure 5, the training and validate accuracy vs 20 epochs graph, shows that the HMPSO has reached above 90% accuracy a few times during the last epochs when training the validating the model, as well as PSO shown in Figure 3 also achieved above 90% accuracies during training and validating but HMPSO is slightly a bit more consistent than PSO, as the second run of PSO drops to an accuracy of 86.59%. According to Figure 5, the training and validation loss versus 20 epochs graph shows a clear trend of the ResNet-18 has been learning effectively. The training loss starts at a relatively high value of 0.3920, steadily decreasing to 0.0316 by the 20th epoch, demonstrating the consistent improvement in fitting the training data. Similarly, the validation loss follows a complementary trend, indicating that the model is not only reducing errors on the training set but is also generalizing well to unseen data. This alignment between training and validation loss trends highlights that the optimization process is functioning effectively, with no apparent overfitting or underfitting during the training period.

As a result, the Hybrid Moss Particle Swarm Optimization (HMPSO) outperforms the Particle Swarm Optimization (PSO) and the Moss Growth Optimization (MGO), achieving the highest average accuracy of **97.23%** and demonstrating greater consistency across runs. The training and validation loss trends confirm effective learning and generalization without overfitting, highlighting HMPSO's potential as a robust hyperparameter optimization method for deep learning tasks.

Table 1 Hyperparameter Tuning and Accuracy Optimization performed by PSO.

Particle Swarm Optimizer (PSO)	Run 1	Run 2	Run 3
Embedding Size	146	224	100
Margin	0.7437	1.0	0.4006
Batch Size	46	98	95
Learning Rate	0.0001	0.01	0.0003
Loss	**0.0495**	**0.4187**	**0.0161**
Accuracy (%)	**98.63**	**86.59**	**99.25**
Avg Accuracy (%)		**94.82**	

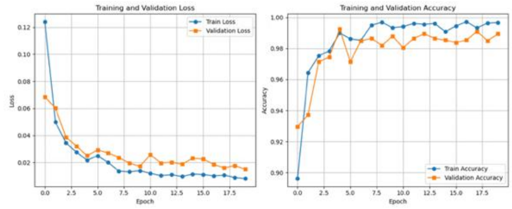


Figure 3 Train and Validate Loss vs Epochs graph (Left), Train and Validate vs Epochs graph (Right) of PSO.

Table 2 Hyperparameter Tuning and Accuracy Optimization performed by MGO

Moss Growth Optimizer (MGO)	Run 1	Run 2	Run 3
Embedding Size	22	72	197
Margin	0.2908	0.9989	0.4819
Batch Size	32	16	83
Learning Rate	0.0069	0.0057	0.0069
Loss	**0.0967**	**0.0734**	**0.1863**
Accuracy (%)	**86.77**	**87.03**	**83.33**
Avg Accuracy (%)		**85.71**	

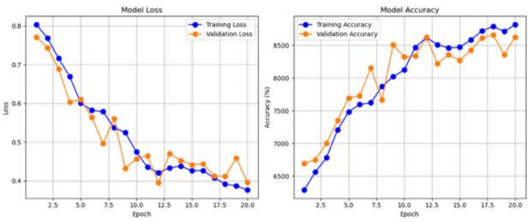


Figure 4 Train and Validate Loss vs Epochs graph (Left), Train and Validate vs Epochs graph (Right) of MGO

Table 3 Hyperparameter Tuning and Accuracy Optimization performed by HMPSO

Hybrid Moss Particle Swarm Optimizer (HMPSO)	Run 1	Run 2	Run 3
Embedding Size	141	64	224
Margin	0.8030	0.9753	0.3226
Batch Size	49	16	102.3274
Learning Rate	0.0001	0.0001	0.0022
Loss	**0.0592**	**0.1134**	**0.0454**
Accuracy (%)	**99.22**	**98.12**	**94.35**
Avg Accuracy (%)		**97.23**	

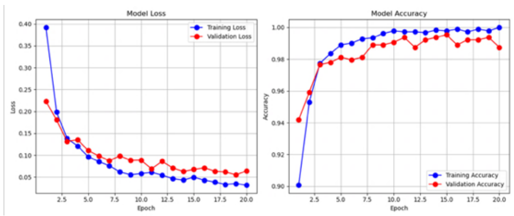


Figure 5 Train and Validate Loss vs Epochs graph (Left), Train and Validate vs Epochs graph (Right) of HMPSO

Table 4 Comparison of Average Accuracy in PSO, MGO and HMPSO with ResNet-18 model

Optimizers	PSO	MGO	HMPSO
Avg Accuracy (%)	94.82	85.71	**97.23**

5. Conclusion

In conclusion, the HMPSO algorithm introduced in this study represents a significant advancement in optimizing hyperparameters for deep learning models, specifically applied to the ResNet-18 architecture for face recognition. By combining PSO and MGO, HMPSO creates a powerful hybrid approach that enhances the accuracy and efficiency of hyperparameter tuning. It focuses on key hyperparameters such as face embedding sizes, the margin from the triple loss function, batch size, and learning rate, which are crucial for model performance. The experimental results demonstrated that HMPSO significantly outperforms traditional optimization methods, showing better accuracy and computational efficiency on the LFW dataset. This combination of PSO's global search and MGO's local search offers a robust solution, improving security, identity verification, and biometric authentication in real-world applications. Additionally, metaheuristic algorithms like PSO and MGO play a crucial role in solving complex optimization problems by exploring large solution spaces more effectively. This hybrid method shows potential for expanding into other domains such as natural language processing, computer vision, and reinforcement learning, where fine-tuning hyperparameters is essential. Overall, this research emphasizes the value of innovative optimization strategies for improving deep learning models, particularly in security and identity verification systems, paving the way for further advancements in various fields requiring intelligent decision-making and pattern recognition.

References

1. D. Kothadiya, A. Chaudhari, R. Macwan, K. Patel, and C. Bhatt, "The convergence of deep learning and computer vision: smart city applications and research challenges," in 3rd International Conference on Integrated Intelligent Computing Communication & Sec
2. R. Archana and P. E. Jeevaraj, "Deep learning models for digital image processing: a review," *Artificial Intelligence Review,* vol. 57, no. 1, p. 11, 2024.
3. I. Dijmărescu, M. Iatagan, I. Hurloiu, M. Geamănu, C. Rusescu, and A. Dijmărescu, "Neuromanagement decision making in facial recognition biometric authentication as a mobile payment technology in retail, restaurant, and hotel business models," Oeconomia C
4. F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 815-823, 2015.
5. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
6. O. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *BMVC 2015-Proceedings of the British Machine Vision Conference 2015*, 2015: British Machine Vision Association.
7. Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in Computer vision–ECCV 2016: 14th European conference, amsterdam, the netherlands, October 11–14, 2016, proceedings, part VII 14, 2016: Springer
8. Z. Zhang and M. R. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," presented at the Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montréal, Canada, 2018.
9. I. Haider, H.-J. Yang, G.-S. Lee, and S.-H. Kim, "Robust human face emotion classification using triplet-loss-based deep CNN features and SVM," *Sensors,* vol. 23, no. 10, p. 4770, 2023.
10. T. Yu and H. Zhu, "Hyper-parameter optimization: A review of algorithms and applications," *arXiv preprint arXiv:2003.05689,* 2020.
11. F. Peres and M. Castelli, "Combinatorial optimization problems and metaheuristics: Review, challenges, design, and development," *Applied Sciences,* vol. 11, no. 14, p. 6449, 2021.
12. A. Sohail, "Genetic algorithms in the fields of artificial intelligence and data sciences," *Annals of Data Science,* vol. 10, no. 4, pp. 1007-1018, 2023.
13. Z. Cai, Y. H. Choo, V. Le, C. P. Lim, and M. Liao, "Enhancing the Whale Optimisation Algorithm with sub-population and hybrid techniques for single-and multi-objective optimisation," *Soft Computing,* vol. 28, no. 5, pp. 3941-3971, 2024.
14. Y. H. Choo, Z. Cai, V. Le, M. Johnstone, D. Creighton, and C. P. Lim, "Enhancing the Harris' Hawk optimiser for single- and multi-objective optimisation," *Soft Computing,* vol. 27, no. 22, pp. 16675-16715, 2023/11/01 2023, doi: 10.1007/s00500-023-08952-w.
15. T. Elsken, J. H. Metzen, and F. Hutter, "Neural architecture search: a survey," *J. Mach. Learn. Res.,* vol. 20, no. 1, pp. 1997–2017, 2019.
16. P. Thanathamathee, S. Sawangarreerak, P. Kongkla, and D. N. M. Nizam, "An optimized machine learning and deep learning framework for facial and masked facial recognition," *Emerging Science Journal,* vol. 7, no. 4, pp. 1173-1187, 2023.
17. T. Ozcan and A. Basturk, "Static facial expression recognition using convolutional neural networks based on transfer learning and hyperparameter optimization," *Multimedia Tools and Applications,* vol. 79, no. 35, pp. 26587-26604, 2020.
18. J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-international conference on neural networks,* 1995, vol. 4: ieee, pp. 1942-1948.
19. B. Zheng, Y. Chen, C. Wang, A. A. Heidari, L. Liu, and H. Chen, "The moss growth optimization (MGO): concepts and performance," *Journal of Computational Design and Engineering,* vol. 11, no. 5, pp. 184-221, 2024, doi: 10.1093/jcde/qwae080.
20. G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments," presented at the Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Ma

Authors Introduction

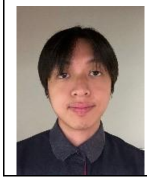
Mr. Hao Feng Chan

He is currently an undergraduate student majoring in Bachelor of Mechatronics Engineering (Honours) at Deakin University, Australia.

Mr. Shakir Hussain Naushad Mohamed

He is currently an undergraduate student majoring in Bachelor of Electrical and Electronics Engineering (Honours) at Deakin University, Australia.

Mr. Dexter Sing Fong Leong



He is currectly an undergrade pursuing Bachelors Of Mechatronic Engineering (Honours) in Deakin University, Australia.

Mr. Chau Wui Chung Alton

He is currently an undergraduate student majoring in Bachelor of Mechanical Engineering (Honours) at Deakin University, Australia.

Mr. Zheng Cai

He is currently a PhD candidate at Deakin University's Institute for Intelligent Systems Research and Innovation (IISRI). His research interests include multiobjective optimisation algorithms such as metaheuristic algorithms and evolutionary algorithms for scheduling problems. He is also exploring the integration of machine learning with optimisation algorithms.

Ms Deng Xinjie

She is pursuing a PhD in Information Technology at the Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, with her research ed on creating lightweight deep learning algorithms for computer vision applications.

Dr. Andi Prademon，Yunus, Ph.D

He is an Assistant Professor at Telkom University and he received his PhD in Engineering from Mie Unviersity, Japan. His research focuses on applied and fundamental machine learning for motion and behavior computing. He also collaborates with industry partners to develop AI-based tools for language modeling and image analytics.

Dr. Takao Ito

He is Professor of Management of Technology (MOT) in Graduate School of Advanced Science and Engineering at Hiroshima University. His current research interests include automata theory, artificial intelligence, systems control, quantitative analysis of interfirm relationships using graph theory, and engineering approach of organizational structures using complex systems theory.

Dr. Yit Hong Choo

He has completed his PhD and is now a Research Fellow in Operations Analytics at Deakin University's Institute for Intelligent Systems Research and Innovation (IISRI), supported by the Rail Manufacturing Cooperative Research Centre (RMCRC). His research focuses on advanced multi-objective optimisation algorithms for complex maintenance scheduling in rolling stock. He collaborates with transportation industry partners to develop AI-based tools for video and image analytics.

Suspicious Behavior Detection Using Computer Vision

Dexter Sing Fong Leong, Hao Feng Chan, Shakir Hussain Naushad Mohamed, Wui Chung Alton Chau
School of Engineering, Deakin University, Australia

Andi Prademon Yunus
Telkom University, Indonesia

Takao Ito
Hiroshima University, Japan

Zheng Cai, Xinjie Deng, Yit Hong Choo*
Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Australia
**Email: s223026243@deakin.edu.au, y.choo@deakin.edu.au*

Abstract

Detecting suspicious activity is a crucial task for public safety. The determination of class suspicious behavior is based on the facial cues of a person. Research has been conducted in this field using computer vision tools. However, accuracy still has room for improvement. Hence, this paper aims to use a novel approach in using other deep learning models to classify behavior as either suspicious or normal based on facial cues. By enhancing the detection process, this paper contributes to improving the reliable and effective surveillance system.

Keywords: Facial Cue, Computer Vision, Deep Learning Models, Suspicious, Surveillance

1. Introduction

Detecting suspicious behavior accurately and efficiently is critical to public safety and security. In the context of increasing global security concerns, surveillance systems equipped with computer vision technology provide an essential system for monitoring public and private spaces. However, while traditional surveillance systems are adept at capturing and recording vast amounts of footage, these systems often fall short of actively analyzing and interpreting behavioral cues that may indicate suspicious activities [1]. Enhancing these systems with the capability to detect subtle facial expressions and cues can significantly improve threat identification and response times, thereby bolstering overall security protocols [2].

The current surveillance system incorporated with computer vision predominantly utilizes object detection and motion tracking technologies [3]. Object detection involves identifying and classifying objects of interest within video feeds, such as people, vehicles, or specific items like weapons. This method uses deep neural networks (DNNs) [4] to automatically classify objects in real time, helping to quickly pinpoint elements that may be relevant to security concerns. Motion tracking, on the other hand, refers to the process of following the movement of identified objects across multiple frames of a video, allowing the system to track individuals or objects as they move through different camera views. This technology is crucial for monitoring activities that occur over time, such as detecting the movement of individuals in restricted areas or tracking the behavior of a person across a public space. While both object detection and motion tracking are foundational for surveillance systems, they primarily focus on recognizing objects and their movement, which is insufficient for identifying subtle behavioral cues that may indicate suspicious activities or criminal intent. Hence, this study focuses on incorporating facial expression cues analysis. There is a substantial gap in the deployment of facial cue analysis within these systems, which could otherwise provide insights into the intent of an individual or emotional state [5] [6]. This gap highlights the need for an integrated approach that combines robust object detection with advanced facial expression analysis.

DenseNet [7, 8] is characterized by the unique structure, which connects each layer to every other layer in a feed-forward manner. This dense connectivity ensures that all layers directly receive feature maps from preceding layers, which promotes feature reuse and substantially reduces the vanishing gradient problem. This architecture arranges the network into several densely connected blocks separated by transition layers that perform convolution and pooling to reduce dimensionality and manage parameter growth.

ResNet18 [9] is designed to facilitate the training of deeper neural networks through residual blocks, which incorporate skip connections. These skip connections enable the network to bypass certain layers by adding the input of the block directly to the output, allowing the network to learn identity functions more effectively. This mechanism mitigates the risk of performance degradation as the network depth increases. By enabling the training of deeper networks without a proportional increase in training difficulty, the ResNet architecture improves both the efficiency and scalability of neural network training.

A further enhancement in the performance of ResNet was done by incorporating Squeeze-and-Excitation (SE) block [10]. The SE block performs dynamic channel wise feature recalibration, significantly improving the ability of the network to emphasize important features while suppressing less useful ones. This process is achieved by first squeezing global spatial information into a channel descriptor through global average pooling, followed by excitation, which recalibrates the feature maps based on the descriptor. This approach helps the network focus on the most relevant features for the task at hand. This mechanism is particularly beneficial for tasks requiring high sensitivity in feature detection, such as facial expression analysis in security systems, where accurately detecting subtle variations in facial expressions is crucial for identifying suspicious behavior.

This study aims to bridge the existing technology gap by employing advanced deep learning models ResNet 18 [9], DenseNet [7, 8], and SEResNet18 [10] to analyze facial cues for suspicious behavior detection. By training these models with custom datasets of micro facial expressions, this research seeks to improve the accuracy and reliability of automated threat detection, thus contributing to safer public environments [11].

2. Methodology

In this study, a robust methodology was employed to enhance the predictive accuracy of neural networks in detecting suspicious behaviors, utilizing facial expressions as primary indicators. A dataset was compiled, encompassing a wide range of facial expressions, categorized into two classes: suspicious (Class 1) and non-suspicious (Class 0). Advanced data management ensured uniformity and quality. Specialized deep learning architectures such as deep learning models ResNet 18 [9], DenseNet [7, 8], and the proposed SEResNet18 [10] were chosen for their effectiveness in image analysis. These models were trained and validated against specific metrics to measure accuracy and generalization across unseen data, contributing to advancements in automated surveillance technologies. The performance of the models [12] will be assessed using Mean Average Precision, Accuracy, Recall, Precision, and confusion matrix metrics [13]. These metrics are essential for evaluating the ability of the model to correctly identify suspicious behaviors while minimizing false positives, thus ensuring the reliability of surveillance interventions.

2.1. *Dataset Preparation*

The preparation of the dataset for this study involved a meticulous and structured process. The initial step comprised the acquisition of images from external datasets, each representing a diverse array of facial expressions. This selection was guided by a review focused on identifying facial cues commonly associated with suspicious behavior.

After acquisition, the images underwent a labeling process, wherein bounding box annotations were meticulously applied to accurately delineate facial features. The enhancement of efficiency in data handling and preparation was done in Roboflow platform [14]. The capability of this platform is known to streamline dataset organization, annotation, class customization, and format conversion, facilitating swift and precise data processing. Data preprocessing techniques such as cropping, resizing, and normalization were conducted on this platform to ensure uniformity across the dataset.

The processed images were then formatted into the Common Objects in Context (COCO) format [15], ensuring integration with the deep learning architectures utilized in this research. A quality assurance protocol was implemented, involving a review and refinement of annotations to ensure accuracy and consistency within the dataset.

The custom dataset [16] comprises 1599 RGB images with a nearly balanced distribution across each class. This dataset was methodically partitioned into training (73%), validation (18%), and testing sets (9%). This distribution aimed to support robust model training, facilitate effective parameter tuning, and enable comprehensive performance evaluation.

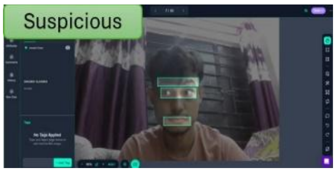


Fig. 1 Example of Suspicious Behavior.

Fig. 2 Example of Non-Suspicious Behavior.

2.2. *Models For Suspicious Behavior Detection*

In this study, three specialized convolutional neural network (CNN) architectures were employed to enhance the predictive accuracy in detecting suspicious behaviors via facial expressions (Fig.1, Fig.2). The models selected for this purpose include DenseNet [7], [8], ResNet 18 [9], and. SEResNet18 [10], each optimized for efficacy in processing complex image data.

DenseNet models, accessible through the 'torchvision.models.densenet' module in PyTorch [17], offer flexibility in configuring network depth and growth rates to meet various computational and application needs. ResNet 18, part of the standard offerings within the 'torchvision.models.resnet' module in PyTorch, is readily available for both research and production use. Additionally, this study introduces SEResNet, an enhanced version of ResNet 18, incorporating SE blocks into each residual unit [10]. The 'torchvision.models.resnet' module was modified to integrate these SE blocks, thus improving model performance.

The models used in this study are trained with key hyperparameters that significantly impact their performance. The learning rate is set to 0.001 and optimized using Adam. The perks of using Adam optimizer are the adaptive learning rate properties, which aids in efficient training. The batch size is set to 32, balancing computational efficiency and memory usage, while 20 epochs provide enough iterations for the model to learn without overfitting. Cross Entropy Loss is used as the loss function for classification tasks, optimizing model performance by comparing predicted probabilities with true labels. Additionally, data augmentation techniques like random horizontal flips and rotations are applied to improve model robustness and prevent overfitting.

2.3. *Metrics For Evaluation*

The evaluation of convolutional neural network models SEResNet 18, DenseNet, and ResNet 18 in this study involves several key metrics including Accuracy, Precision, Recall, F1-Score, and Mean Average Precision (mAP). Each metric contributes to understanding the effectiveness of these models in detecting suspicious behaviors based on facial cues.

This Accuracy quantifies the overall correctness of the models in classifying facial expressions either as suspicious or non-suspicious. This metric involves calculating the ratio of correct predictions (true positives and true negatives) to the total number of cases tested.

Precision [12] reflects the reliability of the model in identifying an expression as suspicious, determined by the ratio of true positives to the combined total of true positives and false positives. High Precision reduces unnecessary alarms in surveillance scenarios. Eq. (1) describes the formula for calculating Precision in a classification context where TP is the correctly predicted positive instances and FP is the incorrectly predicted as positive.

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)} \quad (1)$$

Recall [12], known as sensitivity, measures the capability of model to identify all relevant instances of suspicious expressions. This metric calculates the ratio of true positives to the sum of true positives and false negatives, ensuring comprehensive detection of suspicious activities. Eq. (2) presents the formula for calculating Recall in classification tasks where FN is the actual positive instances incorrectly predicted as negative.

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)} \quad (2)$$

F1-Score [12] represents the harmonic mean of Precision and Recall. This metric evaluates the balance between Precision and Recall, critical for informed decision-making in security contexts. Eq. (3) ensures that the F1-Score considers both the Precision (and the Recall of the test. By balancing these two metrics, the F1-Score serves as a useful measure when you need a single metric to evaluate the overall accuracy of a model, especially in scenarios where uneven class distribution might make Precision or Recall alone misleading. The use of the harmonic mean punishes extreme values, making the F1-Score a more robust measure that requires both Precision and Recall to be relatively high to achieve a high score.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

Mean Average Precision (mAP) [12] evaluates model consistency and reliability by measuring Precision at varying Recall levels, averaged over multiple thresholds. Eq. (4) presents the formulas for calculating Average Precision (AP) and mAP, where N is the number of classes, p(r) is Recall r, Ap indicates Average Precision for each class and mAP indicates Mean Average Precision. Eq. (4) assess object detection performance across classes.

$$AP = \int_0^1 p(r)\, dr$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (4)$$

A confusion matrix [13] describes the performance of a classification model with known true values. It details how predictions are distributed across categories. In this study, the confusion matrix has two classes: Class 1 for suspicious behavior and Class 2 for nonsuspicious behavior. An ideal model would show most detections in

TP (True Positives) and TN (True Negatives), while poor results indicate lower model performance.

3. Results And Discussion

3.1. *Results*

This study evaluates the performance of three convolutional neural network models. The primary model SEResNet 18 will be fairly compared with ResNet 18 and DenseNet for model evaluation. The task of these models is to detect suspicious behavior through facial cues based on the custom dataset. The following results will present Confusion Matrix, Precision, Recall, F1-Score, Accuracy and mAP. These results will reflect the competence of each model and provide a clear conclusion for the study.

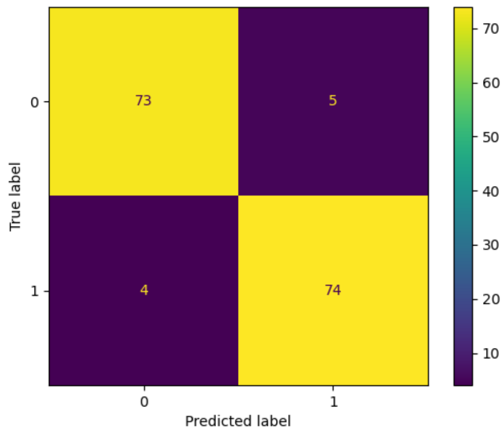


Fig. 3 SEResNet 18 Confusion Matrix.

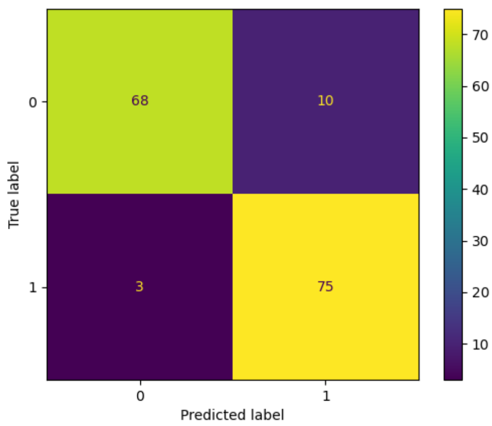


Fig. 4 ResNet 18 Confusion Matrix.

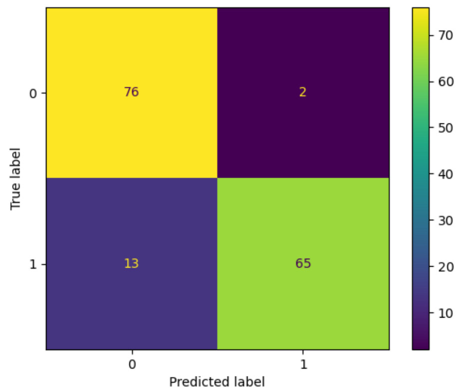


Fig. 5 DenseNet Confusion Matrix.

For SE-ResNet18, Fig. 3 showed 73 true positives and 74 true negatives, with minor misclassifications evident through 5 false positives and 4 false negatives. The performance of ResNet 18 in Fig. 4 was slightly lower with 68 true positives and 75 true negatives but had a higher number of false positives at 10, and fewer false negatives at 3. DenseNet (Fig. 5) demonstrated a strong ability to identify true positives (76) and true negatives (65) but struggled with Recall, evidenced by a higher number of false negatives (13) compared to only 2 false positives.

Table 1. Performance Result of Models.

	Suspicious Class			Accuracy (%)
	Precision (%)	Recall (%)	F1-Score (%)	
SEResNet 18	94	95	94	97
DenseNet	97	83	90	90
ResNet 18	88	96	92	92

SE-ResNet18 (Table 1) demonstrated the highest overall Accuracy at 97%, supported by impressive Precision, Recall, and F1-Scores of 94%, 95%, and 94% respectively. DenseNet followed with an Accuracy of 90%, boasting the highest Precision of 97% but the lowest Recall at 83%, which affected F1-Score of DenseNet, also at 90%. ResNet18 showed robust performance with an overall Accuracy of 92%, and while it had a lower Precision at 88%, it recorded the highest Recall of 96%, culminating in an F1-Score of 92%.

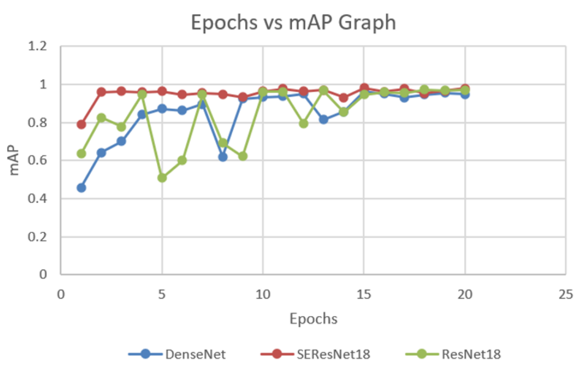


Fig. 6 mAP For Positive Detection Results.

Fig. 6 shows that DenseNet exhibited significant mAP variability, indicating performance instability. In contrast, SE-ResNet18 demonstrated stable, consistent mAP, while ResNet18, though slightly lower in mAP, was more consistent than DenseNet.

3.2. *Discussion*

The evaluation of SEResNet 18, alongside ResNet 18 and DenseNet, reveals a superior capability in detecting suspicious behaviors through the analysis of facial cues using a custom dataset. As the primary model, SEResNet 18 not only achieved the highest Accuracy, Precision, Recall, and F1-Scores but also demonstrated robustness across different testing conditions. This result suggests that the integration of Squeeze-and-Excitation (SE) blocks, which enhance feature recalibration capabilities, significantly contributes to the effectiveness of detection. Despite the standout performance of SEResNet 18, this model did not achieve 100% Accuracy, which underscores the presence of intrinsic challenges and room for improvement. Even though the custom dataset used in this study offers a substantial improvement in clarity compared to more generalized datasets, like crime UCF videos, it still presents limitations. The variation in facial cue clarity, especially under less-than-ideal conditions such as low light or partial obstructions, affects even the most advanced models like SEResNet 18. Enhancing dataset quality with more varied and challenging examples could help improve the robustness of models.

Despite SEResNet 18 overall efficacy, this model showed some fluctuations in Mean Average Precision (mAP) over training epochs, indicative of potential overfitting to the training data or an inability to generalize perfectly to new data. This suggests that while SEResNet 18 effectively learns from the dataset provided, performance of the model could be impacted by the presence of noise or non-representative data. While SEResNet 18 leads in many performance metrics, incorporating features from other architectures like dense connectivity of DenseNet could provide a balance between depth and feature richness, potentially leading to even higher accuracy and stability

In summary, while SEResNet 18 emerges as the most capable model for detecting suspicious behavior, further enhancements to the dataset and model architecture could elevate the accuracy and generalization capabilities of the trained model. Addressing these areas could make SEResNet 18 even more effective, paving the way for the deployment of model in surveillance applications requiring high reliability and Precision.

4. Conclusion

This research has validated the efficacy of advanced CNN models, particularly SEResNet 18, ResNet 18, and DenseNet, in detecting suspicious behaviors through facial cues. SEResNet 18 stood out for the superior performance in Accuracy, Precision, Recall, and F1-Scores, benefiting from SE blocks that enhance feature recognition. Despite these successes, none of the models achieved perfect accuracy, highlighting the need for improved datasets and model architectures that better mirror real-world complexities. Future research should focus on refining datasets and exploring hybrid architectures to enhance the generalization capabilities of these systems. This study underscores the potential of deep learning models to significantly improve automated surveillance, advancing public safety in both theory and application.

References

1. Jiyang Xie, Y.Z., Ruoyi Du, Weiyu Xiong, Yufei Cao, Zhanyu Ma, Dongpu Cao, and Jun Guo, *Deep Learning-Based Computer Vision for Surveillance in ITS: Evaluation of State-of-the-Art Methods.* IEEE Transactions on Vehicular Technology, 2021. 70(4): p. 3027–3040.
2. Harikrishnan, J., Sudarsan, Arya., Ajai, Remya A. S., Sadashiv, Aravind., *Vision-Face Recognition Attendance Monitoring System for Surveillance using Deep Learning Technology and Computer Vision*, in *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*. 2019, IEEE.
3. Haroon Idrees, M.S., and Ray Surette, *Enhancing camera surveillance using computer vision: a research note.* arXiv, 2018. 1808.03998.
4. Szegedy, C., Toshev, A., & Erhan, D, *Deep neural networks for object detection*, in *Advances in Neural Information Processing Systems (NIPS) 2013*. 2013, 1-9. p. 26.
5. Marco Leo, P.C., Pier Luigi Mazzeo, Paolo Spagnolo, Dario Cazzato, Cosimo Distante, *Analysis of Facial Information for Healthcare Applications: A Survey on Computer Vision-Based Approaches.* Information, 2020. 11(128).
6. Zhe Li, T.Z., Xiao Jing, Youning Wang, *Facial expression-based analysis on emotion correlations, hotspots, and potential occurrence of urban crimes.* Alexandria Engineering Journal, 2021. 60: p. 1411–1420.

7. Gao Huang, Z.L., Laurens van der Maaten, Kilian Q. Weinberger. *Densely Connected Convolutional Networks*. in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017. Honolulu, HI, USA: IEEE.
8. Yi Zhu, S.N. *DenseNet for Dense Flow*. in *IEEE International Conference on Image Processing (ICIP)*. 2017. Beijing, China: IEEE.
9. Kaiming He, X.Z., Shaoqing Ren, Jian Sun, *Deep Residual Learning for Image Recognition*, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, IEEE: Boston, MA, USA. p. 770–778.
10. Jie Hu, L.S., Gang Sun, *Squeeze-and-Excitation Networks*, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018)*. 2018, IEEE. p. 7132–7141.
11. Fiza Abdul Razzaq, W.T., Muhammad Abbas Chaudary, Muhammad Waqas, Sumaria Fareed, Shoaib Javaid. *Enhancing Public Safety: Detection of Weapons and Violence in CCTV Videos with Deep*. in *25th International Multitopic Conference (INMIC)*. 2023. IEEE.
12. Rafael Padilla, S.L.N., Eduardo A. B. da Silva. *A Survey on Performance Metrics for Object-Detection Algorithms*. in *Proceedings of the International Workshop on Signal Processing (IWSSIP)*. 2020. Niterói, Brazil: IEEE.
13. Mohammadreza Heydarian, T.E.D., Reza Samavi, *MLCM: Multi-Label Confusion Matrix*. IEEE Access, 2022. 10: p. 19083–19095.
14. Team, R. *Roboflow: End-to-End Computer Vision Workflow*. 2023; Available .
15. Bhadani, R. *How to Work with Object Detection Datasets in COCO Format*. 2022.
16. *Sus_or_non_v1*, D.S.F. Leong, Editor. 2025, Roboflow.
17. Team, P. *PyTorch Vision: Models*. 2023.

Authors Introduction

Mr. Dexter Sing Fong Leong

He is currently an undergraduate pursuing Bachelors Of Mechatronic Engineering in Deakin University, Australia.

Mr. Hao Feng Chan

He is currently an undergraduate student majoring in Bachelor of Mechatronics Engineering (Honours) at Deakin

Mr. Shakir Hussain Naushad Mohamed

He is currently an undergraduate student majoring in Bachelor of Electrical and Electronics Engineering (Honours) at Deakin University, Australia.

Mr. Chau Wui Chung Alton

He is currently an undergraduate student in Bachelor of Mechanical Engineering (Honours) at Deakin University, Australia.

Mr. Zheng Cai

He is currently a PhD candidate at Deakin University's Institute for Intelligent Systems Research and Innovation (IISRI). His research interests include multiobjective optimisation algorithms such as metaheuristic algorithms and evolutionary algorithms for scheduling problems. He is also exploring the integration of machine learning with optimisation algorithms.

Ms Deng Xinjie

She is pursuing a PhD in Information Technology at the Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, with her research centered on creating lightweight deep learning algorithms for computer vision applications.

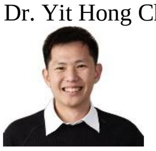
Andi Prademon Yunus, Ph.D

He is an Assistant Professor at Telkom University and he received his PhD in Engineering from Mie University, Japan. His research focuses on applied and fundamental machine learning for motion and behavior computing. He also collaborates with industry partners to develop AI-based tools for language modeling and image analytics.

Dr. Takao Ito

He is Professor of Management of Technology (MOT) in Graduate School of Advanced Science and Engineering at Hiroshima University. His current research interests include automata theory, artificial intelligence, systems control, quantitative analysis of interfirm relationships using graph theory, and engineering approach of organizational structures using complex systems theory.

Dr. Yit Hong Choo



He has completed his PhD and is now a Research Fellow in Operations Analytics at Deakin University's Institute for Intelligent Systems Research and Innovation (IISR). His research focuses on advanced multi-objective optimisation algorithms for complex maintenance scheduling in rolling stock. He also collaborates with transportation industry partners to develop AI-based tools for video and image analytics.

Sign Language Recognition Algorithms Using Hybrid Techniques

Shakir Hussain Naushad Mohamed, Hao Feng Chan, Dexter Sing Fong Leong, Wui Chung Alton Chau
School of Engineering, Deakin University, Australia

Andi Prademon Yunus
Telkom University, Indonesia

Takao Ito
Hiroshima University, Japan

Zheng Cai, Xinjie Deng, Yit Hong Choo*
Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Australia
**Email: y.choo@deakin.edu.au*

Abstract

Sign language recognition is a vital tool for enabling communication with individuals who are hearing impaired. This paper proposes a custom gesture recognition framework designed specifically for sign language interpretation. The proposed model incorporates a deep learning approach trained on a custom dataset. The system achieves robust recognition of complex gestures while maintaining efficiency. This framework emphasizes adaptability to variations in sign language styles.

Keywords: Sign language recognition, pose estimation, gesture recognition, deep learning

1. Introduction

Sign language is an important communication tool for the hearing and speech-impaired community, yet only a small percentage of people know how to use it, creating significant challenges for effective communication. An Australian census revealed that only slightly more than 16,000 people use Auslan, the sign language of the country, highlighting the limited reach and understanding of sign language among the general population[1]. Individuals who rely on sign language often face difficulties when interacting with those who are not proficient in its use. The language encompasses hand gestures, facial expressions, and body movements, providing a visual mode of communication distinct from spoken language.

Effective communication is essential in every society, but for individuals with speech and hearing impairments, it remains a persistent barrier. This gap is especially evident in sectors like education, healthcare, and public services, where individuals from the hearing- and speech-impaired community struggle to communicate with others who do not understand sign language. Traditional communication methods, such as gestures, may be slow and limited in conveying complex ideas, adding to the frustration of those who rely on them.

While sensor-based systems have been explored as potential solutions[2], they are often expensive and complex to implement. In contrast, vision-based systems, which utilize video feeds to capture sign language gestures, have gained popularity due to their accessibility and scalability. These systems are more affordable and less intrusive, as they do not require specialized hardware such as sensor gloves. Using deep learning techniques to process and recognize hand gestures, vision-based systems are a more practical solution to the communication challenges faced by the hearing- and speech-impaired community.

Despite advancements in sign language recognition, challenges persist, particularly in accurately recognizing gestures that vary across individuals. Differences in hand shapes, orientations, and signing styles make it difficult to build robust recognition systems. A key challenge is designing models that generalize well across different signers while maintaining high accuracy. One study successfully demonstrated the application of CNNs in building a sign language recognition system for Indian Sign Language (ISL)[3], demonstrating its potential for accurate communication solutions.

This paper presents a vision-based sign language recognition model that uses a Hybrid CNN-LSTM architecture with an attention mechanism. The model combines the strengths of Convolutional Neural Networks (CNNs) for feature extraction, Long Short-

Term Memory (LSTM) networks for sequential learning, and an attention mechanism to focus on key image regions. It is trained on a custom dataset of American Sign Language (ASL) gestures to achieve accurate recognition of signs. The aim is to enhance communication between the hearing- and speech-impaired community and the public.

To evaluate the model's effectiveness, it is benchmarked against SqueezeNet[4], a lightweight CNN architecture renowned for its efficiency in image classification tasks. Specifically, this study focuses on detecting the ASL hand signs for the numbers “1,” “2,” and “3.” The hybrid model’s performance and efficiency are compared to that of SqueezeNet, providing insights into the benefits of combining CNN, LSTM, and attention mechanisms for sign language recognition.

2. Methodology

The custom dataset contains images of static ASL gestures. These images are preprocessed through resizing, normalization, and augmentation techniques such as rotation and color jittering to ensure the model generalizes well across different conditions. The dataset is split into training, validation, and testing sets, and the performance is evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrices.

For training, the CNN component extracts the relevant features, and the LSTM unit captures any sequential dependencies in the features. The attention mechanism ensures the model focuses on the most informative parts of the input images. The output of the model is classified into predefined classes based on the extracted features, with the training process optimized using stochastic gradient descent and cross-entropy loss for multi-class classification.

The performance of the model is compared with that of benchmark models, specifically SqueezeNet, to evaluate its effectiveness in detecting static gestures. Performance metrics, including mAP, precision, recall, F1-score, and accuracy, are computed to assess the accuracy and efficiency of the model in gesture classification.

2.1. *Dataset preparation*

For this study, a custom dataset was created using Roboflow[5], focusing on images representing static sign language gestures. Roboflow was instrumental in data annotation, augmentation, and preprocessing.

The dataset consists of images representing different 3 sign language classes of the numbers 1,2 and 3 as shown in Fig.1, Fig.2 and Fig.3, respectively and was split into three subsets:

- **Training Set**: Comprising 69% of the total data (1,164 images), this subset was utilized for model training.
- **Validation Set**: Comprising 19% of the total data (324 images), this subset was reserved for hyperparameter tuning and monitoring model performance during training.
- **Test Set**: Comprising 11% of the total data (192 images), this subset was strictly used for the final evaluation of the trained models.

During preprocessing, images were auto oriented to ensure consistent alignment and resized to a uniform dimension of 640x640 pixels, aligning with the input requirements of the benchmarking models used in the later phases of the study. The dataset was exported in COCO format, a widely accepted standard for object detection tasks, facilitating compatibility with various deep learning frameworks.

Fig. 1. Hand signal of number 1 in ASL

Fig. 2. Hand signal of number 2 in ASL

Fig. 3. Hand signal of number 3 in ASL

2.2. *SqueezeNet*

SqueezeNet is a compact convolutional neural network specifically designed to balance high accuracy with minimal model size. It employs "fire modules," which involve an initial squeeze layer utilizing 1x1 convolutions to reduce the input channels, followed by an expand layer with a combination of 1x1 and 3x3 convolutions for enhanced feature extraction. This architectural approach significantly reduces the parameter count and computational demands while maintaining robust performance. Therefore, SqueezeNet serves as an optimal benchmark model for comparative analysis.

2.3. *Base Convolutional Neural Network (CNN)*

A custom CNN architecture was developed to serve as the backbone for feature extraction and classification, specifically designed to optimize feature capture, depth, and regularization for handling the hand sign dataset. The architecture is composed of two primary components: feature extraction layers and fully connected layers dedicated to classification. The feature extraction module comprises four convolutional blocks with increasing depth, ranging from 64 to 512 channels. Early layers employ 5x5 kernels to capture spatial details effectively. Each convolutional block incorporates batch normalization and ReLU activation functions for stability and non-linearity. Max pooling is applied to reduce spatial dimensions, which aids in focusing on prominent features. Flattened features from the convolutional blocks are passed through the fully connected layers, beginning with an input size of 512×4×4. Regularization is implemented using a dropout rate of 0.5 to mitigate overfitting. The final layer produces probabilities for the three hand signal classes.

2.4. *CNN with LSTM and Attention mechanism*

The hybrid CNN architecture combines the base CNN mentioned in the previous section with an LSTM module, and an attention mechanism

CNNs extracts spatial features from images, but these features may still have inherent relationships. LSTMs help model these dependencies, improving feature interpretation[6].

The attention mechanism focuses on the most relevant features by assigning weights to the outputs of the LSTM. By computing a weighted sum of the LSTM outputs, the attention mechanism ensures that the model prioritizes key information while suppressing less important features[7].

Finally, fully connected layers are used to perform the final classification, with dropout applied for regularization to reduce overfitting. This model is designed to efficiently handle static gesture recognition tasks.

2.5. *Training and validation*

Training: The training process begins by setting up a Stochastic Gradient Descent (SGD) optimizer and a cyclic learning rate scheduler to adjust the learning rate during training. The models are trained over 25 epochs, where the cross-entropy loss is computed, and the weights are updated through backpropagation. The training loop includes tracking of training accuracy, loss, and mean average precision (mAP) for each epoch. After every epoch, the model is evaluated on the validation set to check if the validation loss improves, saving the best model based on the lowest validation loss.

Testing: After training, the performance of the model is evaluated on the test set and various metrics were computed, including the confusion matrix, F1-score, ROC curve and overall accuracy.

2.6. *Performance Metrics:*

The effectiveness of the models were evaluated using standard metrics, mathematically defined as follows[8]:

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \quad (1)$$

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FP)}} \quad (2)$$

$$\text{F1} - \text{Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (3)$$

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (4)$$

As shown in Eq. (4), accuracy is the overall proportion of correctly classified instances. As in Eq. (1), precision indicates how much of the result are actual positives out of the predicted positives. As indicated in Eq. (2), recall calculates how much of the actual positives are labelled as true positives. F1-Score as mentioned in Eq. (3), is a function of precision and recall and will provide a value to balance between precision and recall.

These metrics provide a robust assessment of the performance of the models. performance, balancing its ability to correctly classify signs while minimizing false positives and false negatives.

The input data undergoes a series of transformations to enhance model robustness and generalization. These transformations include resizing the images to a consistent size, random horizontal flipping, random rotation, and random colour jittering, which adjusts brightness, contrast, saturation, and hue. These augmentations are applied to all the models, ensuring that the models can effectively handle variations in the input data, such as different orientations and lighting conditions, ultimately improving their performance and ability to generalize to unseen data.

3. Results and Discussion

This section presents a comparative analysis of the base CNN, the proposed hybrid CNN-LSTM with attention model, and SqueezeNet for ASL sign recognition of the numbers "1", "2" and "3". All experiments were conducted on a system equipped with an NVIDIA GeForce RTX 2070 GPU, an Intel Core i7-10750H CPU operating at 2.60 GHz, and 16 GB of RAM.

The performance of the three models is summarized in Table 1, using accuracy, precision, recall, and F1-score as evaluation metrics. Additionally, the relationship between training loss and epochs was used as the primary evaluation metric, providing insight into the convergence behavior of each model during training.

Table 1. Performance metrics of CNN, Hybrid CNN and SqueezeNet

Metric	SqueezeNet	Base CNN	Hybrid CNN
Accuracy	89%	89%	95%
Precision	90%	88.66%	95%
Recall	88.33%	88.33%	95%
F1-Score	88.33%	88%	95%
Final mAP	96.93%	96.25%	99.6%

From the results in Table 1, it is clear that the hybrid CNN-LSTM with attention model significantly outperforms the base CNN and SqueezeNet in all metrics. Its ability to integrate convolutional feature extraction with temporal pattern recognition, combined with attention mechanisms, allows it to focus on the most critical spatial regions, enhancing its precision and recall.

In contrast, the base CNN and SqueezeNet models perform similarly in terms of accuracy, but SqueezeNet exhibits slightly higher precision due to its efficient architecture, which minimizes redundancy in feature extraction. However, its recall and F1-score remain comparable to the base CNN, indicating a potential trade-off between efficient processing and model sensitivity to certain classes.

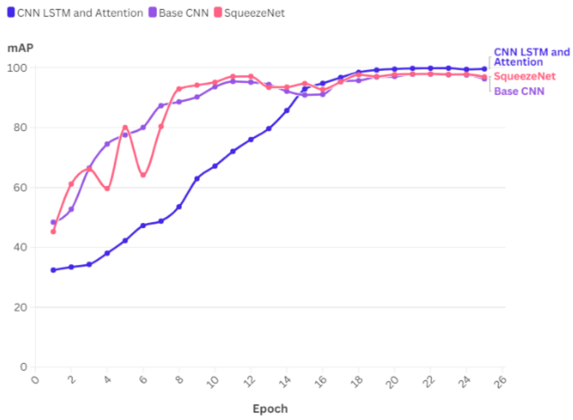


Fig. 4. Epochs vs mAP%

As shown in Fig.4, the mean Average Precision (mAP) scores across epochs of the SqeezeNet, Base CNN and the CNN with LSTM and attention mechanism models, further show the superiority of the hybrid CNN-LSTM with attention model. It consistently outperforms the base CNN and SqueezeNet, particularly in the later stages of training, achieving an mAP of 99.6% by the 25th epoch.

The ability of the hybrid model to achieve high mAP scores at later epochs highlights its strength in balancing precision and recall across varying confidence thresholds. By leveraging attention mechanisms, the model retains essential spatial details critical for distinguishing fine-grained features in the gestures.

In contrast, the base CNN achieves a respectable mAP but begins to level after 15 epochs, reflecting its limited capacity to generalize beyond certain levels of precision and recall. SqueezeNet achieves strong performance initially, peaking at epoch 10, but its mAP scores gradually decline thereafter, indicating a lack of robustness in maintaining balanced precision and recall across varying thresholds.

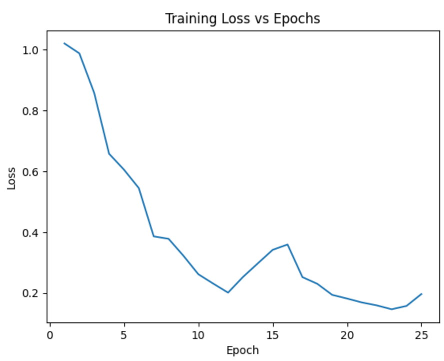


Fig. 5. Training loss vs Epochs Base CNN

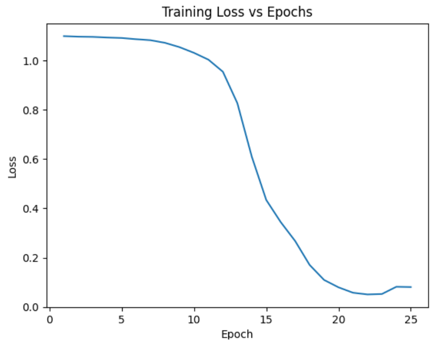


Fig. 6. Training loss vs Epochs Hybrid CNN

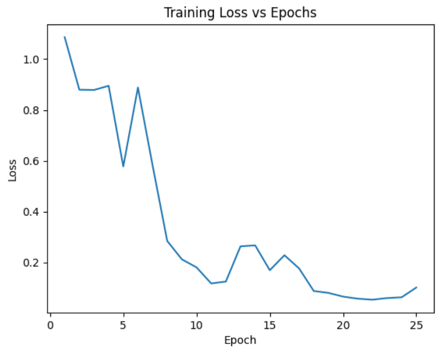


Fig. 7. Training loss vs Epochs SqueezeNet

Figures Fig. 5, Fig. 6 and Fig. 7 show that only the hybrid CNN model has a consistent steady decrease in loss, this can be attributed to its strong optimization capabilities and the combination of convolutional feature extraction, feature map sequence processing by the LSTM and attention mechanisms.

4. Conclusion

In this paper a hybrid deep learning model combining CNNs, LSTMs, and attention mechanisms for recognizing static hand gestures in American Sign Language (ASL) is proposed. The hybrid CNN model demonstrated strong performance across metrics such as accuracy and F1-score when compared to SqueezeNet. Data augmentation improved generalization, enabling adaptability to input variations.

However, limitations include the small size of the dataset and the lack of sequential data, which restricted the ability of the model to capture temporal dependencies and limited its robustness to complex or diverse gestures. These constraints highlight the need for larger, more comprehensive datasets and improvements in handling sequential data for better generalization and performance. Future work could focus on expanding the dataset, incorporating dynamic gesture recognition. Exploring multimodal approaches that integrate gesture recognition with other sensory data could further advance sign language detection systems.

References

1. D. F. Australia. "The Deaf Census: find out about Auslan users." https://www.deafnessforum.org.au/the-deaf-census-find-out-about-auslan-users/ (accessed 2025/01/21, 2025).
2. A. M. Buttar, U. Ahmad, A. H. Gumaei, A. Assiri, M. A. Akbar, and B. F. Alkhamees, "Deep Learning in Sign Language Recognition: A Hybrid Approach for the Recognition of Static and Dynamic Signs," *Mathematics*, vol. 11, no. 17, p. 3729, 2023. [Online]. Available: https://www.mdpi.com/2227-7390/11/17/3729.
3. A. R. Satish Kumar Alaria, Vivek Sharma, Vijay Kumar, "Simulation and Analysis of Hand Gesture Recognition for Indian Sign Language using CNN," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 10, no. 4, pp. 10-14, 2022, doi: https://doi.org/10.17762/ijritcc.v10i4.5556.
4. S. H. Forrest N. Iandola, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, Kurt Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size," presented at the ICLR, 2016. [Online]. Available: https://arxiv.org/pdf/1602.07360.
5. PoseEstimation. "Hand Signals Dataset." Roboflow. https://universe.roboflow.com/poseestimation-nzxk0/hand-signals-m7ruz (accessed 19/01/2025, 2025).
6. Y. He, C. Wen, and W. Xu, "Residual Life Prediction of SA-CNN-BILSTM Aero-Engine Based on a Multichannel Hybrid Network," *Applied Sciences*, vol. 15, no. 2, p. 966, 2025. [Online]. Available: https://www.mdpi.com/2076-3417/15/2/966.
7. Z. L. Jingyi Wang, Qiang Liu, Shu Wu, "Towards Accurate and Interpretable Sequential Prediction: A CNN & Attention-Based Feature Extractor," presented at the Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM '19), Beijing, China, 2019. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3357384.3357887.
8. T. Abdullah All, E. M. Mahir, S. Akhter, and M. R. Huq, "Detecting Fake News using Machine Learning and Deep Learning Algorithms," in *2019 7th International Conference on Smart Computing & Communications (ICSCC)*, 28-30 June 2019 2019, pp. 1-5, doi: 10.1109/ICSCC.2019.8843612. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8843612

Authors Introduction

Mr. Shakir Hussain Naushad Mohamed

He is currently an undergraduate student majoring in Bachelor of Electrical and Electronics Engineering (Honours) at Deakin University, Australia.

Mr. Hao Feng Chan

He is currently an undergraduate student majoring in Bachelor of Mechatronics Engineering (Honours) at Deakin University, Australia.

Mr. Dexter Sing Fong Leong

He is currectly an undergrate pursuing Bachelors Of Mechatronic Engineering (Honours) in Deakin University, Australia.

Mr. Chau Wui Chung Alton

He is currently an undergraduate student in Bachelor of Mechanical Engineering (Honours) at Deakin University, Australia.

Andi Prademon Yunus, Ph.D

He is an Assistant Professor at Telkom University and he received his PhD in Engineering from Mie Unviersity, Japan. His research focuses on applied and fundamental machine learning for motion and behavior computing. He also collaborates with industry partners to develop AI-based tools for language modeling and image analytics.

Dr. Takao Ito

He is Professor of Management of Technology (MOT) in Graduate School of Advanced Science and Engineering at Hiroshima University. His current research interests include automata theory, artificial intelligence, systems control, quantitative analysis of interfirm relationships using graph theory, and engineering approach of organizational structures using complex systems theory.

Mr. Zheng Cai

He is currently a PhD candidate at Deakin University's Institute for Intelligent Systems Research and Innovation (IISRI). His research interests include multiobjective optimisation algorithms such as metaheuristic algorithms and evolutionary algorithms for scheduling problems. He is also exploring the integration of machine learning with optimisation algorithms.

Ms Deng Xinjie

She is pursuing a PhD in Information Technology at the Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, with her research centered on creating lightweight deep learning algorithms for computer vision applications.

Dr. Yit Hong Choo

He has completed his PhD and is now a Research Fellow in Operations Analytics at Deakin University's Institute for Intelligent Systems Research and Innovation (IISRI), supported by the Rail Manufacturing Cooperative Research Centre (RMCRC). His research focuses on advanced multi-objective optimisation algorithms for complex maintenance scheduling in rolling stock. He collaborates with transportation industry partners to develop AI-based tools for video and image analytics.

Geographic and Risk Factor Analysis of Non-Communicable Cardiovascular Diseases in Central Java using Machine Learning

Nurhasanah[1], Andi Prademon Yunus[1*]
[1]*Telkom University, Banyumas, Indonesia*
Email: nurhasanahh@student.telkomuniversity.ac.id, andiay@telkomuniversity.ac.id
[*]*Corresponding Author*

Abstract

Non-communicable diseases (NCDs), especially cardiovascular disease, are a major health problem in Indonesia with a global mortality rate of 17.9 million per year. In Central Java, hypertension cases will reach 8.5 million patients by 2023. This study uses the CART (Classification and Regression Tree) method and geographic mapping with Python to identify cardiovascular disease risk factors. Results showed alcohol stores as the dominant risk factor (54.3%), followed by sweet drinks (25.7%) and smokers (17.1%). The mapping identified the distribution of alcohol stores in 19 regions as a major factor in Central Java, Indonesia.

Keywords: Non-Communicable Diseases, Cardiovascular Diseases, Classification and Regression Tree, Central Java, Risk Factors, Geographic Distribution

1. Introduction

Non-communicable diseases (NCDs) are a major health problem in Indonesia, with the proportion of deaths reaching 76%. These non-communicable diseases mainly occur in low- and middle-income countries [1]. It is estimated that Indonesia experienced a total potential loss of 4.47 trillion US dollars from 2012 to 2030 due to non-communicable diseases. The high prevalence of NCDs can lead to increased demand for health services, more expensive treatment, and increased health expenditure, which in turn can reduce the budget available for investment in more productive activities [2].

Among the main types of NCDs, cardiovascular disease (CVD) is the leading cause of death in Indonesia with a proportion reaching 38%, followed by cancer at 12%, diabetes at 7%, chronic respiratory disease at 6%, and other NCDs at 13%. Cardiovascular disease is the leading cause of death worldwide, with 17.9 million deaths per year [3]. Cardiovascular diseases include various disorders of the heart and blood vessels, such as coronary heart disease, cerebrovascular disease, hypertension, peripheral arterial disease, rheumatic heart disease, congenital heart disease, deep vein thrombosis, and pulmonary embolism [4].

The main factors that cause cardiovascular disease include the consumption of unhealthy foods, lack of physical activity or calorie-burning exercise, alcohol consumption, smoking, and stress levels. Behavioral risk factors, such as unhealthy diet and lack of physical activity, along with smoking and alcohol consumption, are of major concern in cardiovascular disease prevention efforts [5]. The more risk factors a person has, the higher the likelihood of developing cardiovascular disease [6]. Therefore, a thorough analysis of these risk factors is crucial to identify influences and correlations associated with cardiovascular disease. Some regions in Indonesia, such as Central Java Province, record a high prevalence of cardiovascular disease, suggesting the importance of a deeper understanding of the risk factors that contribute to the disease.

Central Java Province is among the top five regions with the highest cardiovascular prevalence (hypertension) [2]. In 2023, the estimated number of people with hypertension in Central Java aged over 15 years reached 8,554,672 people or about 38.2 percent of the population of that age. This figure has increased compared to the previous year, indicating that hypertension is a serious health problem in this region [7].

This makes hypertension a major health problem that needs special attention. Therefore, the selection of Central Java as the location of this study is very appropriate to understand more about the geographical characteristics and risk factors that affect cardiovascular disease by utilizing machine learning or machine learning using the CART (Classification And Regression Tree) algorithm.

CART is used for decision-making related to classification and regression. In classification, CART produces a decision tree that maps observations into classes or categories. In regression, CART produces a decision tree that predicts the numerical or continuous value of the target variable [8]. This research will focus on regression or regression tree, the resulting regression tree is used to predict the numerical or continuous value of the target variable based on the relationship with the independent variables.

Therefore, this study aims to determine the CART method in identifying risk factors with the greatest influence on cardiovascular disease in Central Java. Risk factors with higher feature importance indicate that they have a greater contribution in causing cardiovascular disease. In addition, this research will also be accompanied by visualization of mapping the results of the analysis geographically, so as to know the geographical distribution of cardiovascular disease risk factors. This mapping will be created using the Python programming language and utilizing libraries such as GeoPandas to read and process geospatial data. The results of this research are expected to be useful not only for health practitioners and policymakers in Central Java but also for similar research in other regions facing similar health challenges.

2. Methodology

This section describes the research methodology, including the data collection scheme, data preprocessing, and CART model design.

2.1. *Data Collection*

The cardiovascular disease dataset was obtained from the Central Java Health Office, which includes data from 2017 to 2023. However, in this study, only the 2023 data was used to ensure the relevance and accuracy of the analysis.

Meanwhile, risk factor data was collected through a scraping process from Google Maps using a Chrome extension called Instant Data Scraper. The scraping process was done manually for each region in Central Java, which consists of 29 districts and 6 cities.

Once the risk factor data and the number of cardiovascular disease patients have been combined, the dataset is now ready to be used for further analysis. However, before starting the analysis, another important step is to perform data preprocessing to ensure the data is clean, consistent, and ready to be optimally processed.

2.2. *Modeling*

In the modeling stage, the Decision Tree Regressor is used to predict the target variable, with logarithmic transformation on the target data to reduce the influence of outliers and ensure a more normal distribution. To improve model performance, hyperparameter tuning was performed using GridSearchCV with grid parameters including max_depth, min_samples_split, min_samples_leaf, and max_features. The tuning results showed the best parameters with max_depth 6, max_features 'sqrt', min_samples_leaf 2, and min_samples_split 2, which resulted in a model with lower prediction error based on neg_mean_squared_error evaluation. This combination of parameters allows the model to capture the data patterns better.

2.3. *Evaluation Model*

In this study, the model evaluation used is RMSE (Root Mean Squared Error) and R^2 (R-squared). RMSE measures the average error of the model prediction against the true value, the smaller the RMSE value or closer to 0, the better the model in prediction [9]. Meanwhile, R^2 measures how well the variation in data can be explained by the model, with values ranging from 0 to 1. The higher the R^2 value, the better the model is at explaining variations in the data [10]. The mathematical formulas of RMSE and R^2 are given in Eq. (1) and Eq. (2), respectively

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(y_i - \hat{y}_i)^2} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (2)$$

2.4. *Feature Importance*

Feature Importance measures the contribution of each feature in predicting the target by calculating how much the feature reduces uncertainty in the decision tree. The value ranges from 0 to 1, with higher values indicating greater influence. Feature importance can help identify the risk factors that have the most influence on cardiovascular disease prevalence.

3. Results and Discussions

3.1. *Model Performance Evaluation*

The model evaluation results show an RMSE of 0.33, which indicates a very small prediction error and good predictive ability of the model on the available data. An RMSE close to 0 indicates better model performance, although these results are more relevant for data exploration than evaluation of new data. In addition, the R^2 value of 0.91 indicates that the model is able to explain 91% of the variation in the target data, indicating that the model is effective in identifying data patterns and providing accurate predictions, with only 9% of the variance unexplained.

3.2. *Feature Importance Result*

The feature importance results obtained from the analysis are presented in Table 1 and in the bar chart visualization below.

Table 1. Feature Importance

No.	Feature	Importance
1.	alcohol_store	0.2728
2.	sweet_drinks	0.2231
3.	smokers	0.1970
4.	transport	0.1269
5.	gym	0.0922
6.	fast_food	0.0631
7.	park	0.0217
8.	sports_center	0.0031
9.	tourist_spots	0.0000

The bar chart visualization in Fig. 1 illustrates the contribution of each feature in the model. In this graph, features are displayed on the Y-axis, and important values are shown on the X-axis.

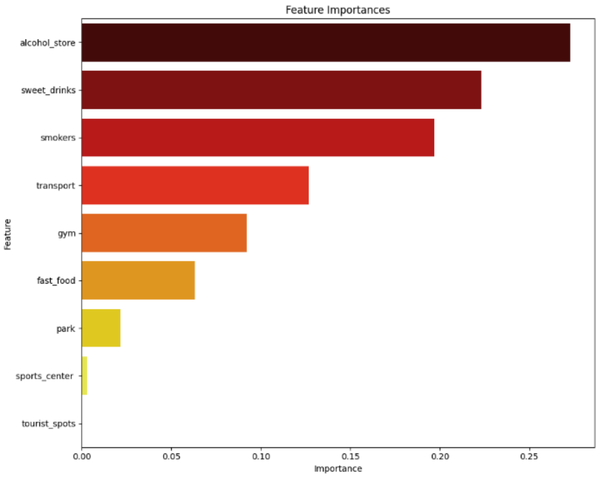


Fig.1. Feature Importance of Risk Factors on CART Model

1. Alcohol stores have the largest contribution (0.2728), indicating that the presence of alcohol stores plays a significant role in the increased risk of cardiovascular disease.
2. Sweet drinks (0.2231) and smokers (0.1970) also have large contributions to the prediction, with consumption of sweet drinks and smoking increasing the risk of cardiovascular disease.
3. Features with small contributions such as sports_center (0.0031) and tourist_spots (0.0000) indicate that sports facilities and tourist attractions have little effect on cardiovascular disease risk.

3.3. *Geographic Distribution*

The mapping of dominant risk factors in Central Java was conducted using matplotlib and geopandas, with Set3 colormaps to distinguish risk factor categories.

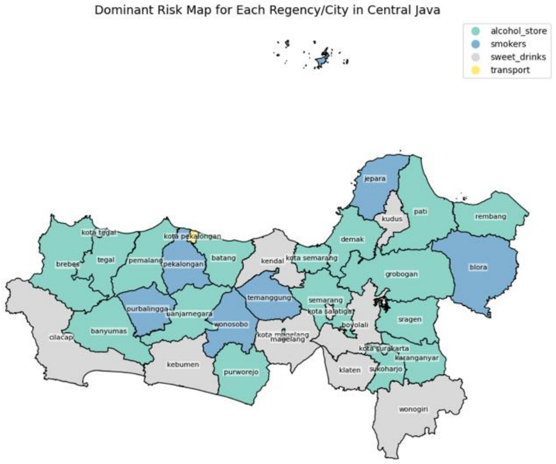


Fig.2. Mapping Results of Dominant Risk Factors in Central Java

The visualization results, as shown in Fig. 2, reveal that the alcohol store factor is the most dominant in 54.3% of the regions, followed by sweet_drinks (25.7%), smokers (17.1%), and transport (2.9%). These findings suggest that alcohol consumption is a major risk factor that needs more attention in cardiovascular disease prevention efforts in Central Java.

4. Conclusion

This study successfully identified and mapped the main risk factors for cardiovascular disease in Central Java using the CART method. Of the 9 factors analyzed, 4 factors were found to have a significant effect: alcohol_store (54.3%), sweet_drinks (25.7%), smokers (17.1%), and transport (2.9%), with alcohol_store as the dominant factor. The mapping shows that alcohol_store is distributed in 19 regions, sweet_drinks in 9 regions, smokers in 6 regions, and transportation in 1 region.

References

1. NCD Data Portal. “Noncommunicable Diseases Data Portal” [online]. Available : https://ncdportal.org/
2. Bloom, David & Chen, Simiao & McGovern, Mark & Prettner, Klaus & Candeias, Vanessa & Bernaert, Arnaud & Cristin, Stéphanie. Economics of Non-Communicable Diseases in Indonesia.
3. World Health Organization: WHO, “Noncommunicable diseases,” Dec. 23, 2024. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases
4. Hassan, C. A. U., Iqbal, J., Irfan, R., Hussain, S., Algarni, A. D., Bukhari, S. S. H., Alturki, N., & Ullah, S. S. (2022). Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers. Sensors (Basel, Switzerland), 22(19), 7227. https://doi.org/10.3390/s22197227
5. World Health Organization: WHO, “Cardiovascular diseases (CVDs),” Jun. 11, 2021. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
6. H. E. Bays et al., “Ten things to know about ten cardiovascular disease risk factors,” American Journal of Preventive Cardiology, vol. 5, p. 100149, Jan. 2021, doi: 10.1016/j.ajpc.2021.100149.
7. D. N. Khasanah, “The Risk Factors Of Hypertension In Indonesia (Data Study Of Indonesian Family Life Survey 5),” Journal of Public Health Research and Community Health Development, vol. 5, no. 2, p. 80, Feb. 2022, doi: 10.20473/jphrecode.v5i2.27923.
8. W. Dong et al., “Risk factors and geographic disparities in premature cardiovascular mortality in US counties: a machine learning approach,” Scientific Reports, vol. 13, no. 1, Feb. 2023, doi: 10.1038/s41598-023-30188-9.
9. W. A. C. Castañeda and P. B. Filho, “Improvement of an Edge-IoT architecture driven by artificial intelligence for Smart-Health chronic disease management,” Sensors, vol. 24, no. 24, p. 7965, Dec. 2024, doi: 10.3390/s24247965.
10. S. Panda, B. Purkayastha, D. Das, M. Chakraborty, and S. K. Biswas, “Health insurance cost prediction using regression models,” 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), pp. 168–173, May 2022, doi: 10.1109/com-it-con54601.2022.9850653.

Authors Introduction

Ms. Nurhasanah

She is an undergraduate student of Informatics Engineering at Telkom University, Indonesia with a GPA of 3.96. She has expertise in data analysis, human resource development, social media specialist, and copywriting.

Dr. Andi Prademon Yunus

He is an Assistant Professor at Telkom University, and he received his PhD in Engineering from Mie University, Japan. His research focuses on applied and fundamental machine learning for motion and behavior computing. He also collaborates with industry partners to develop AI-based tools for language modeling and image analytics.

Analysis of Geographical Characteristics and Risk Factors for Non-Communicable Diseases: Diabetes in Central Java Using Random Forest and SHAP

Ambar Arum Prameswari[1], Andi Prademon Yunus[1*]
[1]*Telkom University, Banyumas, Indonesia*
Email: ambararum@student.telkomuniversity.ac.id, andiay@telkomuniversity.ac.id
[*]*Corresponding Author*

Abstract

Diabetes mellitus is a growing health issue in Indonesia, particularly in Central Java Province. In 2023, it contributed significantly to public health concerns due to its increasing prevalence. This study utilizes the Random Forest algorithm to identify dominant diabetes risk factors and map their geographic distribution across the region. The analysis incorporates community lifestyle indicators such as the number of sweet drink stores, gyms, sports halls, transportation availability, karaoke venues, tourist attractions, fast food outlets, and alcohol stores in each regency/city. Results indicate that the number of tourist attractions (feature importance: 0.2817) and sweet drink stores (feature importance: 0.1502) are the primary global risk factors. Locally, tourism dominates as the key risk factor in 18 regencies/cities (51.4% of the region). The model employs optimal hyperparameter tuning to improve accuracy and utilizes SHAP techniques to evaluate the importance of local features. This study aims to enhance understanding of the key risk factors contributing to diabetes cases and their distribution in Central Java Province.

Keywords: diabetes, risk factors, Random Forest, SHAP, Central Java, non-communicable diseases

1. Introduction

Non-communicable diseases (NCDs) or chronic diseases are among the most significant global health challenges. NCDs encompass various conditions that are not transmitted between individuals but are caused by genetic, physiological, lifestyle, and environmental factors [1]. Major NCDs such as cardiovascular diseases, cancer, chronic respiratory diseases, and diabetes mellitus account for a high number of deaths globally, with an annual toll of 74% of total global deaths or approximately 41 million lives [2].

In Indonesia, NCDs were responsible for 76% of total deaths in 2019, with diabetes mellitus contributing to 7% of those fatalities [3]. Diabetes is a serious chronic condition characterized by elevated blood glucose levels due to the body's inability to produce sufficient insulin or use insulin effectively. Without proper management, diabetes can lead to severe complications such as cardiovascular diseases, kidney failure, and damage to vital organs [4].

The global prevalence of diabetes continues to rise. In 2021, the International Diabetes Federation (IDF) reported approximately 537 million people living with diabetes worldwide, a figure projected to grow to 643 million by 2030 and 783 million by 2045 [5]. Indonesia ranks among the top countries with the highest number of diabetes cases, ranking fifth globally in 2021 with 19.5 million adults affected. This number is expected to increase to 28.6 million by 2045 [5].

The prevalence of diabetes in Central Java also shows an upward trend. In 2021, there were 618,546 diabetes cases in the province, rising to 624,082 cases in 2023 [6] [7]. Although most regencies and cities in Central Java have achieved standardized healthcare services for diabetes patients, this growing trend underscores the need for improved preventive and control measures.

Diabetes mellitus is a chronic disease influenced by numerous risk factors, including unhealthy lifestyle habits such as smoking, physical inactivity, unhealthy diets, and excessive alcohol consumption [8]. These factors lead to physiological changes such as high blood pressure, obesity, elevated blood glucose levels, and high cholesterol, all of which contribute to the onset of diabetes [8].

Psychological conditions, including stress, anxiety, and depression, are also closely associated with NCDs, including diabetes [9]. Given the complexity of these risk factors, identifying dominant risk factors at the local level is essential to support more effective diabetes prevention and control efforts.

This study aims to use data-driven approaches to Identify the key risk factors contributing to diabetes prevalence in Central Java at the global (province-wide) level and analyze the dominant risk factors in each regency/city and city in Central Java.

This study employs the Random Forest algorithm as the primary method. Random Forest is a supervised learning algorithm capable of handling complex data, reducing overfitting, and efficiently determining the importance of each feature [10]. This algorithm is used to calculate feature importance globally, covering the entire Central Java region.

Additionally, this study applies SHAP (SHapley Additive exPlanations) to analyze feature importance locally for each regency/city. SHAP enables a detailed analysis of how each feature influences individual predictions, providing more specific insights[11]. The integration of SHAP with geographical mapping offers a valuable opportunity to visualize the distribution of dominant risk factors across Central Java. This approach provides critical insights into identifying the specific risk factors prevalent in each regency or city, enabling targeted prevention strategies. By focusing on the dominant risk factors within each region, this method ensures that preventive efforts are more precise, effective, and aligned with the unique needs of the local population.

2. Methodology

2.1. *Global Feature Importance*

In this study, we utilize the Random Forest algorithm to identify the dominant risk factors contributing to the prevalence of diabetes in Central Java. Random Forest, first introduced by Leo Breiman [12], is an ensemble learning method based on decision trees. Each tree in the forest relies on a subset of randomly selected variables, and the final prediction is obtained by combining the results from multiple trees [13]. This approach is known for its high accuracy, ability to handle missing data, and capacity to identify important variables [14].

Random Forest works by constructing multiple decision trees, where each tree is trained using a random subset of data and features. The decision tree algorithm makes predictions by recursively splitting the dataset based on feature values. In Random Forest, the choice of features at each node is randomized, meaning only a subset of the features is considered when determining the optimal split. This helps reduce overfitting and increases the robustness of the model [13].

One of the key advantages of Random Forest, as mentioned earlier, is its ability to estimate feature importance, which indicates how much each feature contributes to the model's prediction. In this case, we focus on identifying global feature importance, which measures the overall contribution of each feature to the prediction of diabetes prevalence across Central Java.

To build decision trees, Random Forest employs the Mean Squared Error (MSE) as a measure of impurity when splitting nodes. The goal is to find the optimal split at each node, minimizing the MSE. The formula for MSE is as follows in Eq.(1).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \qquad (1)$$

N is the number of data points in the node, y_i is the actual value of the i-th data point, and $\hat{y}_i$ is the predicted value for the i-th data point[15].

Each time a feature is used to split the data at a node, the impurity (MSE for regression) is reduced. This reduction in impurity is accumulated for each feature across all nodes and decision trees in the Random Forest. The final feature importance is determined by averaging the accumulated impurity reduction for each feature across all trees in the forest. Features that result in the greatest reduction in impurity are considered the most important for predicting the target variable (in this case, diabetes prevalence in Central Java).

2.2. *Local Feature Importance*

In this study, we utilize SHAP (SHapley Additive exPlanations) to identify and explain the contribution of each feature towards the predictions made by the Random Forest model. SHAP is a technique used to provide a transparent explanation of machine learning model predictions by calculating the Shapley value for each feature, which represents its contribution to the model's prediction for each individual instance (data point). This local explanation provides insights into the specific impact of each feature on individual predictions, offering a deeper understanding of the model's decision-making process for each sample in the dataset [11]. The formula for the Shapley value for a feature ***j*** as follows in Eq.(2).

$$Shapley(X_j) = \sum\nolimits_{S \subseteq N\backslash\{j\}} \frac{k!\,(p-k-1)!}{p!}\left(f(S \cup \{j\}) - f(s)\right) \qquad (2)$$

Where p represents the total number of features, $N\backslash\{j\}$ denotes the set of all possible feature combinations excluding feature X_j, S is any subset of features within $N\backslash\{j\}$, $f(S)$ is the model's prediction using only the features in S, and $f(S \cup \{j\})$ refers to the model's prediction when the feature X_j is added to the feature set S [16].

Once the global feature importance has been determined, the next step is to calculate the local feature importance for each regency/city in Central Java using SHAP. This process is as follows:

1. The SHAP explainer is initialized using the trained Random Forest model and the selected feature dataset.
2. The explainer computes the SHAP values, which represent the contribution of each feature to the model's prediction for every individual sample in the dataset (each regency/city).
3. For each sample (regency/city), the SHAP value is computed for every feature, and the features with the largest SHAP values are identified as the dominant risk factors for that specific prediction.
4. After determining the dominant risk factors for each regency/city, the frequency of each feature being the dominant risk factor across all regencies/cities is calculated.

2.3. *Dataset and Variables*

The dataset used in this study includes lifestyle factors of the population, divided into two types: primary data and secondary data. The primary data consists of the number of risk factors present in each regency and city in Central Java, including alcohol stores, sweet drink outlets (such as boba and iced tea), fast food outlets, tourist attractions, fitness facilities (such as sports halls and gyms), public transportation facilities (such as terminals and bus stops), and entertainment venues (such as karaoke). This data was obtained through a scraping process on Google Maps using a tool called Instant Data Scraper Extension. The scraping process was conducted in September 2024. The secondary data used in this study includes the number of diabetes patients in Central Java as of December 2023, which was sourced from the Health Department of Central Java. By incorporating these lifestyle factors, the study offers a more localized and contextual approach to understanding the contribution of lifestyle factors to diabetes prevalence in the region.

2.4. *Model Evaluation and Geo-mapping*

To evaluate the performance of the model, two evaluation metrics were used: R-squared (R^2) and Root Mean Square Error (RMSE). RMSE measures the deviation of the model's predictions from the actual values. A lower RMSE value indicates better model performance. On the other hand, R^2 assesses how well the model explains the variance in the target data. The R^2 value ranges from 0 to 1, where a value closer to 1 indicates better explanatory power. R^2 can also be negative if the model performs poorly. The formulas used for R^2 and RMSE are provided below in Eqs. (3) and (4).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (4)$$

Where n is the number of observations, y_i is the actual value for observation $\hat{y}_i$ is the predicted value for observation $\bar{y}_i$ is the mean of the actual values [17].

For geo-mapping, the goal is to provide a clearer visual representation of the contribution of each feature and its distribution across regencies and cities in Central Java. This process began by utilizing spatial data of Central Java from GeoJSON files. The data was filtered to focus only on Central Java, removing irrelevant entities such as "Waduk Kedungombo" and "WaterBody." The spatial data was then merged with SHAP data, which includes the dominant risk factors for each regency/city.

Before merging, the names of the regencies/cities in both datasets (primary and secondary dataset) were standardized to ensure correct integration. Once the data was successfully merged, a thematic map was created to display the distribution of dominant risk factors across regencies and cities. A unique color was assigned to each category of risk factors using a color map. The data was plotted on the map, with colors representing the dominant risk factor for each regency/city. The names of the regencies/cities were added as labels for easier identification. Finally, an analysis of the thematic map provided additional insights into the geographic distribution of dominant risk factors, which can inform targeted intervention strategies.

3. Results and Discussions

3.1. *Simulations settings*

This section outlines the procedure for constructing a Random Forest model to predict diabetes cases. Initially, the dataset is loaded and preprocessed, which includes handling outliers and applying logarithmic transformations to the target variable in order to mitigate the influence of extreme values. Feature selection is then performed using SelectKBest, retaining only the top k features with the most significant linear correlation to the target variable.

Subsequently, hyperparameter tuning is conducted via GridSearchCV to determine the optimal number of estimators (n_estimators). During this process, 5-fold cross-validation is employed, with R^2 as the evaluation metric. Once the best parameter combination is identified, the model is trained using the full dataset and the optimal number of decision trees.

In the Random Forest algorithm, each decision tree is constructed using bootstrap sampling, automatically handled by the algorithm. Node splitting is based on Mean Squared Error (MSE) for regression tasks, with the best split being the one that yields the highest variance reduction. Upon completion of all decision trees, their individual predictions are aggregated to generate the final model prediction. The model's performance is evaluated using RMSE and R^2, and feature importance is calculated to identify the most influential features in predicting diabetes.

The entire process was executed in Google Collaboratory, a cloud-based platform, using Python as the programming language. The system configuration included 12 GB of RAM, provided by the free-tier account, and an NVIDIA Tesla T4 GPU, which facilitated the acceleration of computations, ensuring efficient data processing and model training.

3.2. *Performance and Analysis*

In this research, we evaluated two main metrics which are R^2 and RMSE. We use K-fold cross validation with average RMSE as the main metric to measure the Random Forest model performance. This evaluation aims to give us an understanding of the hyperparameter tuning to the model performance. We tune the hyperparameter of n_estimator to get the best optimal combination on the model with GridSearchCV. Table 1 shows the evaluation result with various n_estimator with value range from 1 to 201 with the K-fold determined to 3, 5, and 7.

Table 1. Hyperparameter tuning and its impact on the performance of the Random Forest model.

n_estimators range	Best n_estimators	R^2	RMSE	CV RMSE (Avg)
CV (folds)=3				
1-51	50	0.85	0.22	0.62±0.11
51-101	85	0.86	0.21	0.57±0.19
101-151	126	0.86	0.21	0.63±0.09
151-201	156	0.86	0.21	0.63±0.10
CV (folds)=5				
1-51	50	0.85	0.22	0.58±0.09
51-101	85	0.86	0.21	0.57±0.10
101-151	**126**	**0.86**	**0.21**	**0.56±0.10**
151-201	156	0.86	0.21	0.57±0.10
CV (folds)=7				
1-51	50	0.85	0.22	0.60±0.14
51-101	85	0.86	0.21	0.60±0.14
101-151	126	0.86	0.21	0.59±0.14
151-201	156	0.86	0.21	0.59±0.14

The results in Table 1, as shown in the highlighted row, utilized an n_estimators value of 126, resulting in an RMSE of 0.21, an R^2 score of 0.86, and the lowest CV RMSE (Avg) of 0.56 with a standard deviation of 0.10, which delivered the best evaluation results.

The best-performing model was subsequently selected for use in the global feature importance analysis in Central Java. Below are the results of the global feature importance obtained using the Random Forest model.

Table 2. Global feature importance value.

Feature	Importance Value
tourism	0.2817
sweet_drinks	0.1502
gym	0.1378
fast_food	0.1368
sport_hall	0.0833
karaoke	0.0831
transportation	0.0730
alcohol_stores	0.0541

Table 2. highlights the global feature importance for diabetes incidence in Central Java. The "tourist_attraction" feature has the highest importance value (0.2817), followed by "sweet_beverages" (0.1502), and "gym" and "fast_food," which have nearly equivalent values (0.1378 and 0.1368). While the other features demonstrate lower contributions.

Following the global feature importance analysis, local feature importance was assessed using SHAP (SHapley Additive Explanations) to identify the dominant risk factors for each regency and city in Central Java. These factors were then mapped to visualize their distribution. The feature importance values were further illustrated on a map of regencies and cities in Central Java using GeoJSON. This map highlights the prevalence of diabetes diseases cases across the regions, with color intensity representing the distribution of significant risk factors.

The visualization shown in Fig.1 results reveal that the "tourism" feature emerges as the most dominant risk factor, identified in 18 regencies and cities across Central Java, making it the most frequently observed factor both locally and globally. In the global feature importance analysis using Random Forest, it holds the highest importance value of 0.2817, underscoring its significant contribution to diabetes incidence. Locally, its dominance in regions such as Batang, Blora, Boyolali, Brebes, Demak, Kendal, Klaten, Kudus, Magelang, Pati, Pekalongan, Purworejo, Rembang, Sragen, Sukoharjo, Temanggung, and Semarang city aligns with its global significance. The dual role of tourism is evident: while tourist activities can promote physical movement (e.g., walking), they also foster unhealthy dietary habits, with increased availability of sugary drinks, snacks, and fast food around tourist attractions. This highlights the intricate relationship between tourism and public health, making it a critical factor in diabetes prevention strategies.

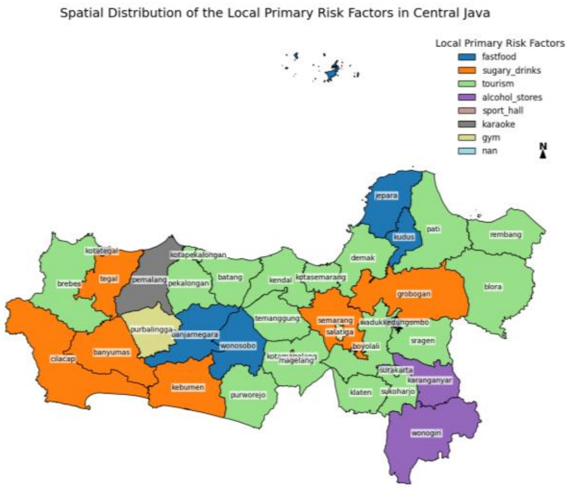


Fig.1 Spatial distribution of local primary risk factors in central java.

The "sweet_drinks" feature, although second in global importance (value: 0.1502), is dominant in only 6 regions, including Banyumas, Cilacap, Grobogan, Kebumen, Tegal, and Semarang. This suggests that while sweet drinks significantly impact diabetes incidence globally, their influence is geographically concentrated. Localized habits, such as a preference for sugary beverages in these regions, may amplify their impact. This discrepancy between global and local significance highlights the need for region-specific interventions to address varying consumption patterns and mitigate the influence of sweet drinks on diabetes.

The "gym" feature ranks third in global importance with a value of 0.1378 and is dominant in 2 regions: Purbalingga and Salatiga. While gyms symbolize access to fitness infrastructure, their limited accessibility to the broader population or unhealthy post-exercise dietary choices may mitigate their potential health benefits. Additionally, gyms tend to attract individuals who are already conscious of health, leaving larger populations unaffected. This explains why gym dominance is localized despite its relatively high global importance.

The "fast_food" feature ranks fourth in global importance with a value of 0.1368 but is dominant in just 4 regions, such as Banjarnegara, Jepara, Kudus, and Wonosobo. This uneven distribution reflects the variable prevalence of fast-food outlets, with higher concentrations in urban or semi-urban areas. Globally, fast food contributes significantly to diabetes due to its high-calorie and low-nutrient content. However, its limited local dominance suggests that other factors may play a larger role in regions with fewer fast-food establishments. This points to the importance of tailored strategies, such as promoting healthy eating habits alongside regulating fast-food expansion.

The "sports_hall" feature is identified as a dominant factor in only one region, contributing a mere 2.9% to the overall distribution. This limited dominance signifies its minimal impact on diabetes incidence at a global level, reflecting its relatively small influence on the overall variance.

In contrast, the "karaoke" feature, despite its lower global importance, is dominant in two regions—Pemalang and Tegal City. Karaoke establishments may indirectly increase diabetes risk through associated behaviors, such as stress relief activities that involve excessive alcohol consumption and smoking, particularly in venues equipped with bar facilities. These lifestyle factors may overshadow the potential stress-reducing benefits of karaoke, rendering it a locally significant yet globally minor contributor to diabetes prevalence.

Similarly, the "alcohol_store" feature holds a lower global importance value of 0.0541 but emerges as a dominant factor in two regions—Karanganyar and Wonogiri. While its contribution to diabetes is modest on a global scale, its local significance underscores the relationship between the prevalence of alcohol stores and increased alcohol consumption in these specific areas.

4. Conclusion

In this study, we employed a Random Forest (RF) model with 126 n_estimators to identify the key risk factors contributing to the prevalence of diabetes in Central Java. The feature importance analysis revealed that "tourism_spot" and "sweet_drinks" are the most significant risk factors, with "tourism_spot" having the greatest impact on diabetes incidence. The SHAP (Shapley Additive Explanations) analysis provided a deeper understanding of the local distribution of these risk factors, showing that "tourism_spot" dominates in 18 regions, while "sweet_drinks" is dominant in 6 regions. Model evaluation results indicated strong predictive accuracy, with an RMSE of 0.21, an R^2 score of 0.86, and the lowest CV RMSE (Avg) of 0.56 with a standard deviation of 0.10, delivering the best evaluation outcomes.

The visualization of dominant risk factors across regencies and cities in Central Java further highlights regional variations in the prevalence of diabetes risk factors. The map predominantly features green areas, indicating "tourism_spot" as the leading risk factor, influenced by both increased physical activity and unhealthy consumption patterns at tourist sites. Additionally, areas marked in orange, such as Banyumas, Cilacap, and Grobogan, suggest that "sweet_drinks" consumption is a dominant risk factor, underlining its significant contribution to diabetes risk. These findings emphasize the complex interplay of lifestyle and environmental factors in shaping diabetes risk across different regions.

Acknowledgments

This work was supported by Telkom University.

References

1. A. Budreviciute, S. Damiati, D. K. Sabir, K. Onder, P. Schuller-Goetzburg, G. Plakys, A. Katileviciute, S. Khoja, and R. Kodzius, "Management and Prevention Strategies for Non-communicable Diseases (NCDs) and Their Risk Factors," *Frontiers in Public Health*, vol. 8, 2020.
2. K. R. Thankappan and G. K. Mini, "Noncommunicable Diseases in the Elderly," in Handbook of Aging, Health and Public Policy: Perspectives from Asia, Singapore: Springer Nature Singapore, 2022, pp. 1-9.
3. World Health Organization, "Noncommunicable Diseases (NCD) Portal," *World Health Organization*. [Online]. Available: https://ncdportal.org/. [Accessed: May 6, 2024].
4. D. J. Magliano, E. J. Boyko, and IDF Diabetes Atlas 10th edition scientific committee, *IDF Diabetes Atlas*, 10th ed. Brussels: International Diabetes Federation, 2021, ch. 1, "What is diabetes?" [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK581938/. [Accessed: May 6, 2024].
5. International Diabetes Federation (IDF), "IDF Diabetes Atlas, 10th ed.," Brussels, Belgium, 2021. [Online]. Available: https://www.diabetesatlas.org. [Accessed: May 6, 2024].
6. Dinkes Jateng, "Profil Kesehatan Jateng 2021," *dinkesjatengprov.go.id*, 2021. [Online]. Available: https://dinkesjatengprov.go.id/v2018/dokumen/Profil_Kesehatan_2021/mobile/index.html. [Accessed: May 7, 2024].
7. Dinkes Jateng, "Profil Kesehatan Jateng 2023," *dinkesjatengprov.go.id*, 2020. [Online]. Available: https://dinkesjatengprov.go.id/v2018/dokumen/1Profil_Kesehatan_2023/mobile/index.html. [Accessed: Oct. 4, 2024].
8. World Health Organization, "Global Health Observatory (GHO) data: Noncommunicable diseases (NCD) risk factors," [Online]. Available: https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/ncd-risk-factors. [Accessed: May 7, 2024].
9. X. Liu, H. Cao, H. Zhu, H. Zhang, K. Niu, N. Tang, et al., "Association of chronic diseases with depression, anxiety and stress in Chinese general population: The CHCN-BTH cohort study," *Journal of Affective Disorders*, vol. 282, pp. 1278-1287, 2021.
10. Q. Zhou, W. Lan, Y. Zhou, and G. Mo, "Effectiveness Evaluation of Anti-bird Devices based on Random Forest Algorithm," 2020 7th International Conference on Information, Cybernetics, and Computational Social

Systems (ICCSS), Guangzhou, China, 2020, pp. 743-748, doi: 10.1109/ICCSS52145.2020.9336891.
11. T. T. H. Le, H. Kim, H. Kang, and H. Kim, "Classification and explanation for intrusion detection system based on ensemble trees and SHAP method," *Sensors*, vol. 22, no. 3, p. 1154, 2022.
12. L. Breiman, "The Random Forest," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
13. T. Zhu, "Analysis on the applicability of the random forest," in *Journal of Physics: Conference Series*, vol. 1607, no. 1, p. 012123, Aug. 2020, IOP Publishing.
14. F. R. Aszhari, Z. Rustam, F. Subroto, and A. S. Semendawai, "Classification of thalassemia data using random forest algorithm," in *Journal of Physics: Conference Series*, vol. 1490, no. 1, p. 012050, Mar. 2020, IOP Publishing.
15. M. T. Islam, M. Raihan, N. Aktar, M. S. Alam, R. R. Ema, and T. Islam, "Diabetes Mellitus Prediction using Different Ensemble Machine Learning Approaches," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, pp. 1-7, 2020.
16. Z. Li, "Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost," *Computers, Environment and Urban Systems*, vol. 96, p. 101845, 2022.
17. G. Grekousis, Z. Feng, I. Marakakis, Y. Lu, and R. Wang, "Ranking the importance of demographic, socioeconomic, and underlying health factors on US COVID-19 deaths: A geographical random forest approach," *Health & Place*, vol. 74, Art. no. 102744, 2022.

Authors Introduction

Ms. Ambar Arum Prameswari

She is currently pursuing a Bachelor of Informatics Engineering at Telkom University Purwokerto, Indonesia. Her research interests focus on machine learning and data analysis.

Dr. Andi Prademon Yunus

He is an Assistant Professor at Telkom University and he received his PhD in Engineering from Mie Unviersity, Japan. His research focuses on applied and fundamental machine learning for motion and behavior computing. He also collaborates with industry partners to develop AI-based tools for language modeling and image analytics.

Geographic Analysis of Risk Factors for Chronic Respiratory Non-Communicable Diseases using Machine Learning

Ayu Susilowati[1], Andi Prademon Yunus[1*]
[1]*Telkom University, Banyumas, Indonesia*
Email: ayususilowati@student.telkomuniversity.ac.id, andiay@telkomuniversity.ac.id
[*]*Corresponding Author*

Abstract

Chronic respiratory diseases (CRDs) are a significant category of non-communicable diseases (NCDs), affecting 235 million asthma patients and 64 million COPD patients globally. In Central Java, CRDs accounted for 6% of total deaths in 2023, with WHO projecting COPD as the third leading cause of death by 2030. This study employs a decision tree machine learning approach to analyze lifestyle behaviors and environmental factors, aiming to identify key CRD risk factors and map their geographic distribution. Results indicate public transportation contributes most significantly (0.3410), followed by smoking habits and NO_2 concentration. The model achieved an RMSE of 0.40 and R^2 of 0.83, reflecting high predictive accuracy. This approach provides insights and enhances healthcare access in high-risk areas.

Keywords: Non-communicable diseases, Chronic respiratory diseases, Decision tree, Geo-mapping

1. Introduction

Non-communicable diseases (NCDs) have emerged as a significant global health challenge, with far-reaching impacts on individuals, families, and societies. Among various NCDs, chronic respiratory diseases (CRDs), such as asthma and chronic obstructive pulmonary disease (COPD), hold a prominent position due to their high prevalence and morbidity rates. According to the World Health Organization (WHO), asthma affects approximately 235 million people worldwide, while COPD impacts over 64 million individuals [1] [2]. These diseases not only pose serious health risks but also impose substantial economic burdens [3].

In Indonesia, particularly in the province of Central Java, CRDs are a major cause of mortality, accounting for 6% of total deaths in 2023 [4]. WHO projects that by 2030, COPD will become the third leading cause of death globally, underscoring the urgency of addressing this issue through data-driven approaches [5]. Referring to global policies, controlling risk factors is a crucial aspect of preventing non-communicable diseases (NCDs) [6]. It is essential to conduct research to identify the risk factors for chronic respiratory diseases to prevent the rising prevalence of NCDs associated with respiratory conditions. One preventive approach involves the application of machine learning techniques.

Machine learning (ML) is becoming an increasingly important tool in health big data analysis due to its ability to identify complex patterns that are difficult to find through conventional statistical methods [7][8]. In the context of chronic respiratory disease risk factors, this study examines the distribution of public facilities—such as transportation, sports, and recreation centers, and access to healthy food—in Central Java. Using geographic analysis within a decision-tree framework, this study aims to elucidate the interactions between risk factors associated with chronic respiratory diseases in Central Java.

By analyzing feature importance, we can identify which predictors, such as air pollution levels, smoking prevalence, and access to health-related facilities, have the most significant impact on the occurrence of respiratory diseases. This approach not only aids in building predictive models with high accuracy but also provides actionable insights for policymakers and public health officials to target interventions effectively [9]. In this study, feature importance derived from a Decision Tree model is utilized to quantify the contribution of each risk factor, allowing for the identification of dominant factors at the district and city levels [10] [11].

Thus, geo-mapping distribution analysis with decision-tree methods offers a strategic approach to uncovering risk factor patterns for chronic respiratory diseases in Central Java, providing an evidence-based foundation for more effective prevention and control measures targeting these conditions.

2. Methodology

2.1. *Feature Importance Using Decision Tree*

In this subsection, we employ a machine learning approach, specifically the decision tree algorithm, to identify and analyze key risk factors contributing to the prevalence of chronic respiratory diseases (CRDs) in Central Java. The decision tree is a supervised learning model that is widely used in classification and regression tasks. It works by splitting the dataset into subsets based on feature values, creating a tree-like structure that aids in understanding the relationships between variables and the target outcome.

The core strength of the decision tree lies in its ability to compute *feature importance*, which quantitatively measures the contribution of each feature in improving the predictive accuracy of the model [12]. To calculate *feature importance*, the decision tree uses the *gini impurity* metric to evaluate the quality of splits at each node. The *gini impurity* is a measure of the likelihood that a randomly chosen data point would be incorrectly classified if it were randomly assigned a label according to the distribution of labels in the dataset. It is calculated as follows in Eq. (1).

$$Gini = 1 - \sum_{i=1}^{n} p_i^2 \quad (1)$$

Where $\boldsymbol{p_i}$ is the proportion of samples belonging to class $\boldsymbol{i}$, and $\boldsymbol{n}$ is the total number of classes. At each split, the decision tree computes the *gini impurity* for the parent node ($\boldsymbol{Gini}_{\textbf{parent}}$) and the resulting child nodes ($\boldsymbol{Gini}_{\textbf{left}}$) and ($\boldsymbol{Gini}_{\textbf{right}}$). The reduction in *Gini impurity*, referred to as the *Gini Gain*, determines the quality of the split. The formula for *Gini Gain* is as follows in Eq. (2).

$$Gini\ Gain = Gini_{\text{parent}} - Gini_{\text{child}} \quad (2)$$

By systematically evaluating the *Gini gain* at each split, the decision tree identifies the feature that most effectively reduces impurity, thereby improving the model's accuracy. This process continues iteratively, building a tree structure that captures the relationships between features and the target variable while ranking feature importance based on their contribution to impurity reduction.

2.2. *Dataset and Variables*

The dataset used in this study consists of 35 observations representing data from various districts and cities in Central Java. It includes variables related to chronic respiratory diseases (CRD) and potential risk factors. Key variables are: CRD_case (number of CRD cases), cigarette (average cigarette consumption per capita), NO_2 and SO_2 (air pollutant concentrations in µg/m³), and various facilities such as transport (public transportation), gym, fast_food outlets, sweet_drink vendors, tourism_spot, alcohol_store, and healthy_food outlets.

The facility data were collected through web scraping of mapping platforms to capture their distribution across districts and cities, providing a comprehensive geographic overview. This dataset integrates environmental and lifestyle factors to analyze their influence on CRD prevalence, enabling the identification of localized risk factors and supporting data-driven policy recommendations for disease prevention and control.

2.3. *Model Evaluation and Geo-mapping*

To evaluate the decision tree model, *Root Mean Square Error (RMSE) and R-squared (R²)* were calculated to measure the model's accuracy. RMSE evaluates the average error between predicted and observed values, while R^2 explains the proportion of variance in the dependent variable that is predictable from the independent variables. The formulas used are as below in Eq. (3) and Eq. (4).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \quad (4)$$

In the Eq. (3) for model evaluation, y_i represents the actual value or the observed data, $\widehat{y_i}$ denotes the predicted value generated by the model, and nnn refers to the total number of observations used in the calculation. These components are crucial for assessing the accuracy of the model by comparing the predicted outcomes with the actual results across all data points. Meanwhile, Eq. (4) evaluation context, $\overline{y_i}$ represents the mean of the actual values. This serves as a baseline to measure the model's performance by comparing the predicted values with the average of the observed data. It is particularly useful in calculating R^2 metrics like the coefficient of determination to evaluate how well the model explains the variance in the actual data.

To visualize the spatial distribution of chronic respiratory disease (CRD) prevalence and its significant risk factors, a geo-mapping approach was employed using GeoJSON data. This process began with integrating data on CRD prevalence and the feature importance of key risk factors into GeoJSON files containing the administrative boundaries of districts and cities in Central Java. Subsequently, a choropleth map was generated to display the CRD prevalence, with varying color intensities representing the prevalence levels across regions. Overlay markers highlighted significant risk factors such as public transportation, cigarette consumption, and NO_2 concentrations. This visualization approach enables a clearer understanding of geographical patterns and the influence of key risk factors, providing valuable insights for targeted intervention strategies.

3. Results and Discussions

3.1. *Simulations settings*

In this section, we are using the decision tree approach as the selected machine learning method to predict chronic respiratory disease (CRD) cases. The model was initialized using the **DecisionTreeRegressor** from the

sklearn.tree library, as the target variable, CRD_case, represents continuous data. To ensure reproducibility, the parameter **random_state=42** was applied, fixing the random seed for consistent results. Model optimization was carried out through hyperparameter tuning using GridSearchCV, which systematically explored various parameter combinations to enhance performance. Key parameters included max_depth, which controls the maximum depth of the tree and the model's complexity; min_samples_split, determining the minimum number of samples needed to split a node; min_samples_leaf, specifying the minimum samples required at a leaf node; and max_features, which limits the maximum number of features considered during splits. The refined parameter combinations identified through this process resulted in an optimized decision tree model capable of effectively predicting and analyzing CRD cases.

3.2. *Performance and Analysis*

The parameter tuning process for the decision tree model was conducted using GridSearchCV, which systematically tested various parameter combinations to identify the optimal configuration. The parameters adjusted included max_depth, min_samples_split, min_samples_leaf, and max_features. Each combination was evaluated based on the model's predictive performance using metrics such as Root Mean Squared Error (RMSE) and R^2 score. The results of the parameter tuning experiments are summarized in Table 1, highlighting the best-performing parameter configuration that achieved the highest evaluation score.

Table 1. Parameter Tuning Results and Best Configuration

Max Depth	Min Samples Split	Min Samples Leaf	Max Features	RMSE	R^2
5	10	4	sqrt	0,75	0,39
10	5	8	None	0,82	0,27
15	5	2	sqrt	0,83	0,26
20	10	6	sqrt	0,85	0,22
10	5	2	$\log^2$	0,42	0,81
5	**5**	**1**	**sqrt**	**0,40**	**0,83**

The results in Table 1 as shown in the highlighted row, utilized a max_depth of 5, min_samples_split of 5, min_samples_leaf of 1, and max_features set to "sqrt," resulting in the lowest **RMSE (0.40)** and highest **R^2 score (0.83)**, which delivered the best evaluation results.

After obtaining the best model through GridSearchCV, the evaluation was conducted using cross-validation and Root Mean Squared Error (RMSE) as performance metrics. Cross-validation ensures the model's ability to generalize new data by splitting the training data into several folds, such as six folds. In each iteration, the model was trained in five folds and tested on one, rotating through all folds. This method provides a more reliable estimate of how the decision tree model would perform on unseen data, thereby reducing the risk of overfitting.

When tested on the full dataset, the trained model achieved an RMSE of 0.40 and an R^2 score of 0.83. The RMSE value of 0.40 reflects the model's low prediction error, while the R^2 score of 0.83 indicates that the model successfully explains 83% of the variance in the target variable (CRD_case). These evaluation results highlight the model's strong predictive performance, with high accuracy and a low error rate, making it a robust approach for understanding chronic respiratory disease prevalence in the given dataset.

Next, feature importance is determined to assess the contribution of each feature in predicting the target variable. Feature importance measures the impact of each feature on the data splits throughout the decision tree. These values are derived from the model selection process and indicate the relative contribution of each feature in segmenting the data. Higher importance values suggest a greater influence of the feature on the model's decision-making process. In this context, the feature importance reflects how significantly each independent variable impacts the prediction of chronic respiratory disease (CRD) cases. The greater the feature's importance, the more substantial its role in shaping the model's decisions. Visualization of feature importance is then performed to facilitate the identification of key features with the most significant contributions to the model. The following Table 2. presents the feature importance values and their corresponding visual representation.

Table 2. Feature Importance Value

Feature	Importance Value
transport	0.3410
cigarette	0.2537
NO_2	0.1310
tourism spot	0.0833
alcohol store	0.0656
SO_2	0.0403
healthy food	0.0352
fast food	0.0269
gym	0.0229
sweet drink	0.0000

From the feature importance Table 2, the visualization is presented on a map of districts and cities in Central Java using GeoJSON. This map illustrates the prevalence of chronic respiratory diseases (CRD) across regions, with color intensity representing the CRD case distribution of significant risk factors, such as public transportation facilities, cigarette consumption rates, and NO_2 concentrations in each area shown.

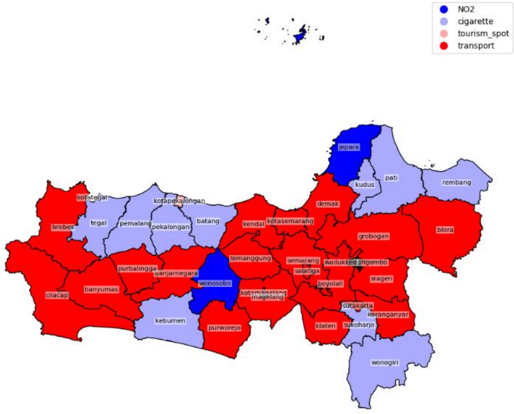


Fig.1 Feature importance of geo-mapping in Central Java

The visualization shown Fig.1 results reveal that the number of public transportation facilities (transport) has the highest feature importance value of **0.3410**, indicating that public transportation is a major risk factor for chronic respiratory diseases (CRD). The map visualization confirms that most regions in Central Java are dominated by this risk factor. This can be interpreted as the contribution of air pollution generated by public transportation vehicles, along with other motorized vehicles such as cars and motorcycles. Areas with dense public transportation facilities are often highly exposed to vehicle emissions, including nitrogen dioxide (NO_2) and fine particulate matter (PM2.5), which are concentrated in high-traffic zones, increasing the risk of respiratory diseases. Long-term exposure to these pollutants is scientifically linked to airway inflammation, decreased lung function, and a heightened risk for individuals with pre-existing chronic respiratory conditions. Furthermore, public transportation systems, particularly in urban areas, tend to have high population densities, which exacerbate exposure to secondhand smoke and other environmental allergens such as dust or mold. These factors may act as triggers that worsen the condition of individuals with chronic respiratory diseases while simultaneously elevating the risk of new cases in vulnerable populations.

Cigarette consumption (cigarette) has a significant importance value of **0.2537**. Smoking is a leading cause of respiratory disorders, including chronic obstructive pulmonary disease (COPD) and asthma. Regions such as Kudus, Pemalang, and Rembang show a dominance of cigarette smoke as a risk factor, indicating a high rate of cigarette consumption in these areas. This not only affects active smokers but also the surrounding community due to passive smoke exposure. The significance of this variable is supported by literature that highlights how the toxic chemicals in cigarette smoke directly damage lung tissue and impair respiratory capacity. This effect is not limited to active smokers but extends to individuals exposed to passive smoke, which contains carcinogenic compounds. Furthermore, children exposed to secondhand smoke are at higher risk of lung growth issues, asthma, and respiratory infections. Additionally, individuals with pre-existing conditions like COPD or asthma may experience exacerbations or worsened symptoms due to smoke exposure.

The NO_2 variable **(0.1310)** also plays a significant role in predicting chronic respiratory diseases. Regions such as Jepara and Wonosobo show NO_2 pollution as a dominant risk factor. Long-term exposure to NO_2 has been scientifically proven to impair lung function and increase the risk of asthma. NO_2 is categorized as a carcinogen and has particularly harmful effects on infants and the elderly. It originates from transportation emissions fueled by fossil fuels, industrial smoke, and environments already contaminated by NO_2. Industrial activities, especially those involving the combustion of fossil fuels such as coal, natural gas, or oil, contribute significantly to NO_2 emissions. Power plants, refineries, and manufacturing factories are examples of industries that generate NO_2.

Other variables, such as tourism spots (tourism_spot) with a value of **0.0833**, found in cities like Pekalongan, and alcohol stores (alcohol_store) with a value of 0.0656, have lower but still relevant contributions. The presence of tourist spots can influence local pollution levels due to increased mobility, while alcohol stores may serve as indicators of urbanization or risky lifestyle patterns. Variables like healthy food (healthy_food), fast food, and gyms show lower importance values (<0.04), indicating minimal impact on the risk of chronic respiratory diseases in Central Java.

4. Conclusion

In this paper, we employed a decision tree model to identify the key risk factors contributing to the prevalence of chronic non-communicable respiratory diseases in Central Java. The model calculates feature importance using Gini impurity, revealing that public transportation, cigarette consumption, and NO_2 concentration are the most significant risk factors, with public transportation having the greatest influence (0.3410). Model evaluation, with an RMSE of 0.40 and R^2 of 0.83, demonstrates strong predictive accuracy.

The visualization of risk factors across districts and cities in Central Java highlights regional variations in risk influences. The map predominantly features red areas, indicating public transportation as the leading risk factor due to air pollution from fossil fuel emissions, notably nitrogen dioxide (NO_2). Regions such as Jepara and Wonosobo show a strong correlation with NO_2 pollution (depicted in blue), reinforcing the link between air pollution and disease prevalence. Additionally, areas marked in purple, such as Kudus, Pemalang, and Rembang, suggest that smoking is the primary risk factor, both through active smoking and passive exposure to secondhand smoke. These findings emphasize the complex interplay of environmental and behavioral factors in shaping respiratory health risks across different regions.

References

1. S. C. Dharmage, J. L. Perret, and A. Custovic, “Epidemiology of asthma in children and adults,” Frontiers in Pediatrics, vol. 7, p. 246, 2019, doi: 10.3389/fped.2019.00246.
2. E. Boers et al., “Global burden of chronic obstructive pulmonary disease through 2050,” JAMA Network Open, vol. 6, no. 12, p. e2346598, 2023, doi: 10.1001/jamanetworkopen.2023.46598.
3. D. E. Bloom, S. Chen, M. Kuhn, M. E. McGovern, L. Oxley, and K. Prettner, “The economic burden of chronic diseases: Estimates and projections for China, Japan, and South Korea,” The Journal of the Economics of Ageing, vol. 17, p. 100163, 2020.
4. H. Y. Lu, C. F. Chen, D. L. Lee, Y. J. Tsai, and P. C. Lin, “Effects of early pulmonary rehabilitation on hospitalized patients with acute exacerbation of chronic obstructive pulmonary disease: A systematic review and meta-analysis,” International Journal of Chronic Obstructive Pulmonary Disease, vol. 18, pp. 881–893, 2023, doi: 10.2147/COPD.S397361.
5. C. T. Wu et al., “Acute exacerbation of chronic obstructive pulmonary disease prediction system using wearable device data, machine learning, and deep learning: Development and cohort study,” JMIR Mhealth and Uhealth, vol. 9, no. 5, p. e22591, 2021, doi: 10.2196/22591.
6. A. Budreviciute et al., “Management and prevention strategies for non-communicable diseases (NCDs) and their risk factors,” Frontiers in Public Health, vol. 8, p. 574111, 2020, doi: 10.3389/fpubh.2020.574111.
7. A. B. Mahammad and R. Kumar, “Machine learning approach to predict asthma prevalence with decision trees,” in 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 263–267, doi: 10.1109/ICTACS56270.2022.9988210.
8. S. B. Khanagar et al., “Developments, application, and performance of artificial intelligence in dentistry: A systematic review,” Journal of Dental Sciences, vol. 16, no. 1, pp. 508–522, 2021.
9. K. Poonpon et al., “Enhancing predictive accuracy in educational assessment: A comparative analysis of machine learning models for predicting student performance,” Review of Contemporary Philosophy, vol. 23, pp. 142–160, 2024.
10. Y. Y. Zhou et al., “Risk factor analysis and clinical decision tree model construction for diabetic retinopathy in Western China,” World Journal of Diabetes, vol. 13, no. 11, pp. 986–1000, 2022, doi: 10.4239/wjd.v13.i11.986.
11. P. Du et al., “The application of decision tree model based on clinicopathological risk factors and pre-operative MRI radiomics for predicting short-term recurrence of glioblastoma after total resection: A retrospective cohort study,” American Journal of Cancer Research, vol. 13, no. 8, pp. 3449–3462, 2023.
12. J. Zhang et al., “Improving wave height prediction accuracy with deep learning,” Ocean Modelling, vol. 188, p. 102312, 2024, doi: 10.1016/j.ocemod.2023.102312.

Authors Introduction

Ms. Ayu Susilowati

She is currently pursuing the Bachelor of Informatics with Honours in Faculty of Informatics, Telkom University Purwokerto, Indonesia. His research interests are Artificial Intelligence, data and bussiness analyst.

Dr. Andi Prademon Yunus

He is an Assistant Professor at Telkom University, and he received his PhD in Engineering from Mie University, Japan. His research focuses on applied and fundamental machine learning for motion and behavior computing. He also collaborates with industry partners to develop AI-based tools for language modeling and image analytics.

Role-Play Prediction using Ontology-Based Graph Convolutional Network Model

Asyafa Ditra Al Hauna[1], Andi Prademon Yunus[1*], Siti Khomsah[1], Yit Hong Choo[2], Masanori Fukui[3]
[1]Telkom University, Banyumas, Indonesia, [2]Deakin University, Australia, [3]Iwate Prefectural University, Japan
Email: alditra@student.telkomuniversity.ac.id, andiay@telkomuniversity.ac.id, sitijk@telkomuniversity.ac.id, y.choo@deakin.edu.au, fukui_m@iwate-pu.ac.jp
[*]Corresponding Author

Abstract

Current applications of large language models often assign tasks without consideration of how LLMs understand a given prompt. Simple commands sometimes do not guarantee desired responses, as LLMs are systems based on mathematical modeling and cannot cognitively be capable of understanding commands. Hence, a method is required to guide LLMs in performing tasks appropriately. This paper presents a method to develop model-based automation of role selection supported by ontology. This can allow for more accurate and relevant role recommendations than if done manually. As such, this optimization at hand improves the performance of LLMs for specific tasks and overcomes the limitations of previous studies that define the roles by hand.

Keywords: Graph Convolutional Network, Large Language Models, Ontology, Role-play

1. Introduction

In recent years, the development of Large Language Models (LLMs) has demonstrated remarkable progress, with researchers competing to advance their findings as breakthroughs in artificial intelligence, particularly in natural language preprocessing. This advancement is primarily driven by innovations in transformer architecture, which enables models to analyze context and semantics through the attention mechanism. Such innovations allow LLMs to exhibit reasoning abilities that can approximate human-like reasoning despite being grounded in mathematical modeling. One prominent example of an advanced LLM is GPT-4, which has made significant strides in adapting to specific tasks through fine-tuning, reinforcement learning with human feedback (RLHF), and domain-specific training [1].

Research on LLMs has shown a notable impact in the technology field and other domains. In education, LLMs promise to enable intelligent tools such as personalized learning [2]. Similarly, in medicine, LLMs can serve as agents that work alongside professionals to analyze complex problems and information [3]. Concurrently, there has been a surge in studies exploring optimization techniques for LLM performance, particularly in specialized domains. One relatively new optimization technique focuses on prompt engineering. Given that an LLM's performance heavily depends on the given prompts, prompt engineering seeks to enhance outputs by designing prompts that lead to the expected response [4][5].

A specific optimization approach within this framework is role-play prompting, which aims to improve LLM performance without requiring intensive or complex interventions. This technique narrows the scope of the LLM's response by assigning a relevant role aligned with the task. For example, a study investigating the impact of assigning roles within prompts reported

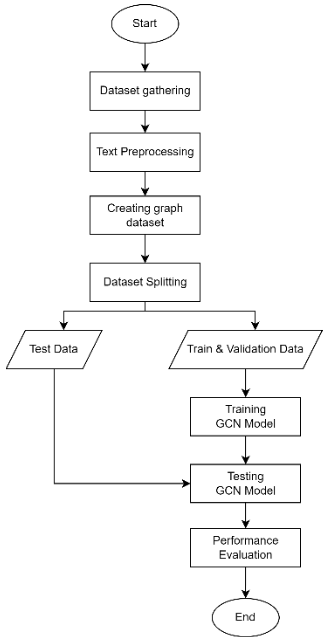


Fig. 1 Methodology flowchart

that manually assigning roles based on task benchmarks improved response accuracy by over 10% [6]. However, because the role assignment in this study was done manually, there is room for automation in determining relevant roles.

This research proposes developing a deep learning model capable of predicting roles based on tasks. Specifically, the focus is on predicting roles such as mathematician, nurse, education teacher, and recorder. A dataset comprising task-relevant information and ontology data, including skills, abilities, and role definitions, will be used to train a Graph Convolutional Network (GCN). The model aims to accurately predict the most relevant role based on the given task.

2. Methodology

The objective of this study is to train a GCN model to predict roles that are relevant to the given task. The procedures undertaken in this research are illustrated in Fig. 1.

2.1. Dataset Gathering

In line with the focus of this research, the model to be developed will be trained using datasets aimed at predicting four roles: mathematician, nurse, education teacher, and recorder. The datasets employed in this study serve as benchmarks commonly used to evaluate reasoning capabilities and the performance of large language models (LLMs). For the mathematician role, the Algebra Question Answering with Rationales (AQUA-RAT) and MultiArith datasets are utilized, containing story problems and mathematical equations [7][8]. To predict the nurse's role, the MedQA dataset is selected as it encompasses medical terminology and concepts [9]. For the teacher's role, the BigBench Date and StrategyQA datasets are employed because they include foundational yet comprehensive questions that reflect the broad knowledge expected of a teacher. Lastly, for the recorder role, the BigBench Object Tracking dataset is used, as it contains questions related to tracking events or occurrences. The ontology applied in this research is Occupation Ontology (OccO), which provides a structured framework consisting of the skills, abilities, and definitions specific to each role [10]. Each of those datasets was gathered from the repository listed in the corresponding paper.

2.2. Text Preprocessing

Text preprocessing is a pivotal step in the field of text data processing, as it directly impacts the overall performance and accuracy of machine learning models [7]. This process must be tailored to the unique characteristics and behaviors of the collected data to ensure effective outcomes, for the details see Table 1. One preprocessing technique involves replacing digits in text data with the word "numeric" using regular expressions. This step is specifically applied to datasets categorized under the mathematician class, as it aids the classifier in recognizing that the term "numeric" is contextually relevant to this category.

Another essential method includes the removal of dates formatted as MM/DD/YYYY, which is executed through regular expressions. This approach is applied exclusively to datasets pertaining to the teacher's class, as dates in this context are considered irrelevant for classification tasks. Similarly, named entity recognition (NER) provided by flair is utilized to identify and remove personal names from the data [8], followed by applying regular expressions. However, this step excludes data from the MedQA class, as it lacks identifiable named subjects.

Furthermore, all datasets undergo lemmatization, a process that identifies the root form of each word, ensuring uniformity in textual representation. This step leverages the lemmatizer provided by the Stanza [9]. Lastly, keyword extraction is conducted across applicable datasets using the KeyBERT algorithm [10] which employs sentence transformers to identify key terms effectively.

Table 1. Text preprocessing for each dataset

Stage / dataset	Replace Digit	Removing Date	Replacing Name	Lemmatization	Keyword Extraction
AQUA-RAT	1	0	1	1	1
Mulriarith	1	0	1	1	1
MedQA	0	0	0	1	1
StrategyQa	0	1	1	1	1
BigBench Date	0	1	1	1	1
BigBench Object Tracking	0	0	1	1	1

For ontology data, the preprocessing stage involves applying stopword removal and lowercasing techniques. These preprocessing steps collectively optimize the datasets, preparing them for subsequent graph dataset creation and task modelling.

2.3. Creating graph dataset

All data from the text preprocessing process is represented as a graph. Each word or keyword is treated as a node in this representation, while the relationships between words and their corresponding datasets are depicted as edges. The first step involves creating six central nodes, each representing one of the dataset names. Subsequently, the text data from each dataset and ontology is iterated to construct a subgraph. In this subgraph, each word is represented as a node, which is then connected to the corresponding central node associated with the dataset.

Since the GCN model requires two inputs, numerical data representing the attributes of each node and an edge index that captures the relationships between nodes [11], the data represented by a node in this study consists of text. It will undergo a word embedding process to convert it into numerical form, this process leveraging sentence transformer [12]. Meanwhile, the edge index will be derived from the iterative process of constructing each subgraph.

2.4. Dataset splitting

In this study, data splitting was conducted using a masking technique on graph-structured data. This masking procedure involves assigning a binary value, either zero or one, to each node in the graph. A zero value indicates that the corresponding node will not be included in the training data, whereas a value of one signifies its inclusion. For the initial experimental phase, the data was divided into three subsets with a ratio of 60%, 20%, and 20%, respectively, representing the training, validation, and test datasets.

2.5. Model building and training

In the development stage of the baseline model, the primary objective is to establish a foundation for assessing the level of complexity required to learn patterns inherent in the data. This model development process leverages GCNs to capture features from graph-structured data. Batch normalization layers are employed to normalize the input of each layer for every batch during training to force the model for faster convergence and introduce slight regularization. Activation functions, such as tanh and ReLU, are incorporated to enable the model to identify non-linear patterns effectively. Additionally, dropout layers are utilized as a regularization technique to mitigate the risk of overfitting, ensuring the model generalizes well to unseen data. Experiment configuration during the training state is shown in Table 2. In addition, the t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm is employed to help visualize the logit of each class generated by the model. It aims to determine how well the model improves classification during training [13].

Table 2. Training configuration

Parameter	Configuration
Epoch	500
GCN layer input size	768
GCN layer hidden size	64
Loss Function	Cross Entropy
Optimizer	Adamax
Learning Rate	$1e^{-3}$
Dropout Probability	0.5

2.6. Model evaluation

In the model evaluation phase, the confusion matrix and ROC-AUC curve are employed to assess the classification performance of the node classifier [14]. The confusion matrix serves as a tool to analyze the classifier's performance by evaluating predictions across four key indicators: accuracy, precision, recall, and F1 score. Accuracy measures the proportion of correctly classified data relative to the total dataset. Precision evaluates the ratio of correctly predicted positive instances to the total instances predicted as positive by the model. Conversely, recall calculates the ratio of correctly predicted positive instances to all actual positive instances in the dataset. The F1 score provides a harmonic mean of precision and recall, evaluating the model's performance when precision and recall are equally important. Additionally, the ROC-AUC curve is utilized to evaluate the classifier's performance based on the false positive rate (FPR) and true positive rate (TPR), visualized on the x-axis and y-axis, respectively, to depict the trade-off between sensitivity and specificity.

3. Results and Discussions

3.1. Training and Testing Model

The training stage is implemented according to the configuration specified in the methodology section. The t-SNE graphs before and after training are compared to suggest how well the model learns the features, as shown in Fig. 2 and Fig. 3.

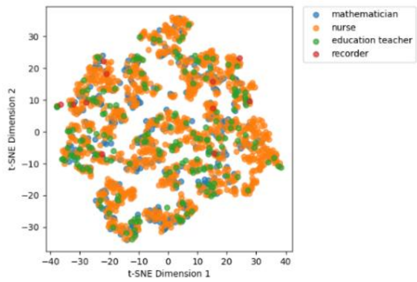


Fig. 2 t-SNE plot before training

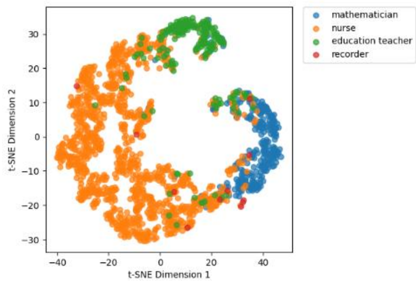


Fig. 3 t-SNE plot after training

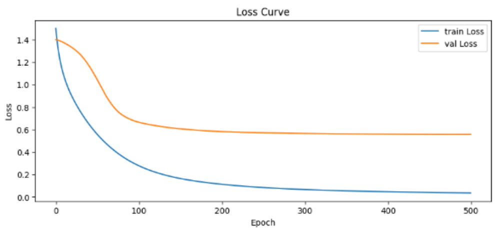


Fig. 4 Loss curve

Based on those graphs, the model can classify nodes well in mathematician, nurse, and education teacher classes but is suboptimal at classifying nodes in the recorder class. The model loss curve during training in Fig. 4 shows that the model learns well due to consistently decreasing training loss, indicating it converges sufficiently. At the beginning of training, the validation loss continues to decrease, but until around the 100th epoch, the validation loss stops decreasing even though the training loss continues to decrease.

3.2. Model Evaluation

The evaluation results on the test data, as presented in Table 3, indicate that the model demonstrates satisfactory performance in classifying data belonging to the mathematician and nurse categories. However, its performance is less than optimal when classifying instances in the education teacher category, and it proves to be particularly ineffective in accurately classifying data associated with the recorder category.

Table 3. Classification report

Class \ Metrics	Precision	Recall	F1-Score
Mathematician	0.74	0.83	0.78
Nurse	0.90	0.86	0.88
Education teacher	0.35	0.38	0.36
Recorder	0.0	0.0	0.0

Furthermore, Fig. 5 corroborates these findings, as the most prominent curves are observed in the mathematician and nurse classes. However, it is important to note that the data distribution within each class does not influence performance evaluation based on the curve area. It explains why the education teacher and recorder classes exhibit relatively good curve areas despite their lower classification performance, exceeding 0.5.

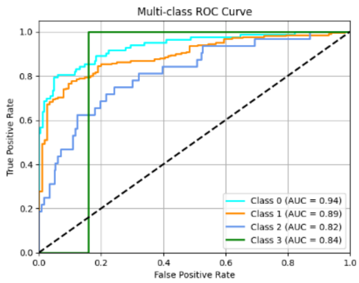


Fig. 5 ROC AUC Curve (class 0: mathematician, class 1: nurse, class 2: education teacher, class 3: recorder)

4. Conclusion

This study proposes a method to optimize prompts for Large Language Models (LLMs). As an initial experiment, the research focuses on predicting four roles: mathematician, nurse, education teacher, and recorder. A GCN algorithm was employed to develop a model capable of predicting relevant roles. Additionally, OccO, an ontology containing occupational role information, was utilized as supplementary data to help the model effectively capture features associated with specific roles. Experimental results indicate that the GCN-based model demonstrated suboptimal performance, particularly in predicting the roles of education teacher and recorder. These findings highlight limitations in the model's ability to generalize across all classes. Future work will enhance the model by integrating other advanced feature engineering techniques to capture text data representations better. Exploring various model architectures and hyperparameter configurations will also be essential to align model complexity with the dataset's characteristics. Furthermore, addressing class imbalance issues should be prioritized to ensure balanced data distribution, thereby improving the model's learning capability.

Acknowledgement

We extend our heartfelt gratitude to Telkom University for its unwavering support during this research. It is important to note that this work did not receive any financial assistance from funding agencies. Nonetheless, we deeply appreciate the researchers and collaborators whose invaluable contributions and insights greatly enhanced the quality of this study. Their dedication and support played a vital role in the successful completion of this research.

References

1. OpenAI *et al.*, GPT-4 Technical Report, 2023, doi: https://doi.org/10.48550/arXiv.2303.08774.
2. W. Gan, Z. Qi, J. Wu, and J. C. W. Lin, Large Language Models in Education: Vision and Opportunities, in *Proceedings - IEEE International Conference on Big Data*, 2023, pp. 4776–4785, doi: https://doi.org/10.48550/arXiv.2311.13160.
3. X. Tang *et al.*, MedAgents: Large Language Models as Collaborators for Zero-shot Medical Reasoning, ACL, 2024, pp. 599–621, doi: https://doi.org/10.18653/v1/2024.findings-acl.33.
4. Y. Zhou *et al.*, Large Language Models Are Human-Level Prompt Engineers, *ICLR*, 2023, doi: https://doi.org/10.48550/arXiv.2211.01910.
5. H. Sun *et al.*, AutoHint: Automatic Prompt Optimization with Hint Generation, 2023, doi: https://doi.org/10.48550/arXiv.2307.07415.
6. A. Kong *et al.*, Better Zero-Shot Reasoning with Role-Play Prompting, *Proc. NAACL*, vol. 1, pp. 4099–4113, 2024, doi: https://doi.org/10.48550/arXiv.2308.07702.
7. A. R. Baskara, M. A. S. Jati, M. Maulida, Y. Sari, N. F. Mustamin, and E. S. Wijaya, Classification of User Reviews for Software Maintenance in Indonesian Language Using IndoBERT-BiLSTM (Case Study: MyPertamina), *8th International Conference on Informatics and Computing, ICIC 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, doi: http://dx.doi.org/10.1109/ICIC60109.2023.10381946.
8. A. Akbik, T. Bergmann, D. Blythe, K. Rasul, S. Schweter, and R. Vollgraf, FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP, NAACL, 2019, pp. 54–59, doi: https://doi.org/10.18653/v1/N19-4010.
9. P. Qi, Y. Zhang, Y. Zhang, J. Bolton, and C. D. Manning, Stanza : A Python Natural Language Processing Toolkit for Many Human Languages, ACL, 2020, pp. 101–108, doi: https://doi.org/10.18653/v1/2020.acl-demos.14.
10. Maarten Grootendorst, KeyBERT: Minimal keyword extraction with BERT, *Zenodo*: 0.8.0, 2020, 10.5281/zenodo.4461265.
11. W. L. Hamilton, Graph Representation Learning, Synthesis Lectures on Artificial Intelligence and Machine Learning, 2020, Vol. 14, No. 3, pp. 47-53, isbn: 978-3031004605.
12. N. Reimers and I. Gurevych, Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, 2019, doi: https://doi.org/10.48550/arXiv.1908.10084.
13. W. Li, J. E. Cerise, Y. Yang, and H. Han, Application of t-SNE to Human Genetic Data, J Bioinform Comput Biol, 2017, http://dx.doi.org/10.1101/114884.
14. T. Fawcett, An introduction to ROC analysis, *Pattern Recognit Lett*, 2006, vol. 27, pp. 861–874, doi: https://doi.org/10.1016/j.patrec.2005.10.010.

Authors Introduction

Mr. Asyafa Ditra Al Hauna

He is currently pursuing the Bachelor of Informatics Engineering with Honours in Faculty of Informatics, Telkom University, Indonesia. His research interests are Machine Learning, Deep Learning, and Natural Language Processing.

Dr. Andi Prademon Yunus

He is an Assistant Professor at Telkom University, and he received his PhD in Engineering from Mie University, Japan. His research focuses on applied and fundamental machine learning for motion and behavior computing. He also collaborates with industry partners to develop AI-based tools for language modeling and image analytics.

Ms. Siti Khomsah, M. Cs

She is an Lecturer in Faculty of Informatics at Telkom University in Indonesia. She received her Master in Computer Science from Universitas Gajah Mada. Her research interests are Text Mining, Data Analytics, and Big Data Analytics.

Dr. Yit Hong Choo

He has completed his PhD and is now a Research Fellow in Operations Analytics at Deakin University's Institute for Intelligent Systems Research and Innovation (IISR). His research focuses on advanced multi-objective optimisation algorithms for complex maintenance scheduling in rolling stock. He also collaborates with transportation industry partners to develop AI-based tools for video and image analytics.

Dr. Masanori Fukui

He is an associate professor at Iwate Prefectural University and received his PhD in Engineering from Hiroshima Unviersity, Japan. His research interests include creativity education, problem-posing, computational thinking, and learning technology.

Experimental Exploration of Neural Style Transfer: Hyperparameter Impact and VGG Feature Dynamics in Batik Motif Generation

Happy Gery Pangestu[1], Andi Prademon Yunus[1*], Siti Khomsah[1], Yit Hong Choo[2], Takao Ito[3]
[1]*Telkom University, Banyumas, Indonesia*
[2]*Deakin University, Australia*
[3]*Hiroshima University, Japan*
Email: happygerypangestu@student.telkomuniversity.ac.id, andiay@telkomuniversity.ac.id, sitijk@telkomuniversity.ac.id, y.choo@deakin.edu.au, itotakao@hiroshima-u.ac.jp
Corresponding Author

Abstract

Innovating traditional batik designs while preserving their cultural essence remains a significant challenge in the intersection of heritage conservation and computational creativity. This study addresses this challenge by optimizing Neural Style Transfer (NST), a deep learning technique to synthesize batik motifs that harmonize structural fidelity and stylistic authenticity. Focusing on hyperparameter adjustments tailored to batik's abstract geometries, we systematically evaluate the impact of layer selection in VGG and pooling operations (max-pooling vs. average-pooling) on style-content synthesis. Experiments reveal that shallow layers (e.g., conv2, conv4) preserve explicit motifs and edge details (SSIM 0.85), while deeper layers (conv16, conv24) generate abstract textures. Average-pooling demonstrates superior stability, achieving smoother convergence (loss stabilized at 0.5 vs. 3.0 for max-pooling) and higher structural coherence (SSIM 0.6963 vs. 0.6634), whereas max-pooling introduces fragmented artifacts due to gradient explosion. The optimized framework, validated through quantitative metrics (MSE Loss 0.02, SSIM 0.82) and qualitative artisan evaluations, successfully transferring the style image into the image, while maintaining the original content of the image, with adjusted hyperparameters. This work advances AI-driven tools for batik preservation, offering a scalable methodology to sustain Indonesia's intangible heritage in the digital era..

Keywords: Batik preservation, Neural Style Transfer, hyperparameter exploration, VGG, pooling operations, cultural coherence.

1. Introduction

Batik, a traditional Indonesian textile art form, is renowned for its intricate motifs and cultural symbolism. Each pattern reflects regional philosophies, histories, and identities, making batik a cornerstone of Indonesia's intangible heritage [1]. Despite its UNESCO recognition in 2009 [2], batik designs remain challenging to innovate while preserving their cultural identity. Artisans and designers must carefully balance traditional values with contemporary aesthetics to create motifs that stay meaningful in a modern context while maintaining cultural authenticity. This delicate balance highlights the complexity of embracing innovation while safeguarding batik's rich heritage.[3].

The tension between tradition and modernity defines batik's creative process, as artisans strive to innovate motifs for contemporary markets while preserving their symbolic essence. Batik patterns encode cultural narratives and spiritual values, demanding careful reinterpretation to avoid diluting their heritage. Globalization and shifting aesthetics pressure designers to modernize motifs through abstract geometries or minimalist palettes yet rigid adherence to tradition risks stagnation. The lack of systematic frameworks forces reliance on trial-and-error methods, while digital tools prioritize efficiency over cultural depth, often sidelining traditional techniques like hand-drawn *canting* or natural dyes. Bridging this divide requires merging technological advancements, such as AI-driven design, with respect for traditional knowledge, ensuring batik evolves as a living art form rooted in cultural identity.[4]. Current practices often rely on manual reinterpretation of historical patterns, a time-intensive process constrained by limited technical guidance and creative inspiration [5]. While digital tools like graphic software have streamlined design workflows, they lack the capacity to autonomously generate culturally coherent patterns or synthesize traditional and contemporary elements [6]. Consequently, there remains a critical need for methodologies that augment human creativity while safeguarding batik's cultural integrity.

Recent advances in computational creativity, particularly Artificial Intelligence (AI), present transformative opportunities for cultural heritage preservation. Among these, Neural Style Transfer (NST), first introduced by Gatys et al. in their foundational work "A Neural Algorithm of Artistic Style" [7], has emerged as a pivotal technique for artistic synthesis. NST algorithm leverages deep convolutional neural networks to decompose images into separable representations of content (structural features) and style (textural patterns), enabling the synthesis of novel artworks by recombining these elements across distinct sources. This framework, which optimizes a generated image to match the content of one input and the stylistic attributes of another via Gram matrix-based style loss, has since inspired adaptations in diverse artistic domains. In batik research, NST has been experimentally applied to generate hybrid designs by fusing traditional motifs with modern visual elements. For example, [8] propose a local style transfer model for batik patterns with enhanced edges and use the Stable Diffusion AI painting tool for style transfer Results: The local style transfer model generated images with good performance in detail texture and color space, achieving a peak signal-to-noise ratio (PSNR) of 25.3 dB and a structural similarity index measure (SSIM) of 0.85. However, the Stable Diffusion tool had limitations in inheriting style and content details. The results is local style transfer model generated images with good performance in detail texture and color space, achieving a peak PSNR of 25.3 dB and a structural similarity index measure (SSIM) of 0.85.. Similarly, [9] demonstrates the use of CNN for batik style transfer, showing that the technique can effectively combine the content of an image with the style of traditional batik, producing visually appealing and culturally coherent designs.

To enhance the application of NST for batik motif synthesis, this study focuses on systematically optimizing hyperparameters critical to style-content adaptation. Batik's abstract geometries and symbolic textures, which diverge from photorealistic imagery, necessitate tailored adjustments to standard NST workflows. Specifically, we investigate (1) selection layers in VGG to determine their impact on preserving batik's structural and stylistic nuances; and (2) pooling operations to evaluate their role in feature aggregation and texture synthesis. By testing these hyperparameters, we aim to develop an NST framework that aligns with batik's artistic principles. The efficacy of the optimized pipeline is validated through quantitative metrics MSE Loss and SSIM.

2. Methodology

The Neural Style Transfer framework begins with feature extraction, where a pretrained convolutional neural network (CNN) is employed to derive hierarchical representations from batik images. These features capture multi-scale attributes ranging from coarse patterns to fine-grained textures, collectively defining the visual lexicon of batik motifs. For style images, the Gram matrix a statistical measure of channel-wise feature correlations is computed from the extracted representations. The Gram matrix $\boldsymbol{G^l_{ij}}$ for layer $\boldsymbol{l}$ is calculated as:

$$G^l_{ij} = \sum_k F^l_{ik} F^l_{jk} \tag{1}$$

where $\boldsymbol{F^l_{ik}}$ and $\boldsymbol{F^l_{jk}}$ denote the activations of the $\boldsymbol{i}$-th and $\boldsymbol{j}$-th filters at spatial position $\boldsymbol{k}$ in layer $\boldsymbol{l}$. This matrix quantifies stylistic textures by capturing interdependencies between feature channels. To assess the impact of domain adaptation, this study compares feature extraction performance between generic pretrained CNNs (e.g., ImageNet-trained models) and counterparts fine-tuned on batik datasets, evaluating their ability to preserve culturally significant visual elements during style transfer.

The Gram matrix calculation serves as the cornerstone for encoding stylistic attributes. When a style image is processed through the CNN, the resulting feature tensor (dimensioned as batch size × channels × height × width) is reshaped into a two-dimensional matrix. The Gram matrix is derived by computing the inner product of this reshaped matrix with its transpose, producing a channel-wise correlation matrix. Each element is normalized by the total number of elements in the original tensor $\boldsymbol{b} \times \boldsymbol{c} \times \boldsymbol{h} \times \boldsymbol{w}$, ensuring scale invariance across varying image resolutions.

Loss computation integrates two components, a content loss and style loss. Content loss quantifies structural fidelity using the mean squared error (MSE) between feature maps of the generated image xx and the original content image $\vec{\boldsymbol{p}}$ expressed as :

$$L_{\text{content}}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} (F^l_{i,j} - P^l_{i,j})^2 \tag{2}$$

where $F^l_{i,j}$ and $P^l_{i,j}$ represent activations of the i-th filter at position j in layer l for the generated and content images, respectively. Style loss evaluates stylistic alignment by comparing Gram matrices of the generated image x and reference style image a [10]. The layer-wise style error E_l is calculated as:

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G^l_{ij} - A^l_{ij})^2 \tag{3}$$

where N_l is the number of filters, M_l is the feature map size, and G^l_{ij} and A^l_{ij} are Gram matrices of the style reference and generated images. The total style loss aggregates contributions across layers:

$$L_{\text{style}}(\vec{a}, \vec{x}) = \sum_{l=0}^{L} w_l e_l \tag{4}$$

with w_l weighting layer-specific importance. The total loss combines these terms as:

$$L_{\text{total}}(\vec{p}, \vec{a}, \vec{x}) = \alpha L_{\text{content}}(\vec{p}, \vec{x}) + \beta L_{\text{style}}(\vec{p}, \vec{x}) \tag{5}$$

where α and β are hyperparameters balancing structural preservation and stylistic adaptation. A total variation regularization term L_{TV} is incorporated to enhance spatial smoothness.

Pixel optimization iteratively minimizes the total loss using the adam optimizer, which adaptively adjusts learning rates per parameter based on gradient momentum $\boldsymbol{m_t}$ and variance $\boldsymbol{v_t}$:

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \varepsilon}} \tag{6}$$

where η is the initial learning rate $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected estimates of the first and second moments, and ϵ prevents division by zero. This approach accelerates convergence while mitigating oscillations common in gradient-based optimization. Over successive epochs, pixel values are updated via backpropagation, aligning the generated image with target content and style attributes.

Evaluation employs quantitative and qualitative metrics, the equation of MSE can be seen in the equation below:

$$MSE = \frac{1}{n}\sum_{i}^{n}(\hat{y}_i - y_i)^2 \quad (7)$$

Where $\hat{y}_i$ is generated pixel images and y_i is reference images [11]. The Structural Similarity Index (SSIM) assesses perceptual fidelity through luminance $(\mu_x\mu_y)$, contrast (σ_x, σ_y), and structural coherence (σ_{xy}):

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (8)$$

where C_1, C_2 stabilize the denominator. These metrics are complemented by expert-led qualitative assessments to ensure cultural relevance and aesthetic integrity in synthesized batik patterns.

3. Results and Discussions

3.1. *Testing Scenario*

The experimental framework evaluated NST performance across multiple scenarios involving diverse batik motifs, CNN architectures, and hyperparameter configurations. Content-style pairings were designed to test both representational fidelity. Key experimental configurations are summarized in Table 1.

Table 1. Testing Scenario of Neural Style Transfer.

Sub	Pooling	Layers		Weight	
		Content	Style	α	β
S1	Max	[2, 8, 16]	[2, 8, 16]	1	1e8
S2	Max	[2, 8, 16]	[2, 8, 16]	1	1e8
S2	Avg	16	8	1	1e8
S4	Max	16	8	1	1e8
S5	Avg	8	8	1	1e8
S6	Max	8	8	1	1e8
S7	[Avg, Max]	16	8	1	1e8
S8	[Avg, Max]	16	8	1	1e8
S9	[Avg, Max]	16	8	1	1e8

3.2. *Performance analysis*

In the implementation of Neural Style Transfer (NST), the selection of layers within the Convolutional Neural Network (CNN) model significantly influences the outcomes of image synthesis. This dependence arises because each layer governs the characteristics of the features extracted during the process. The content representations derived from these layers, which encode structural and textural attributes of the input image, are visually illustrated in Fig. 1, Fig. 2, and Fig. 3.

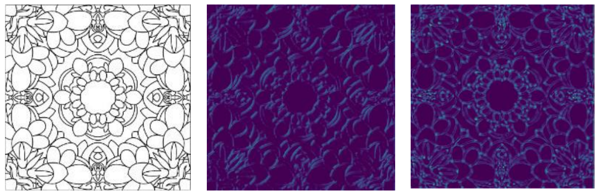


Figure 1. Compare feature map shallow layer of the pattern image.

When NST utilizes shallower layers, the resulting features tend to be simple and focus on the basic elements of the image, such as lines, edges, and basic geometric patterns. This makes the style transfer results more literal and visually recognizable.

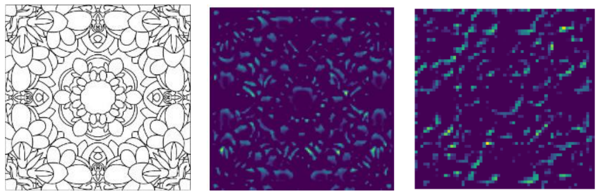


Figure 2. Compare feature map inner layer of the pattern image.

In contrast, the use of deeper layers produces more abstract and complex features. Deep layers capture the representation of the image, such as shape, texture, and deeper concepts. To directly determine the effect of layer depth on the results of style transfer, we also conducted experiments with various layers that would be applied to 2 batik images. The results of the combination of batik motifs can be seen in the following image:

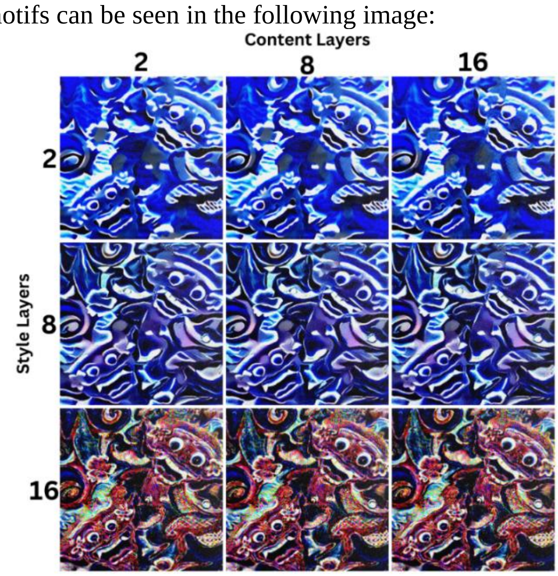


Figure 3. Comparison of layer difference results.

As can be seen in Fig. 3 (Sub 1) above, the effect of layer depth variation on neural style transfer, with a focus on the difference in influence between the content layer and the style layer. The test results show that the variation of the content layer depth (2, 8, 16) does not significantly affect the visual results, indicating that the content representation can be maintained even when using deeper layers. In contrast, the variation of the style layer depth shows a significant effect on the final result, where increasing the style layer depth produces increasingly complex patterns. This can happen because the ability of deeper layers in CNN can capture more complex style features. Therefore, determining the weight value in the loss function is an important factor to balance between maintaining the content structure and applying details to the style.

In the second experiment, we used other batik motifs to test the results of the influence of different layers. The batik used for this test was peacock batik as content and also megamendung batik as style. The batik image and style transfer results can be seen in the image below:

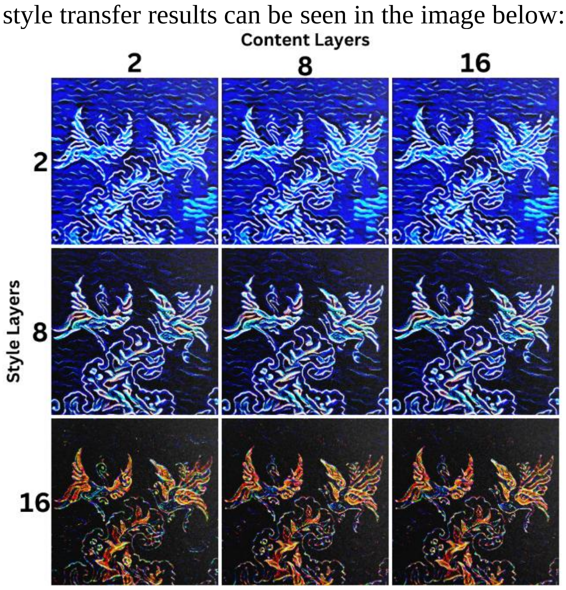


Figure 4. Comparison of difference layer results.

It can be seen in Fig. 4 (Sub 2), the results of combining batik motifs in the second experiment to determine the effect of different layers on the features extractor. The results show that selecting the initial layer up to the first 1/4 of the network provides more visible or more representative transfer results, conversely if the feature map taken from the layer is too deep, the features captured will also tend to be more abstract. The test results show that the selection of layers needs to be adjusted to the desired artistic goals. Shallow layers are more suitable for producing batik with explicit patterns and more representative style transfer results, while deep layers will create designs with a style that tends to be abstract.

We performed a thorough test by comparing the SSIM values for each generated image when utilizing average pooling and max pooling strategies in order to assess the effect of pooling selection on the performance of NST. The purpose of this analysis was to determine how the pooling decision impacts the integrity and quality of the stylized outputs. Our selected batik pattern is megamendung for the style input and barong images for the content input in this experiment. The images below serves as a visual reference for the components involved in the transfer process and displays the content and style images utilized in this experiment:

Figure 5. Comparison of difference layer results.

The left image shows the content source featuring a detailed barong motif, chosen for its intricate textures to test neural style transfer capabilities. The right image displays the style source with a traditional megamendung batik pattern, known for its flowing cloud-like design. The results can be seen in the image below:

Figure 6. Comparison of difference layer results.

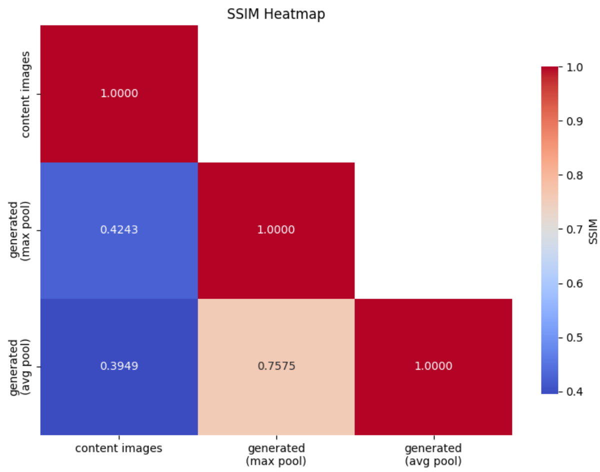


Figure 7. Comparison SSIM of layer difference results.

Fig. 6 and Fig. 7 (Sub 3 and 4) present the experimental results evaluating the impact of distinct pooling operations (max-pooling vs. average-pooling) on style transfer outcomes. The content image, depicting a barong batik motif, and the style image, featuring a megamendung pattern, were processed through the CNN architecture. Synthesized outputs using max-pooling and average-pooling exhibited measurable differences in visual characteristics, quantified via the Structural SSIM. Quantitative analysis revealed that the max-pooling-derived output achieved an SSIM score of 0.4243 relative to the original content, while the average-pooling output scored 0.3949. Notably, the SSIM between the two synthesized outputs was 0.7575, indicating substantial structural similarity despite differing pooling strategies. This suggests that both techniques preserve analogous visual traits in the final stylized image.

Next, we conducted testing using bird-patterned batik and blue megamendung batik. The content and style images can be seen in the image below:

Figure 8. Content (left) and style (right) images.

In this test, the layers used were shallow layers, namely layers 2 and 4. The results presented in the image below:

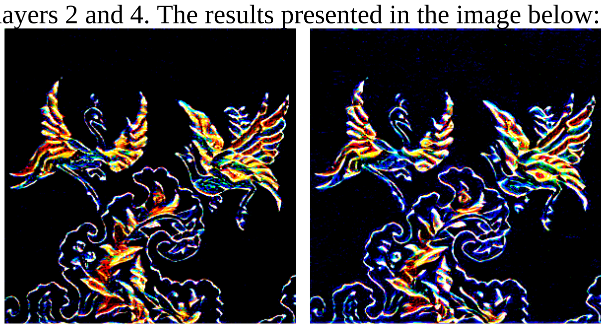


Figure 9. The results generated with max pooling (left) average pooling (right) with epochs 100.

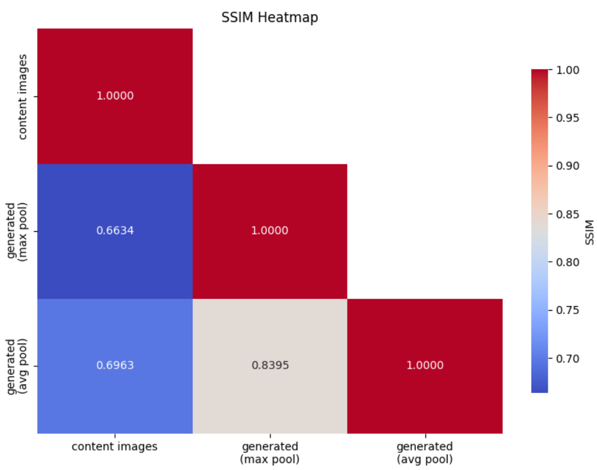


Figure 10. SSIM heatmap of test result image.

Based on the results of the second experiment shown in Fig. 9 and 10 (Sub 5 and 6), the effect of using different pooling layers in the CNN architecture for the style transfer process can be seen. The content image showing the peacock batik motif and the style image with the mega mendung motif are used as input. The results generated using max pooling and average pooling with epochs 100 show slightly different visual characteristics in terms of color. The results generated with max pooling have a structural similarity level of 0.6634 to the original image, while the results with average pooling show an SSIM value of 0.6963. Both generated results have a fairly high similarity to each other with an SSIM value of 0.8395, indicating good consistency of results between the two pooling techniques. Compared to the previous experiment, these results show an increase in the SSIM value indicating better style transfer in peacock batik.

It can be seen in Fig. 4, the results of combining batik motifs in the second experiment to determine the effect of different layers on the features extractor. The results show that selecting the initial layer up to the first 1/4 of the network provides more visible or more representative transfer results, conversely if the feature map taken from the layer is too deep, the features captured will also tend to be more abstract. The test results show that the selection of layers needs to be adjusted to the desired artistic goals. Shallow layers are more suitable for producing batik with explicit patterns and more representative style transfer results, while deep layers will create designs with a style that tends to be abstract.

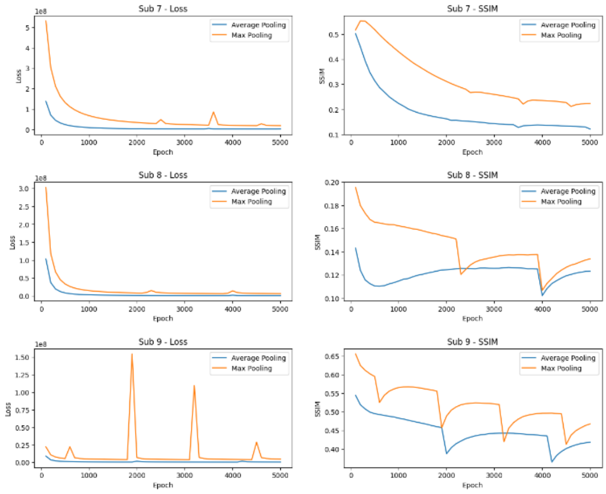


Figure 11. Comparison of loss and SSIM max pooling and average pooling.

The comparative evaluation of max-pooling and average-pooling in NST revealed significant differences in stability and performance (Sub 7, 8, and 9). Max-pooling exhibited exploding gradients, evidenced by erratic loss trajectories (e.g., initial loss of 3.0 dropping sharply to near-zero by epoch 5,000 in Image 6), attributed to its selective gradient amplification through dominant activations. In contrast, average-pooling demonstrated faster and smoother convergence (e.g., loss stabilizing at 0.5 for Image 6), owing to uniform gradient distribution across pooled regions. While max-pooling occasionally achieved marginally higher Structural Similarity Index (SSIM) scores (e.g., Image 22: 0.65 vs. 0.60), it introduced fragmented artifacts, compromising visual coherence. Average-pooling consistently prioritized holistic texture blending, achieving superior stability (e.g., Image 44 SSIM 0.16 vs. max-pooling's 0.12) and mitigating high-frequency distortions. These results underscore average-pooling's robustness for stable training and culturally nuanced batik synthesis, balancing edge preservation with reliable convergence.

4. Conclusion

The experimental investigation into NST for batik motif synthesis underscores the critical interplay between layer selection and pooling strategies in balancing artistic fidelity and computational stability. Shallow layers (e.g., layers 2, 4, 8) excel at preserving explicit patterns and structural details, producing literal style transfers aligned with traditional batik aesthetics. In contrast, deeper layers (e.g., layers 16, 24) capture abstract textures and complex geometries, enabling creative reinterpretations but risking loss of motif recognizability. Notably, style layer depth significantly influences output complexity, while content layer depth exhibits minimal impact, highlighting the importance of prioritizing style representation in parameter tuning. Pooling operations further delineate performance trade-offs: max-pooling, despite occasional edge sharpness (e.g., SSIM 0.65 in structured motifs), suffers from exploding gradients and fragmented artifacts due to selective gradient amplification. Average-pooling ensures stable convergence, smoother texture blending (e.g., SSIM 0.6963 in peacock batik), and superior preservation of holistic patterns, making it ideal for culturally nuanced applications. These findings advocate for average-pooling coupled with shallow style layers to harmonize stability, aesthetic coherence, and computational efficiency in batik-inspired NST algorithm.

Acknowledgements

This work was supported by Telkom University.

References

1. A. Amanzholova And I. Wirasari, "Batik Innovation: A Harmony Of Indonesian Batik Motifs And Kazakh Motifs," *International Journal Of Multicultural And Multireligious Understanding* , Vol. 11, Pp. 356–362, 2024.
2. R. Febriani, L. Knippenberg, And N. Aarts, "The Making Of A National Icon: Narratives Of Batik In Indonesia," *Cogent Arts Humanit*, Vol. 10, No. 1, 2023.
3. M. Xiao, "Innovative Applications And Market Impact Of Indonesian Batik In Modern Fashion," Studies In Art And Architecture, Vol. 3, No. 2, Pp. 62–66, Jun. 2024.
4. N. M. Faizah, "Transformation Of Batik As A Symbol Of Nationalism And National Identity," Kamara Journal, Vol. 1, Pp. 24–35, 2024.
5. V. Jenny Basiroen, M. Purbasari Wahidiyat, F. Marcelinna Suhendra, And D. Carolina, "Bridging Tradition And Innovation: Exploring Design Thinking For Lasem Batik Tulis Motif Creation," In E3s Web Of Conferences, Edp Sciences, Sep. 2023.
6. K. W. Mas'udah And R. D. Issafira, "Improving Skills In Batik Design Through Digital Application For A Pesantren Community In Jombang," Asean Journal Of Community Engagement, Vol. 6, No. 2, Dec. 2022.
7. L. A. Gatys, A. S. Ecker, And M. Bethge, "A Neural Algorithm Of Artistic Style," Aug. 2015.
8. J. Zhang And J. Yan, "Style Transfer Technology Of Batik Pattern Based On Deep Learning," Journal Of Fiber Bioengineering And Informatics, Vol. 16, No. 1, Pp. 57–67, 2023.
9. G. Guntur, P. Ponimin, And M. A. J. Purnomo, "Innovation And Creativity In Batik Motif Design: A Study Of Students' Art Theses," *Creativity Studies*, Vol. 16, No. 2, Pp. 668–681, Oct. 2023.
10. L. A. Gatys, A. S. Ecker, And M. Bethge, "Image Style Transfer Using Convolutional Neural Networks."
11. S. Sima, T. Neda, And Akbar Siami, *The Performance Of Lstm And Bilstm In Forecasting Time Series*. 2019.

Authors Introduction

Happy Gery Pangestu

He is currently pursuing the Bachelor of Data Science in Telkom University, Indonesia. His research interests are machine learning, deep learning, Computer Vision, Natural Language Processing.

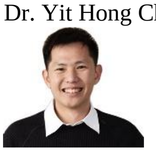
Andi Pranedemon Yunus, Ph.D

He is an Assistant Professor at Telkom University and he received his PhD in Engineering from Mie Unviersity, Japan. His research focuses on applied and fundamental machine learning for motion and behavior computing. He also collaborates with industry partners to develop AI-based tools for language modeling and image analytics.

Siti Khomsah, M.Cs

She is an Lecturer in Faculty of Informatics at Telkom University in Indonesia. She received her Master in Computer Science from Universitas Gajah Mada. Her research interests are Text Mining, Data Analytics, and Big Data Analytics.

Dr. Yit Hong Choo



He has completed his PhD and is now a Research Fellow in Operations Analytics at Deakin University's Institute for Intelligent Systems Research and Innovation (IISR). His research focuses on advanced multi-objective optimisation algorithms for complex maintenance scheduling in rolling stock. He also collaborates with transportation industry partners to develop AI-based tools for video and image analytics.

Dr. Takao Ito

He is Professor of Management of Technology (MOT) in Graduate School of Advanced Science and Engineering at Hiroshima University. His current research interests include automata theory, artificial intelligence, systems control, quantitative analysis of interfirm relationships using graph theory, and engineering approach of organizational structures using complex systems theory.

Exploring Non-Communicable Disease Risk Factors on Cancer Rates in Central Java Using Random Forest and SHAP

Novi Ramadani, Andi Prademon Yunus*
Telkom University, Banyumas, Indonesia
Email: novrmd@student.telkomuniversity.ac.id, andiay@telkomuniversity.ac.id
**Corresponding Author*

Abstract

Non-communicable diseases (NCDs), including cancer, account for over 75% of global deaths, with cancer being the second leading cause. In Indonesia, cancer ranks third, and Central Java has a high prevalence of 1.7 per 1,000 population in 2023. This study analyzed cancer incidence in Central Java using a Random Forest model combined with SHapley Additive exPlanations (SHAP) to assess local feature importance. Key risk factors identified include the presence of sugar-sweetened beverage outlets, which affect 28.6% of districts, air pollution from SO_2 and NO_2, which impacts 22.9% and 11.4% of districts, respectively, and the presence of slum areas, which is associated with higher cancer risks in 8.6% of districts. These findings offer insights into targeted public health strategies aimed at reducing cancer incidence and improving community health.

Keywords: Non-communicable diseases, Chronic respiratory diseases, Decision tree, Geo-mapping

1. Introduction

Non-communicable diseases (NCDs), including heart disease, stroke, cancer, diabetes, and chronic respiratory diseases, are responsible for over 75% of global deaths. Among these, cancer stands as the second leading cause of mortality worldwide, claiming approximately 10 million lives annually [1]. In Indonesia, cancer ranks as the third leading cause of death, following cardiovascular diseases and maternal health issues [2]. Based on data from the Indonesian Ministry of Health in 2023, Central Java, in particular, is among the top five provinces with the highest cancer prevalence, recorded at 1.7 per 1,000 population.

The main risk factors for non-communicable diseases, including cancer, encompass unhealthy eating habits (such as consuming junk food), smoking, alcohol use, physical inactivity, and conditions like high blood pressure, high blood glucose, high cholesterol, and obesity. Socio-economic and environmental factors, such as poverty, limited healthcare access, low public health spending, air pollution, climate change, and sun exposure, also contribute [3], [4]. Understanding these risk factors is crucial for designing effective public health interventions. This study examines the spatial distribution and relative importance of cancer risk factors in Central Java, focusing on the presence of fast-food outlets, tourist attractions, slum areas, transportation hubs, alcohol stores, sugary drinks outlets, gyms, sports halls, smoking prevalence, and air pollution (NO_2 and SO_2).

Machine Learning (ML) has emerged as a transformative tool for analyzing complex datasets, enabling researchers to uncover non-linear relationships and patterns [5]. Among ML algorithms, Random Forest (RF) is widely regarded for its robustness and versatility. RF operates by constructing multiple decision trees and combining their outputs, improving accuracy, and mitigating overfitting. Moreover, RF provides insights into feature importance, identifying key predictors of outcomes [6][7]. In this study, RF is employed to evaluate the global importance of cancer risk factors, complemented by SHapley Additive exPlanations (SHAP) to assess local variations and spatial distributions. SHAP values explain the contribution of each feature to the model's predictions for individual data points, offering insights into the influence of features on the model's decisions for each sample [8].

This research sheds light on the spatial disparities of cancer risk factors in Central Java, offering valuable insights for tailored public health strategies. By utilizing advanced computational methods like RF and SHAP, this study demonstrates the potential of Machine Learning in addressing intricate public health challenges and supporting data-driven decision-making.

2. Material and Method

2.1. Material

2.1.1. Cancer Data

Cancer cases data were obtained from the Health Office of Central Java Province, which includes information on the incidence of cancer, covering leukemia, cervical cancer, breast cancer, retinoblastoma, and colorectal cancer in each administrative region. The administrative regions consist of 29 regencies and 6 cities in Central Java. This data serves as the foundation for understanding the spatial distribution of cancer cases in the region.

2.1.2. Risk Factors Data

Risk factor data were collected through various methods. Information related to public facilities and community behaviors, such as sports hall, gym, tourist attractions, fast food outlets, sugary drinks outlet, alcohol store, and transportation hub, was obtained through web scraping techniques from Google Maps. Additionally, air pollution data (NO_2 and SO_2) and smoking prevalence data were collected from Central Bureau of Statistics (BPS) reports. Information regarding slum area size was obtained from Central Java Province Settlement Area Information System reports.

2.2. Method

2.2.1. Random Forest (RF)

Random Forest (RF) is a non-parametric machine learning method for classification and regression [8]. It builds multiple decision trees from randomly selected samples in the training data. The main steps of RF are:

1. Randomly select n samples from the training set with replacement (usually 2/3 of the data). The remaining third is used for out-of-bag (OOB) estimation.
2. For each sample, randomly select a subset of variables and create a decision tree.
3. Trees grow to their maximum size without pruning.
4. The prediction/classification is based on the majority vote (for classification) or average (for regression) from all trees.

The OOB method is also utilized to evaluate the significance of each independent variable. A common approach to measure importance is by calculating the increase in Mean Squared Error (%IncMSE). This technique involves randomly permuting the values of each variable in the OOB sample and then calculating the resulting OOB error. If the OOB error rises with the permuted values, it suggests that the variable plays a significant role. The greater the increase in error, the more crucial the variable is for predicting the dependent variable [9].

To evaluate the performance and goodness-of-fit of the RF model, the following common metrics are computed: Root Mean Square Error ($RMSE$) (1) and coefficient of determination (R) (2).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (2)$$

Where y_i is the actual value of observation for sample i, $\hat{y}_i$ is the predicted value for observation i, and $\bar{y}$ is the average value of the dependent variable, and n refers to the total sample.

2.2.2. SHAP (Shapley Additive Explanations)

SHAP (Shapley Additive Explanations) method is employed to measure feature importance at the local level within predictive models, providing clearer insights into the contribution of each feature to the model's decision-making process. In this study, SHAP can be used to evaluate how specific risk factors influence cancer risk predictions in different regions of Central Java.

SHAP generates locally additive feature attributions, meaning that the contribution of each feature to the model's prediction can be calculated and summed up to determine the difference between the actual prediction and the model's average prediction, as described in

$$\hat{y}_i = shap_0 + shap(X_{1i}) + shap(X_{2i}) + \cdots + shap(X_{pi}) \quad (3)$$

Where $shap_0 = E(\hat{y})$ is the average prediction for all observations, and $shap(X_{ji})$ represents the SHAP value of the j^{th} feature for observation i, indicating the feature's marginal contribution to the prediction. The sum of all SHAP values equals the difference between the actual prediction and the average prediction. SHAP values maintain the properties of Shapley values. Additionally, the absolute SHAP value indicates the magnitude of the feature's influence on the model's prediction, making it a useful measure of feature importance [10].

3. Results

Before fitting the RF model and to prevent overfitting, we used Grid Search to find the optimal values for the hyperparameters. After testing various combinations of hyperparameter values through K-fold cross-validation, we obtained the following settings for the RF model: The best hyperparameter configuration, determined through 3-fold cross-validation, resulted in an of $\boldsymbol{R^2}$ 0.84 and an $\boldsymbol{RMSE}$ of 0.39. The optimal configuration used the following values: n_estimators = 80, max_depth = 7, min_samples_split = 2, min_samples_leaf = 1, and max_features = sqrt.

The importance of the independent variables in the Random Forest (RF) model is illustrated in Fig. 1. Variables with a higher percentage increase in the mean squared error (%IncMSE) are considered more important. The top five variables contributing to cancer rates are SO_2, sugary drinks, smoking prevalence (%), NO_2, and slum area.

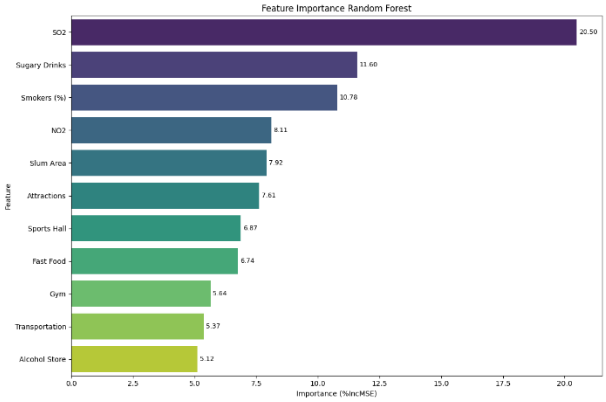


Fig. 1. RF feature importance

We also determined the proportion of counties sharing the same primary local risk factor, defined as the factor with the highest importance (Table 1). For instance, sugary drinks are the factor with the highest importance in 28.6% of counties.

Table 1. Counties sharing the same primary factor, identified as the one with the highest importance.

Local Primary Factor	**Share of Counties (%)**
Sugary Drinks	28.6%
SO_2	22.9%
NO_2	11.4%
Slum Area	8.6%
Attractions	8.6%
Smokers (%)	5.7%
Fast Food	5.7%
Sports Hall	5.7%
Alcohol Store	2.9%
Transportation Hub	0%
Gym	0%

Unsurprisingly, factors such as consumption habits, air pollution, and environmental conditions, including sugary drinks (28.6%), SO_2 (22.9%), NO_2 (11.4%), and slum area (8.6%), rank as the most influential factors contributing to cancer incidence in 71.5% of regions in Central Java (Table 1, Fig. 2). Additionally, attractions (8.6%), followed by smokers (5.7%), fast food outlets (5.7%), sports hall (5.7%), and alcohol stores (2.9%), are also identified as primary risk factors. The spatial distribution of these factors reveals intriguing patterns (Fig. 2).

Sugary drinks dominate as the primary factor in Banyumas, Boyolali, Brebes, Grobogan, Kebumen, Kendal, Klaten, Semarang City, Semarang, and Tegal. SO_2 emerges as the most influential risk factor in Banjarnegara, Batang, Blora, Surakarta City, Tegal City, Pemalang, Purworejo, and Rembang. NO_2 is dominant in Cilacap, Magelang, Pati, and Purbalingga, while slum areas significantly impact Karanganyar, Kota Salatiga, and Wonosobo. Meanwhile, attractions are the leading factor in Pekalongan, Sragen, and Sukoharjo. Smokers (10%) influence Demak and Wonogiri, fast food outlets dominate in Jepara and Pekalongan City, sports halls are influential in Magelang City and Kudus, and alcohol stores are significant in Temanggung.

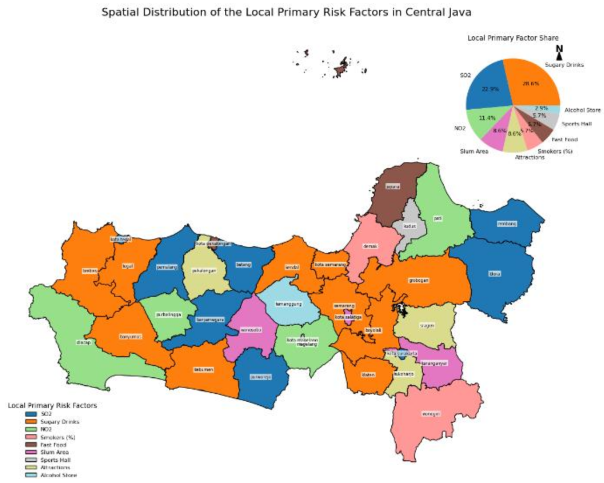


Fig. 2. Spatial distribution of key factor importance

4. Discussion

This cross-sectional ecological study analyzed cancer incidence rates at the county level in Central Java in 2023. Using a Random Forest model (R^2=0.84, RMSE=0.39), the study effectively captured complex, non-linear relationships between local risk factors and cancer incidence. It aimed to identify the key environmental, socio-economic, and lifestyle factors influencing cancer rates. Below, we explore these factors in detail and emphasize their significance.

4.1 Sugary Drinks

Sugary drinks are a primary local risk factor affecting 28.6% of Central Java's districts, including Banyumas, Boyolali, Brebes, Grobogan, Kebumen, Kendal, Klaten, Kota Semarang, Semarang, and Tegal. Long-term high sugar consumption increases the risk of non-communicable diseases like cancer, obesity, and type 2 diabetes [11]. Strategies such as reducing sugar intake, clear nutritional labeling, and public health campaigns are essential to address this issue.

4.2 SO_2

SO_2 impacts 22.9% of districts, primarily from fossil fuel combustion in power plants and vehicles. Prolonged exposure to SO_2 is linked to tissue damage and oxidative stress, increasing cancer risk [12]. Industrial areas show the highest SO_2 influence, highlighting the need for stricter emission regulations and investment in eco-friendly technologies to improve public health.

4.3 NO_2

Nitrogen dioxide (NO_2) is a dominant risk factor in Cilacap, Magelang, Pati, and Purbalingga, largely due to vehicle emissions and industrial activity. Long-term NO_2 exposure is associated with systemic inflammation and increased cancer risk [13]. Policies targeting emission reduction and better air quality monitoring systems are critical to mitigate health risks.

4.4 Slum Area

Slum areas, characterized by overcrowding, poor sanitation, and limited access to clean water [14], represent a dominant risk factor in 25.7% of districts. These areas often experience higher pollution levels and limited healthcare access, which exacerbate cancer risks. Urban development programs aimed at improving infrastructure and health services are crucial to mitigating these challenges.

5. Conclusions

This study examined cancer incidence rates in Central Java and identified key risk factors using a Random Forest model, complemented by SHapley Additive exPlanations (SHAP) for detailed local feature importance analysis. The findings revealed that local environmental, social, and lifestyle factors contribute significantly to the variation in cancer rates across regions. Sugar-sweetened beverage consumption affected 28.6% of districts, highlighting the importance of public health strategies to promote healthier dietary habits. Air pollution from SO_2 and NO_2 emissions was a major risk in 22.9% and 11.4% of districts, respectively, emphasizing the need for stricter pollution control policies and cleaner technologies. Additionally, slum areas were associated with higher cancer risks in 8.6% of districts, underscoring the importance of improving living conditions and access to healthcare.

These findings provide valuable insights for policymakers to design targeted interventions that address local risk factors, reduce cancer incidence, and promote healthier communities.

References

1. World Health Organization: WHO. (2024, December 23). *Noncommunicable diseases*. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases.
2. World Health Organization: WHO. *Noncommunicable Diseases Data Portal*. https://ncdportal.org/.
3. Manderson, L., & Jewett, S. (2023). Risk, lifestyle, and non-communicable diseases of poverty. Globalization and Health, 19(1). https://doi.org/10.1186/s12992-023-00914-z.
4. Piovani, D., Nikolopoulos, G. K., & Bonovas, S. (2022). Non-Communicable Diseases: the Invisible Epidemic. *Journal of Clinical Medicine*, *11*(19), 5939. https://doi.org/10.3390/jcm11195939
5. Grekousis, G. (2018). Artificial neural networks and deep learning in urban geography: A systematic review and meta-analysis. *Computers Environment and Urban Systems*, *74*, 244–256. https://doi.org/10.1016/j.compenvurbsys.2018.10.008.
6. Breiman, L. (2001). Random forest. *Machine Learning*, *45*(1), 5-32.
7. Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, 2024, 69–79. https://doi.org/10.58496/bjml/2024/007.
8. Rodríguez-Pérez, R., & Bajorath, J. (2020). Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *Journal of Computer-Aided Molecular Design*, 34(10), 1013–1026. https://doi.org/10.1007/s10822-020-00314-0.
9. Georganos, S., Grippa, T., Gadiaga, A. N., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., & Kalogirou, S. (2019). Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, 36(2), 121–136. https://doi.org/10.1080/10106049.2019.1595177.
10. Li, Z. (2022). Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. *Computers Environment and Urban Systems*, *96*, 101845. https://doi.org/10.1016/j.compenvurbsys.2022.101845.
11. De Lorgeril, M., Salen, P., & Rabaeus, M. (2020c). Sugary drinks and cancer risk. *Translational Cancer Research*, 9(5), 3172–3176. https://doi.org/10.21037/tcr-2020-003.
12. Xu, Q., Lu, C., Murithi, R. G., & Cao, L. (2021b). Increase associated risk of gynaecological cancer due to long-term exposure to high concentration of atmospheric SO_2 industrial pollutant. *Indoor and Built Environment*, 31(8), 2183–2192. https://doi.org/10.1177/1420326x211003655.
13. Amadou, A., Praud, D., Coudon, T., Deygas, F., Grassot, L., Dubuis, M., Faure, E., Couvidat, F., Caudeville, J., Bessagnet, B., Salizzoni, P., Leffondré, K., Gulliver, J., Severi, G., Mancini, F. R., & Fervers, B. (2022b). Long-term exposure to nitrogen dioxide air pollution and breast cancer risk: A nested case-control within the French E3N cohort study. *Environmental Pollution*, *317*, 120719. https://doi.org/10.1016/j.envpol.2022.120719.
14. Ssemugabo, C., Nalinya, S., Lubega, G. B., Ndejjo, R., & Musoke, D. (2020b). Health Risks in our Environment: Urban slum Youth' perspectives using photovoice in Kampala, Uganda. *Sustainability*, *13*(1), 248. https://doi.org/10.3390/su13010248.

Authors Introduction

Novi Ramadani

She is currently pursuing a Bachelor of Informatics with at the Faculty of Informatics, Telkom University Purwokerto, Indonesia. Her research interests include Artificial Intelligence, data analysis, and business analytics.

Andi Prademon Yunus, Ph. D

He is an Assistant Professor at Telkom University, and he received his PhD in Engineering from Mie University, Japan. His research focuses on applied and fundamental machine learning for motion and behavior computing. He also collaborates with industry partners to develop AI-based tools for language modeling and image analytics.

Addressing Noise Challenges in CNN-based Pneumonia Detection: A Study using Primary Indonesian Thoracic Imagery

Wahyu Andi Saputra[1], Andi Prademon Yunus[1*]
[1]*Telkom University, Banyumas, Indonesia*
Email: andiwahyu@telkomuniversity.ac.id, andiay@telkomuniversity.ac.id

Abstract

Accurate pneumonia diagnosis is vital, especially in resource-limited areas like Indonesia. While CNNs show promise for automated detection using chest X-rays, real-world image quality affects their performance. This study addresses this challenge by using a primary dataset—images directly from Indonesian patients—to avoid the biases of pre-processed secondary data. This ensures our findings are relevant to the Indonesian context. We tested how different noise types (salt-and-pepper and Gaussian) impact the accuracy of several common CNN architectures. These noise types mimic common image imperfections. Our analysis reveals that noise degrades the CNN's ability for 3% to 5% performance. This highlights the need for better pre-processing methods and potentially specialized CNN designs to handle noisy images. Ultimately, our work improves our understanding of deploying CNNs for pneumonia diagnosis in real-world settings, leading to more reliable and helpful diagnostic tools. Using primary data from diverse populations is crucial for building trustworthy AI in healthcare.

Keywords: pneumonia classification, CNN, salt-and-pepper noise, gaussian noise

1. Introduction

Pneumonia remains a significant global health challenge, particularly in resource-limited areas such as Indonesia, where access to advanced diagnostic tools and medical expertise is often restricted. The World Health Organization (WHO) identifies pneumonia as one of the leading causes of mortality in children under five years old, underscoring the urgent need for efficient and accurate diagnostic methods [1]. In such contexts, the development and deployment of automated diagnostic systems, particularly those leveraging deep learning technologies like Convolutional Neural Networks (CNNs), offer a promising solution to enhance healthcare delivery and outcomes.

Convolutional Neural Networks have emerged as a powerful tool in the field of medical imaging, capable of learning complex patterns and features directly from raw data. Their application in diagnosing diseases from chest X-rays has been extensively researched, with numerous studies demonstrating their potential to achieve diagnostic accuracy comparable to that of human experts [2]. CNNs are particularly attractive due to their ability to process large datasets and their adaptability to various imaging modalities. However, the performance of CNNs is heavily contingent on the quality of the input data, a factor that poses significant challenges in real-world clinical settings.

In clinical environments, especially those with limited resources, the quality of medical images can be compromised by several factors, including equipment limitations, varying imaging protocols, and environmental conditions [3][4]. These factors often introduce noise and artifacts into the images, which can adversely affect the performance of CNNs. Common types of noise encountered in medical imaging include salt-and-pepper noise, characterized by random occurrences of black and white pixels, and Gaussian noise, which manifests as variations in pixel intensity following a Gaussian distribution. Such noise can obscure critical features necessary for accurate diagnosis, leading to potential misclassification and reduced diagnostic reliability.

To ensure robust model performance, this study utilizes primary Indonesian patient image data, prioritizing authenticity over the potential biases of secondary datasets. However, the inherent limitations of primary data necessitate data augmentation. Augmentation artificially expands the training dataset, creating a diverse training environment for the CNN [5], [6]. This mitigates overfitting and improves the model's ability to generalize to unseen data, ultimately enhancing its reliability in real-world Indonesian clinical settings. The combined approach addresses both data quality and model generalizability.

The central focus of this study is to quantify the impact of different noise types on the accuracy of CNN architectures commonly used for pneumonia diagnosis. By simulating noise conditions that mimic real-world image imperfections, we aim to systematically evaluate how these disturbances affect CNN performance. The findings from our analysis highlight the critical need for improved pre-processing methods that can effectively mitigate the impact of noise on medical images. Techniques such as noise reduction algorithms, image enhancement, and normalization can play a vital role in preserving the integrity of image features essential for accurate diagnosis. Additionally, there is a compelling case for the development of specialized CNN architectures that incorporate mechanisms for noise robustness, such as attention mechanisms or noise-resistant feature extraction layers.

By enhancing our understanding of how noise affects CNN performance, this study contributes to the broader goal of deploying reliable and effective AI tools in healthcare settings. The insights gained can inform the design of diagnostic systems that are not only accurate but also resilient to the variability and imperfections inherent in clinical imaging. This is particularly important in regions like Indonesia, where the successful integration of AI technologies into healthcare systems can significantly improve diagnostic capacity and patient outcomes.

The use of primary data from diverse populations is a crucial step toward building trustworthy AI systems in healthcare. It ensures that the models developed are representative and applicable to the populations they are intended to serve. This approach aligns with the broader objectives of ethical AI development, which prioritize transparency, fairness, and inclusivity. By focusing on primary datasets, this study addresses the critical issue of data representativeness, thereby enhancing the credibility and acceptance of AI-driven diagnostic tools in clinical practice.

This study provides valuable insights into the challenges and opportunities associated with using CNNs for pneumonia diagnosis in resource-limited settings. By systematically evaluating the impact of noise on CNN performance and emphasizing the importance of primary data, we lay the groundwork for future research and development efforts aimed at creating robust, reliable, and contextually relevant AI solutions for healthcare. The findings not only advance the field of medical imaging but also contribute to the broader discourse on the responsible and effective deployment of AI technologies in global health.

2. Noise Emergence in Primary Dataset Acquisition

The retrieval of primary datasets in medical imaging, particularly for tasks such as pneumonia classification, is inherently susceptible to the presence of noise. This noise can arise from a multitude of sources, each contributing to the degradation of image quality and the subsequent performance of Convolutional Nseural Networks (CNNs). In the context of chest X-ray images, several key factors contribute to the emergence of noise.

One of the primary causes of noise in radiographic images is the inherent physical properties of the imaging process itself. For instance, in conventional X-ray techniques, secondary radiation and scattered radiation from the patient can introduce significant noise into the images. Additionally, digital radiography systems are prone to noise from various elements of the system, including the CCD camera, imaging screen, X-ray source, and controller circuits. These sources can lead to the presence of Poisson noise, salt and pepper noise, and speckle noise, which manifest as random fluctuations in pixel intensity or as white and black pixels scattered across the image. Furthermore, electronic interferences in receiver circuits and radiofrequency emissions due to thermal effects can also contribute to noise in modalities like MRI3. Fig. 1 illustrates the presence of noise based on different milliampere-seconds (mAs) of the x-ray image asset [7].

The impacts of noise on the performance of CNNs in pneumonia classification are multifaceted and profound. Noise can obscure critical features necessary for accurate diagnosis, leading to misclassification and reduced diagnostic reliability. For example, salt and pepper noise can introduce random black-and-white pixels that may mimic or obscure pathological features, while Gaussian noise can blur the image, reducing the contrast and making it harder for the CNN to distinguish between different regions of the lung. This degradation in image quality can result in a significant drop in the accuracy of CNN models. Studies have shown that even moderate levels of noise can reduce the performance of CNNs by 2.5-5%, highlighting the need for robust pre-processing techniques and noise-robust learning methods to mitigate these effects.

Fig. 1 Noise presence on high milliampere-second (left) and low milliampere-second (right) [7]

Moreover, the presence of noise can exacerbate the issue of label noise, where incorrect annotations due to noise can further complicate the training process of CNNs. Label noise is a significant challenge in medical image analysis, as it can lead to overfitting to noisy labels, thereby reducing the generalizability and reliability of the models1. The inter-observer variability among medical experts can also compound this issue, making it essential to develop methods that can estimate and mitigate noise rates in the dataset. By understanding the causes and impacts of noise, researchers can develop more effective strategies for noise reduction and robust learning, ultimately enhancing the performance and reliability of CNNs in real-world clinical settings.

In conclusion, the emergence of noise in primary dataset retrieval is a critical issue that must be addressed to ensure the accurate and reliable classification of pneumonia images using CNNs. By identifying the sources of noise and understanding its impacts on image quality and model performance, researchers can develop tailored approaches to mitigate these effects. This includes the use of advanced noise reduction techniques, noise-robust learning methods, and sparse regularization to enhance the robustness of CNN models against various types of noise. Such efforts are crucial for the deployment of trustworthy AI-driven diagnostic tools in clinical environments.

3. Dataset's Noise Simulation

3.1. *Dataset Collection*

The primary dataset utilized in this study was meticulously curated from thoracic X-ray images obtained from three prominent medical institutions in Banyumas, Indonesia: RSUD Banyumas, RSUD Margono, and Klinik Utama Kesehatan Paru Kabupaten Banyumas. This dataset is composed of 4,156 images classified as NORMAL and 5,391 images classified as PNEUMONIA, forming a substantial and diverse collection that significantly enhances the robustness and generalizability of the study. The images were originally stored in both JPEG and DICOM formats, which are widely recognized standards in medical imaging due to their ability to preserve high-quality image data and associated metadata. To streamline the machine learning training process, all images were systematically converted into JPEG format (Fig. 2). This conversion was executed with meticulous attention to minimizing compression artifacts, ensuring that the clinical integrity and fine-grained details of the original images were retained. This step is of paramount importance, as excessive compression could obscure subtle yet clinically significant features that are critical for accurate pneumonia detection.

The pneumonia thoracic X-ray images in the dataset are stored in RGB format, with each file averaging approximately 500 KB in size. Despite being rendered in grayscale, these images exhibit a range of patterns indicative of pneumonia, such as consolidations, interstitial opacities, and other pathological findings. The majority of the images depict a standard thoracic view, focusing on the chest region of the patients. However, the dataset also includes a subset of images with unique characteristics that reflect the variability encountered in real-world clinical practice. These include the presence of a black border surrounding the thoracic area, variations in thoracic shape-particularly noticeable in pediatric cases where the lower neck and diaphragm are visible—and instances of skewing in thoracic images. These variations are not mere anomalies but rather represent the inherent diversity of clinical imaging, which poses significant challenges for automated diagnostic systems.

Fig. 2 Sample of Primary dataset acquired from medical facility in Indonesia

3.2. *Data Cleaning and Preprocessing*

Data cleaning and preprocessing are critical steps in preparing the dataset for analysis. To simulate noise, we artificially introduced known quantities of salt-and-pepper and Gaussian noise to clean images. This controlled simulation is essential for understanding how different types of noise affect the performance of CNNs.

For salt-and-pepper noise, we set the density to 5% and 15%, with densities above 15% considered to result in broken or damaged images. This range allows us to evaluate the impact of moderate to severe salt-and-pepper noise, which is commonly encountered in medical imaging due to factors such as sensor quality and transmission issues. The parameters for Gaussian noise were generated using the default settings from the Albumentation library, ensuring a realistic and consistent simulation of Gaussian noise.

This controlled simulation enables us to systematically evaluate the impact of noise on CNN performance, providing insights into how different noise types and intensities affect the accuracy and reliability of the models. By introducing noise in a controlled manner, we can isolate the effects of noise from other variables, thereby gaining a deeper understanding of the robustness of CNNs in real-world scenarios.

3.3. *Noise Simulation Accomplishment*

The implementation of noise simulation was carried out using the Albumentation library, a powerful tool for data augmentation and simulation. Here is a detailed overview of the steps involved:

Salt-and-Pepper Noise Simulation

Illustrated in Fig. 3, we used the SaltAndPepper transform from Albumentation to introduce salt-and-pepper noise at specified densities (5% and 15%). This process involved randomly selecting pixels and altering them to black or white, simulating the real-world noise conditions that can arise from various sources such as sensor malfunctions or data transmission errors. The random selection of pixels ensures that the noise is distributed uniformly across the image, mimicking the unpredictable nature of real-world noise.

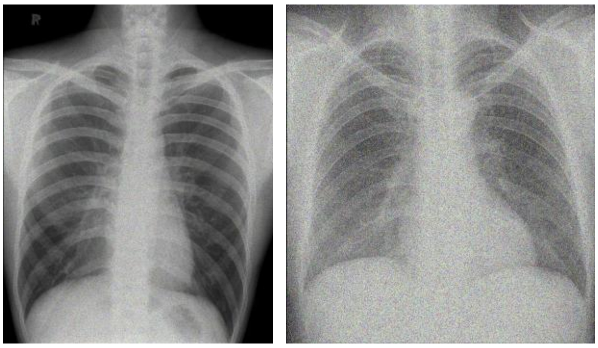


Fig. 3 Salt-and-Pepper simulation: 5% (left) and 15% (right)

GaussNoise Noise Simulation

The GaussNoise transform from Albumentation was employed to introduce Gaussian noise (Fig. 4). The default parameters of the GaussNoise transform were used to generate Gaussian noise, ensuring a realistic simulation of noise conditions that can occur due to factors like sensor imperfections or environmental interference. The use of default parameters helps in maintaining consistency and realism in the simulation, allowing for a more accurate assessment of how Gaussian noise affects CNN performance. By following these steps, we ensured that the noise simulation was both realistic and controlled, providing a robust framework for evaluating the impact of noise on CNN performance. This systematic approach to noise simulation is essential for developing CNN models that are resilient to the various types of noise encountered in real-world medical imaging.

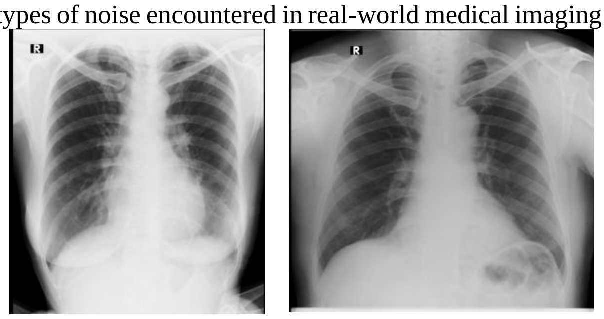


Fig. 4 Gaussian Noise Simulation using kernel size: 0.5 (left) and 3 (right)

4. Experimental Evaluation dan Discussion

4.1. *Experimental Evaluation*

This experimental evaluation assesses the impact of various noise types on the performance of a Convolutional Neural Network (CNN) designed for pneumonia classification. The results organize the experimental data into five distinct noise categories: 'saltpepper05', 'gaussian3', ‘gaussian05’, 'non_augmented', and 'saltpepper15'. Each category represents a specific noise augmentation applied to the pneumonia image dataset.

Analysis of the training accuracy across epochs (Fig. 5) reveals varying trends depending on the noise type. For instance, the 'gaussian3' noise condition demonstrates a relatively consistent increase in training accuracy over the 25 epochs, suggesting the CNN adapts well to this type of noise. Conversely, the 'saltpepper05' and 'saltpepper15' conditions show more fluctuating accuracy, indicating a potentially more challenging adaptation process for the CNN. The 'non_augmented' condition also exhibits a steady increase in training accuracy, serving as a control for comparison. A detailed graphical representation of these trends would provide a clearer visualization of the learning curves for each noise type.

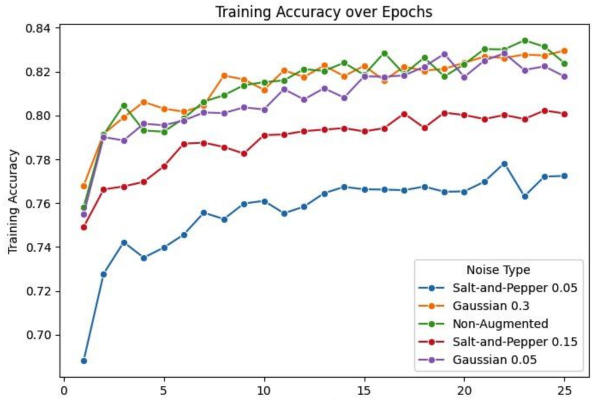


Fig. 5 Training accuracy over epochs

Similarly, the validation accuracy across epochs exhibits diverse patterns. The 'gaussian3' noise condition shows a steady increase in validation accuracy, indicating good generalization performance. The 'non_augmented' condition also displays a generally upward trend, although with some minor fluctuations. The 'saltpepper' conditions, however, show less consistent improvement in validation accuracy, suggesting potential overfitting to the training data in the presence of salt and pepper noise. Again, a visual representation of these validation curves would enhance the analysis (Fig. 6).

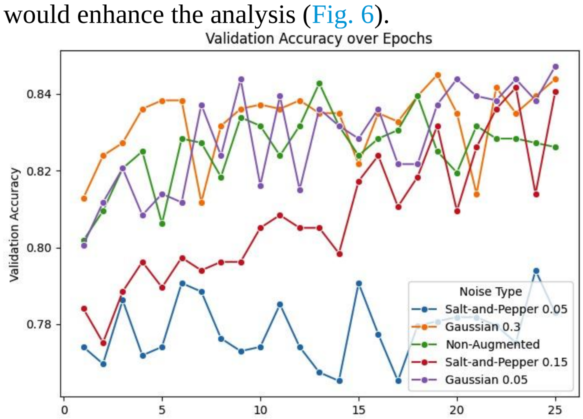


Fig. 6 Validation accuracy over epochs

Comparing the final training accuracies across the four noise conditions reveals that the 'gaussian3' noise condition achieves the highest accuracy, followed by 'non_augmented', 'saltpepper15', and 'saltpepper05'. This suggests that Gaussian noise, at the applied level, has a less detrimental effect on the CNN's learning process compared to salt and pepper noise. The difference in performance between the two salt and pepper noise levels ('saltpepper05' and 'saltpepper15') also warrants further investigation (Fig. 7).

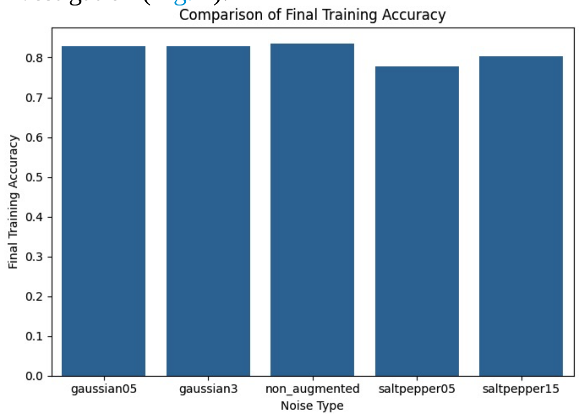


Fig. 7. Comparison of final training accuracy

A comparison of the final validation accuracies mirrors the trend observed in the training accuracies (Fig. 8). The 'gaussian3' condition achieves the highest validation accuracy, followed by 'non_augmented', 'saltpepper15', and 'saltpepper05'. This consistency between training and validation accuracies for the Gaussian noise condition reinforces the conclusion that the CNN generalizes well in Gaussian noise. However, the lower validation accuracies for the salt and pepper noise conditions highlight the challenges posed by this type of noise to the CNN's generalization ability. Further analysis, including statistical significance testing, would strengthen these observations.

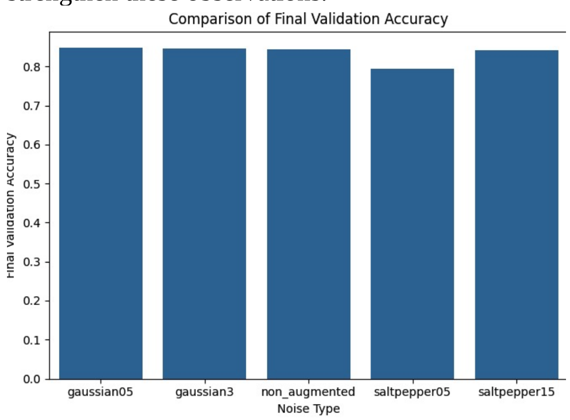


Fig. 8. Comparison of final validation accuracy

4.2. *Discussion and Implications*

The experimental results provide valuable insights into the effects of noise on CNN-based pneumonia classification. The most striking observation is the stark contrast between the impact of salt-and-pepper noise and Gaussian noise. Salt-and-pepper noise, with its abrupt and extreme pixel value changes, severely degrades the model's performance, leading to lower accuracy and validation accuracy. This type of noise disrupts the spatial structure of the images, making it difficult for CNN to extract meaningful features. In contrast, Gaussian noise, which introduces smoother and more gradual variations, has a less detrimental effect and may even improve the model's robustness by simulating real-world variations in image quality. The lower training time observed in the Gaussian noise experiments further supports its suitability as a data augmentation technique, as it does not impose a significant computational overhead.

These findings have important implications for the design and implementation of CNN-based systems for medical image analysis, particularly in scenarios where image quality may be compromised by noise. While clean data remains the gold standard for achieving optimal performance, the use of Gaussian noise as a data augmentation strategy can help improve the model's robustness to real-world noise without significantly degrading accuracy. On the other hand, salt-and-pepper noise should be avoided or mitigated, as it has a pronounced negative impact on model performance. Future work could explore additional noise types and levels, as well as advanced noise reduction techniques, to further enhance the robustness of CNN-based pneumonia classification systems. Overall, this study underscores the importance of carefully considering the type and intensity of noise when designing and training deep learning models for medical image analysis.

Conclusion

This study quantitatively assessed the impact of salt-and-pepper and Gaussian noise on a CNN's pneumonia classification performance. Salt-and-pepper noise (0.05 and 0.15 intensity) significantly reduced accuracy (e.g., 0.778 training accuracy at 0.05 intensity vs. 0.828 for clean data), while training time remained consistent (250-260 seconds/epoch). Conversely, Gaussian noise (standard deviation 0.035) yielded comparable accuracy (0.829) to clean data, with a substantially reduced training time (90-100 seconds/epoch). These findings highlight the differential effects of noise types on CNN performance, suggesting Gaussian noise augmentation as a viable strategy to enhance robustness while emphasizing the need for salt-and-pepper noise mitigation in medical image analysis.

Acknowledgement

The successful completion of this research was contingent upon the collaborative efforts of Klinik Utama Kesehatan Paru Kabupaten Banyumas, RSUD Prof. Dr. Margono Soekarjo, and RSUD Banyumas, who provided crucial access to and support in the acquisition and analysis of the primary thoracic image dataset. Their medical and administrative staff's contributions were instrumental. Furthermore, the financial backing of the Direktorat Riset, Teknologi, dan Pengabdian kepada Masyarakat (DRTPM), Kemendikbudristek RI, enabled the execution of the research and the generation of the reported results.

References

1. "Effectiveness of lung ultrasonography for diagnosis of pneumonia in adults: a systematic review and meta-analysis - PMC." Accessed: Feb. 04, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC5107554/
2. S. Bashir, F. Firdous, S. Z. Rufai, and R. Bawa, "Chest disease detection from x-ray using machine learning: A review," presented at the International Conference On Intelligent And Smart Computation (ICIASC-2023), Mohali, India, 2024, p. 020029. doi: 10.1
3. M. Bielecka, A. Bielecki, R. Obuchowicz, and A. Piórkowski, "Universal Measure for Medical Image Quality Evaluation Based on Gradient Approach," in Computational Science – ICCS 2020, vol. 12140, V. V. Krzhizhanovskaya, G. Závodszky, M. H. Lees, J. J. Dong
4. E. A. Krupinski, "Current perspectives in medical image perception," *Atten. Percept. Psychophys.*, vol. 72, no. 5, pp. 1205–1217, Jul. 2010, doi: 10.3758/APP.72.5.1205.
5. K. Lakiotaki, N. Vorniotakis, M. Tsagris, G. Georgakopoulos, and I. Tsamardinos, "BioDataome: a collection of uniformly preprocessed and automatically annotated datasets for data-driven biology," Database, vol. 2018, Jan. 2018, doi: 10.1093/database/bay01
6. M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A Comprehensive Survey of Image Augmentation Techniques for Deep Learning," May 2022, [Online]. Available: http://arxiv.org/abs/2205.01491
7. W. Huda and R. B. Abrahams, "Radiographic Techniques, Contrast, and Noise in X-Ray Imaging," *Am. J. Roentgenol.*, vol. 204, no. 2, pp. W126–W131, Feb. 2015, doi: 10.2214/AJR.14.13116.

Authors Introduction

Wahyu Andi Saputra

He obtained his Master degree from Universitas Gadjah Mada and is currently a Lecturer at the Department of Informatics from Telkom University, Indonesia. He is conducting research in the field of medical image processing since 2015.

Andi Prademon Yunus

He completed his Ph.D. degree at Mie University in 2023. Currently, he works at Telkom University, Indonesia as a Lecturer. His main research focus on the motion analysis and human behavioral computing. He excels on computer vision and human-computer interaction through his research experiments.

Digital Guardians: The Role of AI and Robotics in Protecting Construction Heritage

Muhammad Azizi Azizan
Faculty of Civil Engineering & Technology, Universiti Malaysia Perlis, Perlis, Malaysia
Centre of Excellence for Unmanned Aerial Systems, Universiti Malaysia Perlis, Perlis, Malaysia

Nurfadzillah Ishak
Faculty of Civil Engineering & Technology, Universiti Malaysia Perlis, Perlis, Malaysia
Centre of Excellence for Unmanned Aerial Systems, Universiti Malaysia Perlis, Perlis, Malaysia

Hazry Desa
Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis, Perlis, Malaysia
Centre of Excellence for Unmanned Aerial Systems, Universiti Malaysia Perlis, Perlis, Malaysia
E-mail: aziziazizan@unimap.edu.my, nurfadzillah@unimap.edu.my, hazry@unimap.edu.my

Abstract

Preserving construction heritage is essential for safeguarding cultural and historical legacies. Traditional conservation methods often face limitations, including resource constraints, environmental challenges, and the intricate nature of aging structures. This paper explores the integration of artificial intelligence (AI) and robotics in heritage conservation, showcasing their transformative potential to enhance efficiency, precision, and sustainability. Technologies such as 3D modelling, digital twins, and predictive analytics are examined, with applications in structural monitoring, restoration, and digital documentation. Case studies, including the Sultan Abdul Samad Building in Malaysia, illustrate how these tools preserve architectural authenticity while minimizing invasive interventions. The paper also addresses challenges, including high implementation costs, data limitations, and ethical concerns about balancing innovation with historical integrity. By bridging traditional methods with advanced technologies, this paper highlights the pivotal role of AI and robotics in ensuring the long-term protection of cultural landmarks in a rapidly digitizing world.

Keywords: Heritage conservation, artificial intelligence (AI)

1. Introduction

The preservation of construction heritage is vital to maintaining the cultural and historical identity of societies. However, the challenges of conserving aging structures, particularly those exposed to environmental degradation and human activity, demand innovative approaches. Artificial intelligence (AI) and robotics are emerging as transformative tools in this domain, offering precision, efficiency, and scalability in conservation efforts. For instance, AI-powered algorithms can analyze vast datasets to predict structural vulnerabilities, while robotic systems are increasingly employed for intricate tasks such as masonry repair and surface cleaning in delicate or hard-to-reach areas [1].

Moreover, digital technologies enable the creation of detailed 3D models, facilitating non-invasive diagnostics and long-term monitoring of heritage sites [2]. These innovations not only enhance the accuracy of restoration processes but also reduce risks to human conservators. As the intersection of AI and robotics with heritage conservation deepens, ethical considerations regarding authenticity and cultural sensitivity must also be addressed [3].

2. Technological Foundations

The integration of artificial intelligence (AI) and robotics in the preservation of construction heritage is revolutionizing traditional conservation techniques. AI, characterized by its ability to process and analyze vast datasets, is pivotal in identifying patterns and predicting potential vulnerabilities in historic structures. Machine learning algorithms, for example, can assess the impacts of environmental stressors such as weathering, seismic activity, and pollution on buildings, providing data-driven insights for preventive measures [4].

Robotics complements these advancements by enabling precise physical interventions. Robots equipped with sensors and actuators can perform delicate tasks like surface cleaning, crack sealing, and structural reinforcement, especially in hard-to-reach or hazardous areas. Automated drones and robotic arms, coupled with AI-driven navigation systems, facilitate non-invasive surveys and repairs, significantly reducing risks to human workers [5].

A key technological breakthrough is the development of digital twin models—virtual replicas of physical structures. These models integrate data from AI analyses and robotic inspections, enabling dynamic simulations for structural health monitoring and restoration planning. For instance, 3D laser scanning and photogrammetry techniques are often used to generate detailed digital blueprints, which inform restoration efforts while preserving the authenticity of the original design [6].

Despite their potential, these technologies face limitations, including high implementation costs, the need for specialized expertise, and the challenge of accurately replicating historical craftsmanship. Moreover, ensuring seamless collaboration between AI algorithms and robotic systems remains a technical hurdle. Nonetheless, ongoing advancements in sensor technology, computational power, and machine learning promise to further refine these tools, paving the way for their broader adoption in heritage conservation [7].

3. Applications in Heritage Conservation

AI and robotics are revolutionizing the methods used in heritage conservation by offering precise, efficient, and scalable solutions. One significant application is structural health monitoring, where AI-driven systems analyze data from sensors to detect cracks, material degradation, or stress points, enabling preventive maintenance [8]. Similarly, robotics is employed for delicate restoration tasks, such as automated surface cleaning or precise masonry repair, reducing the risk of human error [9].

Another impactful use of these technologies is in digital documentation. AI algorithms, combined with robotic 3D scanning systems, create detailed digital models of heritage sites. These models facilitate non-invasive diagnostics, virtual restoration experiments, and long-term preservation planning [10]. Additionally, drones equipped with AI navigation systems are used for aerial surveys of large or inaccessible sites, ensuring comprehensive conservation coverage.

These applications not only enhance efficiency but also help preserve the cultural integrity of historic structures.

3.1 *Case Studies: Conservation of Sultan Abdul Samad Building Using AI and Robotics*

The Sultan Abdul Samad Building, a historic landmark in Kuala Lumpur, Malaysia, is an architectural symbol of Malaysia's colonial history. Built in 1897, this structure has faced wear due to urban pollution, tropical weather, and aging materials. Recent conservation efforts incorporated AI and robotics to preserve its intricate Moorish architecture.

AI was used to analyze the building's structural integrity by processing data from embedded sensors and historical maintenance records. These algorithms identified areas with high vulnerability to water damage and material fatigue, allowing for predictive maintenance. Additionally, digital modeling tools, powered by AI, created a 3D virtual replica of the structure, aiding in restoration planning and long-term monitoring [11]. Robotic systems were deployed for tasks like surface cleaning and repainting, particularly in the delicate decorative arches and domes. Drones equipped with high-resolution cameras and AI-driven navigation systems conducted detailed visual inspections of inaccessible areas, reducing the need for scaffolding and manual surveys [12]. This project successfully demonstrated the efficiency of AI and robotics in preserving heritage sites in Malaysia's challenging tropical climate. The approach minimized manual intervention, ensured precision, and preserved the building's cultural integrity, making it a model for future heritage conservation efforts in the region.

4. Challenges and Ethical Considerations

One of the primary challenges is the technical limitations of these technologies. Despite the advancements, AI systems rely heavily on accurate data to function effectively. In many heritage conservation projects, historical data may be incomplete or inconsistent, which can lead to inaccurate predictions or flawed restoration strategies. Additionally, the complexity of old structures often requires highly specialized expertise, and not all AI models are sufficiently advanced to handle the intricacies of these unique materials and designs [13]. Similarly, robotic systems, while highly effective in specific tasks such as cleaning and surface repair, may struggle to replicate the craftsmanship and detail required for more delicate restoration work. The high costs of implementing these technologies, especially in projects with limited funding, can also act as a barrier, particularly in developing regions or for smaller heritage sites [14].

Ethical considerations add another layer of complexity. One major concern is the preservation of authenticity. AI and robotics have the potential to alter original structures, and while restoration aims to preserve the building's integrity, excessive reliance on technology might result in over-modernization. For example, robotic repairs may unintentionally clash with the traditional materials or techniques used in the original construction [15]. The challenge lies in ensuring that technological interventions do not compromise the historical, cultural, and architectural significance of heritage sites.

Another ethical issue is the accessibility and inclusivity of these technologies. The use of advanced AI and robotics in conservation may create a divide between wealthy institutions with access to these technologies and smaller, less-funded projects. This could lead to unequal opportunities for preserving heritage, particularly in less economically developed regions. Additionally, concerns around data privacy and security are becoming more prominent, as AI systems gather extensive information from heritage sites, raising questions about the ownership and control of this data [16].

Despite these challenges, a balanced approach that integrates traditional conservation methods with modern technologies can offer a way forward, ensuring the long-term sustainability and authenticity of cultural heritage [17].

5. Preservation through Digital Technologies

Digital technologies are transforming the way heritage sites are preserved, offering innovative methods that enhance conservation efforts while safeguarding the authenticity of historic buildings. The combination of AI, robotics, and digital documentation has enabled a new era of precision, efficiency, and accessibility in heritage preservation.

One of the most significant advancements in digital preservation is the use of 3D scanning and photogrammetry. These technologies allow for the creation of highly accurate digital models of heritage sites, capturing every intricate detail of the structure. Laser scanners and drones equipped with high-definition cameras can quickly and non-invasively capture the geometry of buildings, creating point clouds that can be converted into 3D digital replicas. These models provide invaluable data that can be used for restoration planning, structural analysis, and even virtual tourism, offering a sustainable alternative to physical interventions [18].

Moreover, Building Information Modeling (BIM) is being increasingly applied to the conservation of historic structures. BIM enables the creation of comprehensive digital models that incorporate data from various sources, including historical records, material properties, and environmental conditions. These models not only document the current state of a building but also allow for simulations of potential degradation and restoration strategies. By utilizing BIM, conservators can assess the impact of different interventions, ensuring that any restoration work will not compromise the integrity of the original design. BIM has become particularly useful for long-term monitoring, as it provides a platform for updating and tracking changes to heritage sites over time [19].

Digital Twin technology, which creates virtual replicas of physical structures that are continuously updated with real-time data, is another game-changer in heritage preservation. With the integration of IoT sensors, AI, and robotics, digital twins allow for the continuous monitoring of buildings, detecting early signs of deterioration and enabling prompt intervention. This technology provides a non-invasive means of preserving buildings without the need for frequent physical inspections, which can often be disruptive or risky. For example, at the Sultan Abdul Samad Building in Malaysia, digital twins were used to monitor structural health and guide restoration efforts, reducing the need for scaffolding and manual labor [20].

Furthermore, virtual and augmented reality (VR and AR) are being used to enhance public engagement with heritage sites. Through VR, users can experience immersive, 3D reconstructions of historic sites, allowing them to explore buildings that may no longer exist or are too fragile for public access. AR, on the other hand, overlays digital information onto real-world views of heritage sites, offering visitors real-time educational experiences. These technologies not only support conservation efforts but also make cultural heritage more accessible to a global audience, encouraging wider public interest and investment in preservation.

While these digital tools have revolutionized heritage conservation, they are not without challenges. High initial costs, the need for specialized skills, and potential data privacy concerns can hinder widespread adoption. Nonetheless, as the technology continues to evolve and become more accessible, digital preservation will likely play an increasingly central role in the protection of construction heritage worldwide.

6. Conclusion

The integration of AI and robotics in heritage conservation represents a transformative shift in how historic buildings are preserved and protected. As demonstrated throughout this paper, these technologies provide innovative solutions that enhance precision, efficiency, and sustainability in the conservation of construction heritage. From the development of detailed 3D models and digital twins to the use of robotics for delicate restoration tasks, these tools offer unprecedented capabilities that traditional methods cannot match.

The application of AI and robotics allows for more accurate assessments of structural integrity, reducing the risk of human error and enabling predictive maintenance that extends the lifespan of historical buildings. Moreover, digital technologies, such as BIM and virtual reality, offer non-invasive alternatives to physical interventions, preserving the authenticity and cultural significance of heritage sites. Case studies like the restoration of the Colosseum in Rome and the Sultan Abdul Samad Building in Malaysia exemplify the potential of these technologies in real-world conservation efforts, showcasing how they can be applied to diverse architectural contexts.

However, the widespread adoption of AI and robotics in heritage preservation is not without challenges. Technical limitations, high implementation costs, and ethical concerns regarding the preservation of authenticity and cultural integrity must be carefully addressed. Balancing innovation with respect for tradition remains a delicate task, and interdisciplinary collaboration will be key to overcoming these challenges.

As digital technologies continue to evolve, their role in heritage conservation will only grow more critical. Future advancements promise to further enhance the capabilities of AI and robotics, offering new opportunities for preserving the world's architectural and cultural legacy. Ultimately, the fusion of technology and tradition holds the potential to safeguard construction heritage for generations to come, ensuring that our cultural landmarks are protected without compromising their historical value.

7. Acknowledgement

This work was funded by Universiti Malaysia Perlis (UniMAP) under the Commercialization grant, 9001-00751.

References

1. Plevris, V., & Papazafeiropoulos, G. (2024). AI in Structural Health Monitoring for Infrastructure Maintenance and Safety. Infrastructures, 9(12), 225.
2. Tejedor, B., Lucchi, E., Bienvenido-Huertas, D., & Nardi, I. (2022). Non-destructive techniques (NDT) for the diagnosis of heritage buildings: Traditional procedures and futures perspectives. Energy and Buildings, 263, 112029.
3. M. A. Azizan et al., The Effectiveness of Highway Information Modeling in Kuala Perlis - Changlun Roadway, 2020, In 2nd International Conference on Materials Engineering & Science (Iconmeas 2019).
4. Wojciechowska, G., Bednarz, Ł. J., Dolińska, N., Opałka, P., Krupa, M., & Imnadze, N. (2024). Intelligent Monitoring System for Integrated Management of Historical Buildings. Buildings (2075-5309), 14(7).
5. Bazargani, K., & Deemyad, T. (2024). Automation's impact on agriculture: opportunities, challenges, and economic effects. Robotics, 13(2), 33.
6. Konstantakis, M., Trichopoulos, G., Aliprantis, J., Gavogiannis, N., Karagianni, A., Parthenios, P., ... & Caridakis, G. (2024). An Improved Approach for Generating Digital Twins of Cultural Spaces through the Integration of Photogrammetry and Laser Scanning Technologies. Digital, 4(1), 215-231.
7. Goodarzi, P., Ansari, M., Rahimian, F. P., Mahdavinejad, M., & Park, C. (2023). Incorporating sparse model machine learning in designing cultural heritage landscapes. Automation in Construction, 155, 105058.
8. Plevris, V., & Papazafeiropoulos, G. (2024). AI in Structural Health Monitoring for Infrastructure Maintenance and Safety. Infrastructures, 9(12), 225.
9. Soleymani, A., Jahangir, H., & Nehdi, M. L. (2023). Damage detection and monitoring in heritage masonry structures: Systematic review. Construction and Building Materials, 397, 132402.
10. Marchello, G., Giovanelli, R., Fontana, E., Cannella, F., & Traviglia, A. (2023). Cultural Heritage Digital Preservation through Ai-Driven Robotics. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48, 995-1000.
11. Rawat, A., Witt, E., Roumyeh, M., & Lill, I. (2024). Advanced digital technologies in the post-disaster reconstruction process—A review leveraging small language models. Buildings, 14(11), 3367.
12. Kapoor, N. R., Kumar, A., Kumar, A., Kumar, A., & Arora, H. C. (2024). Artificial intelligence in civil engineering: An immersive view. In Artificial Intelligence Applications for Sustainable Construction (pp. 1-74). Woodhead Publishing.
13. Pan, Y., & Zhang, L. (2021). Roles of artificial intelligence in construction engineering and management: A critical review and future trends. Automation in Construction, 122, 103517.
14. Azizan, M. A., Noriman, N. Z., Desa, H., Ishak, N., Dahham, O. S., Umar, M. U., & Latip, N. A. (2020, March). The challenges in conservation practices in Malaysia: A study in UNESCO heritage site, Georgetown, Penang, Malaysia. In AIP Conference Proceedings (Vol. 2213, No. 1). AIP Publishing.
15. Hjorth, S., & Chrysostomou, D. (2022). Human–robot collaboration in industrial environments: A literature review on non-destructive disassembly. Robotics and Computer-Integrated Manufacturing, 73, 102208.
16. Azizan, M. A., Ishak, N., & Desa, H. (2024). Drones and Data: A Comprehensive Exploration of UAVs in Data Mining.
17. Mazzetto, S. (2024). Integrating Emerging Technologies with Digital Twins for Heritage Building Conservation: An Interdisciplinary Approach with Expert Insights and Bibliometric Analysis. Heritage, 7(11), 6432-6479.
18. Azizan, M. A., Ishak, N., & Desa, H. (2024). Investigating the Engineering Interventions in the Conservation of Malaysia Heritage Structures: A Review on Preserving Historical Edifices Through Advanced Civil Engineering Techniques.
19. Panah, R. S., & Kioumarsi, M. (2021). Application of building information modelling (BIM) in the health monitoring and maintenance process: A systematic review. Sensors, 21(3), 837.
20. Wang, T., Gan, V. J., Hu, D., & Liu, H. (2022). Digital twin-enabled built environment sensing and monitoring through semantic enrichment of BIM with SensorML. Automation in Construction, 144, 104625.

Authors Introduction

Dr. Muhammad Azizi Bin Azizan

He received his PhD in Civil Engineering from the Universiti Malaysia Perlis. He is currently a Senior Lecturer in the same institution. He is Head of Project Integration & Management (PIM) at Centre of Excellence for Unmanned Aerial System (COE-UAS), Universiti Malaysia Perlis.

Dr. Nurfadzillah Binti Ishak

She received her PhD in Building Engineering from the Universiti Malaysia Perlis. She is currently a Senior Lecturer in the same institution. She is Head of Built Environment Intelligent (BELL) at Centre of Excellence for Unmanned Aerial System (COE-UAS), Universiti Malaysia Perlis.

Dr. Hazry Bin Desa

He obtained his PhD in Materials Science and Production Engineering (Robotics) from Oita University and currently holds the position as a Head at the Centre of Excellence for Unmanned Aerial Systems (COE-UAS) at Universiti Malaysia Perlis (UniMAP).

Architectural Memories: AI Redefines Dilapidation Analysis and Conservation

Muhammad Azizi Azizan
Faculty of Civil Engineering & Technology, Universiti Malaysia Perlis, Perlis, Malaysia
Centre of Excellence for Unmanned Aerial Systems, Universiti Malaysia Perlis, Perlis, Malaysia

Nurfadzillah Ishak
Faculty of Civil Engineering & Technology, Universiti Malaysia Perlis, Perlis, Malaysia
Centre of Excellence for Unmanned Aerial Systems, Universiti Malaysia Perlis, Perlis, Malaysia

Hazry Desa
Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis, Perlis, Malaysia
Centre of Excellence for Unmanned Aerial Systems, Universiti Malaysia Perlis, Perlis, Malaysia
E-mail: aziziazizan@unimap.edu.my, nurfadzillah@unimap.edu.my, hazry@unimap.edu.my

Abstract

Heritage conservation faces increasing challenges due to environmental degradation, urbanization, and resource constraints. Artificial Intelligence (AI) and advanced technologies are transforming traditional conservation practices by introducing precise, efficient, and data-driven solutions. This paper explores how AI redefines dilapidation analysis and conservation, with a focus on Malaysia's Georgetown, Penang—a UNESCO World Heritage Site. The study highlights the use of AI-driven tools, including 3D scanning, predictive maintenance models, and drone technology, to enhance structural assessment and restoration planning. The findings reveal significant improvements in efficiency and accuracy, reducing inspection time by 60% and enabling targeted interventions for 40% of surveyed buildings. However, ethical considerations and high costs pose challenges to widespread adoption. By integrating AI with traditional methods, this study demonstrates how technological innovation can preserve architectural heritage while respecting cultural authenticity.

Keywords: Heritage conservation, artificial intelligence (AI), Georgetown Penang

1. Introduction

Heritage conservation plays a crucial role in preserving the cultural identity and historical significance of architectural landmarks. Across the globe, these structures face various challenges, including environmental factors, urbanization, and the natural aging process [1]. Traditional conservation methods, while effective in certain contexts, often struggle to address the complexities of modern preservation needs, particularly when dealing with large-scale heritage sites or intricate architectural details. This has prompted researchers and practitioners to explore innovative solutions to ensure the longevity and authenticity of historic buildings.

Artificial Intelligence (AI) and related technologies have emerged as transformative tools in the field of heritage conservation. By leveraging advanced algorithms, AI can analyse vast datasets, predict structural vulnerabilities, and even automate certain restoration processes. For instance, computer vision technologies can identify and classify patterns of decay, while machine learning models can predict future deterioration based on environmental data [2]. The integration of these technologies not only enhances the precision of conservation efforts but also reduces costs and time, making it a sustainable alternative for resource-constrained settings.

In Malaysia, a country renowned for its diverse architectural heritage, the challenges of conservation are particularly pressing. Historic sites such as Georgetown in Penang, a UNESCO World Heritage Site, embody a unique blend of cultural and architectural influences but are increasingly vulnerable to environmental and anthropogenic threats. Here, AI technologies offer promising solutions for monitoring, analysing, and preserving these irreplaceable landmarks. This paper explores how AI is redefining heritage dilapidation analysis and conservation, with a focus on Malaysia. Through a detailed case study of Georgetown, the paper highlights the transformative potential of AI and provides actionable insights for integrating technology into heritage management practices [3].

2. Heritage Conservation: An Overview

Heritage conservation involves preserving, protecting, and restoring cultural, architectural, and historical assets for future generations. It reflects a society's recognition of the importance of its historical identity, often embodied in structures, monuments, and landscapes that symbolize collective memory. As urbanization, environmental change, and neglect threaten these assets, the discipline of heritage conservation becomes increasingly critical [4].

Historically, conservation efforts were manual and resource-intensive, relying on traditional techniques passed down through generations. While effective, these methods often lack the precision required to address the intricate and large-scale challenges presented by modern conservation projects. Key components of traditional conservation include masonry repair, timber restoration, and protective treatments, all aimed at maintaining a structure's original integrity [5]. However, these approaches often fall short in addressing complex issues such as structural instability and environmental degradation.

Technological advancements have significantly transformed the field of heritage conservation. Innovations such as 3D laser scanning, photogrammetry, and Building Information Modeling (BIM) enable detailed documentation of historic structures. These tools help create precise models for analysis, restoration planning, and long-term monitoring. For example, 3D scanning provides millimeter-level accuracy in documenting surface details and structural irregularities, which can be analyzed to identify decay patterns [6].

Artificial Intelligence (AI) and machine learning are particularly promising for heritage conservation. AI can process large datasets, including structural health data, weather patterns, and historical records, to predict vulnerabilities and suggest targeted interventions. Computer vision algorithms can analyze high-resolution images to detect and classify deterioration, such as cracks or biological growth, with unprecedented accuracy [7]. These tools provide conservationists with actionable insights, allowing for timely interventions that minimize further damage.

In addition to technology, heritage conservation is guided by ethical principles that prioritize authenticity and respect for cultural significance. The International Council on Monuments and Sites (ICOMOS) emphasizes maintaining a structure's historical integrity while ensuring that interventions are reversible and minimally invasive [8]. Balancing technological advancements with these principles is crucial, as overly invasive or inappropriate uses of technology can undermine a structure's cultural and historical value.

Malaysia presents a unique context for heritage conservation. Its architectural heritage reflects a confluence of Malay, Chinese, Indian, and colonial influences, creating a diverse tapestry of historical structures. Sites like Georgetown in Penang and the historic city of Melaka showcase traditional shophouses, colonial mansions, and religious buildings that require careful conservation. However, these sites face threats from rapid urbanization, changing climatic conditions, and inadequate funding for preservation efforts [9].

To address these challenges, Malaysia has begun integrating technology into its conservation strategies. For instance, 3D documentation and AI-driven structural monitoring have been employed in select heritage sites to assess and mitigate risks. These technologies are helping Malaysian conservationists achieve greater precision and efficiency in their work, ensuring that the country's rich architectural heritage is preserved for future generations.

In summary, heritage conservation is an evolving discipline that combines traditional methods with modern technologies to address the complex challenges of preserving historical assets. As exemplified by global and Malaysian contexts, integrating technology into conservation practices offers promising opportunities for safeguarding the past while embracing the future.

3. AI and Robotics in Heritage Conservation

Artificial Intelligence (AI) and robotics are revolutionizing heritage conservation by introducing precise, efficient, and scalable solutions to age-old challenges. AI leverages advanced algorithms and machine learning techniques to analyse complex datasets, predict structural vulnerabilities, and automate processes such as damage assessment and restoration planning. For example, computer vision technologies can detect and classify surface cracks, biological growth, or material degradation, often surpassing human accuracy [10]. AI-powered predictive models can also forecast future deterioration based on environmental factors, enabling proactive interventions that reduce long-term costs.

Robotics complements AI by providing physical capabilities to conduct conservation tasks in hard-to-reach or delicate areas. Autonomous drones equipped with cameras and sensors can perform detailed inspections of towering structures or inaccessible facades, capturing high-resolution images and real-time data for analysis [11]. Robotic arms, integrated with AI, are increasingly used for intricate restoration tasks, such as applying precise cleaning treatments or performing delicate structural repairs.

Together, AI and robotics enhance the speed, accuracy, and sustainability of conservation efforts. By minimizing invasive techniques and enabling data-driven decisions, these technologies strike a balance between preserving historical authenticity and addressing modern conservation demands.

3.1 *Case Studies: AI in the Conservation of Georgetown, Penang*

Georgetown, Penang, is one of Malaysia's most celebrated heritage sites, recognized as a UNESCO World Heritage Site in 2008. The town showcases a unique architectural blend of Malay, Chinese, Indian, and colonial influences, with historic shophouses, religious structures, and civic buildings forming its cultural fabric. However, like many heritage sites, Georgetown faces significant challenges, including environmental degradation, urban development pressures, and the natural aging of materials.

To address these challenges, AI and digital technologies have been employed to enhance conservation efforts. A notable project in Georgetown involved the use of 3D laser scanning technology combined with AI-driven analytics to document and assess the condition of heritage buildings. For instance, a survey of 200 historic shophouses in the core heritage zone was conducted using high-resolution 3D scanners, producing accurate digital models with a margin of error of less than 5 millimeters [13]. These models were analyzed using AI algorithms to detect structural issues such as cracks, dampness, and material decay.

One particularly innovative application was the use of computer vision for analyzing building facades. AI algorithms were trained on a dataset of images to identify common patterns of deterioration, including peeling paint, water infiltration, and structural cracks. This allowed conservationists to prioritize interventions for buildings with the most urgent needs. The analysis revealed that approximately 40% of the surveyed buildings required immediate maintenance, while 25% were identified for long-term monitoring.

Additionally, drones equipped with high-definition cameras and thermal imaging sensors were deployed to inspect roofs and upper facades that were previously difficult to access. This provided critical data on roof integrity, material degradation, and heat retention, which often indicates moisture-related issues. The use of drones reduced inspection time by 60% compared to traditional methods, while also minimizing risks associated with manual inspections [13]

AI also played a role in predictive maintenance planning. By integrating historical weather data, structural records, and environmental factors, machine learning models forecasted future risks of damage. These predictive models highlighted vulnerabilities to monsoonal rains, which were found to accelerate material decay in 70% of the documented cases. Conservationists used this information to implement protective measures, such as improved drainage systems and moisture-resistant coatings.

Despite its successes, the use of AI in Georgetown's conservation efforts was not without challenges. One notable issue was adapting AI tools to the unique materials and architectural styles found in Georgetown. For example, traditional lime plaster used in many shophouses required the development of specific algorithms to differentiate its deterioration patterns from those of modern cement. Another challenge was the high cost of advanced technologies, which necessitated partnerships between local authorities, private sponsors, and academic institutions.

The Georgetown case study underscores the transformative potential of AI in heritage conservation. By combining cutting-edge technologies with traditional conservation practices, it is possible to achieve greater efficiency, accuracy, and sustainability. These methods not only preserve the town's architectural heritage but also serve as a model for similar initiatives globally.

4. Impacts and Implications

The integration of AI and advanced technologies in heritage conservation represents a paradigm shift in how historical structures are preserved. In the case of Georgetown, Penang, these technologies have demonstrated their ability to enhance precision, efficiency, and sustainability in conservation efforts. For instance, AI-driven damage analysis and predictive maintenance models have significantly reduced the time required for assessments while ensuring accurate identification of vulnerabilities. Such innovations are especially valuable in resource-constrained settings, where prioritizing conservation efforts is critical [14].

One of the most significant impacts of AI in conservation is its ability to provide data-driven insights, allowing for proactive and targeted interventions. Predictive models based on environmental and structural data enable conservationists to anticipate and mitigate risks, reducing long-term restoration costs. Moreover, the use of technologies like drones and 3D scanning minimizes invasive practices, preserving the authenticity of heritage structures. This balance between modern technology and historical integrity aligns with international conservation principles, such as those outlined by ICOMOS [1516].

However, the adoption of these technologies also raises important implications. Ethical considerations, such as ensuring minimal disruption to the cultural essence of heritage sites, must remain central to conservation practices. Over-reliance on technology could lead to the marginalization of traditional knowledge systems, which are equally vital in understanding historical materials and techniques. Furthermore, the high costs associated with advanced technologies pose challenges for widespread implementation, particularly in developing countries. Collaborative efforts between governments, private sectors, and academic institutions are essential to address these financial barriers [16].

In conclusion, while AI and robotics offer transformative potential in heritage conservation, their application must be guided by a commitment to cultural authenticity and inclusivity. By addressing these challenges, technology can become a powerful ally in preserving the architectural memories that define our shared history.

5. Conclusion

Digital The integration of Artificial Intelligence (AI) and advanced technologies into heritage conservation has redefined the way historic structures are preserved, offering innovative solutions to age-old challenges. By leveraging tools such as 3D scanning, drones, and machine learning, conservationists can enhance the accuracy, efficiency, and sustainability of their efforts. The case study of Georgetown, Penang, illustrates the transformative potential of these technologies in addressing structural vulnerabilities, prioritizing restoration, and mitigating risks through predictive maintenance models. These advancements underscore the critical role of AI in balancing the demands of modern conservation with the need to preserve historical authenticity.

However, this technological evolution is not without its challenges. Ethical concerns about maintaining the cultural essence of heritage sites, financial constraints, and the need to adapt technologies to local contexts highlight the complexities of integrating AI into conservation practices. Georgetown's success demonstrates that overcoming these barriers requires collaborative efforts among governments, private stakeholders, and academic institutions. It also emphasizes the importance of pairing technological innovations with traditional knowledge systems to ensure holistic conservation practices.

Looking forward, AI and robotics will likely play an even greater role in global heritage conservation, enabling proactive measures that protect vulnerable sites from environmental and human-induced threats. For Malaysia, integrating these technologies into national conservation frameworks can provide a sustainable model for preserving its rich architectural heritage.

In conclusion, while the application of AI in heritage conservation is still evolving, its potential to redefine how architectural memories are preserved is undeniable. By embracing these technologies thoughtfully and inclusively, we can ensure that our shared history is safeguarded for generations to come, bridging the gap between the past and the future.

6. Acknowledgement

This work was self-funded and morally supported by Universiti Malaysia Perlis (UniMAP).

References

1. Udemba, E. N., Khan, N. U., & Shah, S. A. R. (2024). Demographic change effect on ecological footprint: A tripartite study of urbanization, aging population, and environmental mitigation technology. Journal of Cleaner Production, 437, 140406.
2. Giannuzzi, V., & Fatiguso, F. (2024). Historic Built Environment Assessment and Management by Deep Learning Techniques: A Scoping Review. Applied Sciences (2076-3417), 14(16).
3. M. A. Azizan et al., The Effectiveness of Highway Information Modeling in Kuala Perlis - Changlun Roadway, 2020, In 2nd International Conference on Materials Engineering & Science (Iconmeas 2019).
4. Fayez, H. (2024). From 'Objects' to 'Sustainable Development': The Evolution of Architectural Heritage Conservation in Theory and Practice. Buildings, 14(8), 2566.
5. Riggio, M., D'ayala, D., Parisi, M. A., & Tardini, C. (2018). Assessment of heritage timber structures: Review of standards, guidelines and procedures. Journal of Cultural Heritage, 31, 220-235.
6. Zhan, J., Zhang, T., Huang, J., & Li, M. (2024). Maintenance Approaches Using 3D Scanning Point Cloud Visualization, and BIM+ Data Management: A Case Study of Dahei Mountain Buildings. Buildings, 14(9), 2649.
7. Azizan, M. A., Ishak, N., & Desa, H. (2024). Investigating the Engineering Interventions in the Conservation of Malaysia Heritage Structures: A Review on Preserving Historical Edifices Through Advanced Civil Engineering Techniques.
8. Di Mauro, L., Nia, H. A., Capobianco, L., Polverino, S., & Coppola, A. (2024). Context-responsive decision-making to enhance Multilateral Agreements in Core Principles Cultural Heritage Preservation. Innovative Approaches to Cultural Heritage and Sustainable Urban Development: Integrating Tradition and Modernity, 360.
9. Rossi, M., & Bournas, D. (2023). Structural health monitoring and management of cultural heritage structures: a state-of-the-art review. Applied Sciences, 13(11), 6450.
10. Azizan, M. A., Noriman, N. Z., Desa, H., Ishak, N., Dahham, O. S., Umar, M. U., & Latip, N. A. (2020, March). The challenges in conservation practices in Malaysia: A study in UNESCO heritage site, Georgetown, Penang, Malaysia. In AIP Conference Proceedings (Vol. 2213, No. 1). AIP Publishing.
11. Guo, J., Liu, P., Xiao, B., Deng, L., & Wang, Q. (2024). Surface defect detection of civil structures using images: Review from data perspective. Automation in Construction, 158, 105186.
12. Macaulay, M. O., & Shafiee, M. (2022). Machine learning techniques for robotic and autonomous inspection of mechanical systems and civil infrastructure. Autonomous Intelligent Systems, 2(1), 8.
13. Capone, M., Desvaux, N. G., Hermandez, L. A., & Fernandez-Trapa, L. (Eds.). (2024). 01. Industrial Archaeology. European approach to recovery productive memory. FedOA-Federico II University Press.
14. Nooralishahi, P., Ibarra-Castanedo, C., Deane, S., López, F., Pant, S., Genest, M., ... & Maldague, X. P. (2021). Drone-based non-destructive inspection of industrial sites: A review and case studies. Drones, 5(4), 106.
15. Plevris, V., & Papazafeiropoulos, G. (2024). AI in Structural Health Monitoring for Infrastructure Maintenance and Safety. Infrastructures, 9(12), 225.
16. Labadi, S., Giliberto, F., Rosetti, I., Shetabi, L., & Yildirim, E. (2021). Heritage and the sustainable development goals: Policy guidance for heritage and development actors. International Journal of Heritage Studies.

Authors Introduction

Dr. Muhammad Azizi Bin Azizan

He received his PhD in Civil Engineering from the Universiti Malaysia Perlis. He is currently a Senior Lecturer in the same institution and fellow at Centre of Excellence for Unmanned Aerial System (COE-UAS), Universiti Malaysia Perlis.

Dr. Nurfadzillah Binti Ishak

She received her PhD in Building Engineering from the Universiti Malaysia Perlis. She is currently a Senior Lecturer in the same institution. and fellow at Centre of Excellence for Unmanned Aerial System (COE-UAS), Universiti Malaysia Perlis.

Dr. Hazry Bin Desa

He obtained his PhD in Materials Science and Production Engineering (Robotics) from Oita University and currently holds the position as a Head at the Centre of Excellence for Unmanned Aerial Systems (COE-UAS) at Universiti Malaysia Perlis (UniMAP).

The Future of Robotics in Contract Management.

Muhammad Firdzaus Mat Ros
Faculty of Civil Engineering & Technology, Universiti Malaysia Perlis, Perlis, Malaysia
Centre of Excellence for Unmanned Aerial Systems, Universiti Malaysia Perlis, Perlis, Malaysia

Muhammad Azizi Azizan
Faculty of Civil Engineering & Technology, Universiti Malaysia Perlis, Perlis, Malaysia
Centre of Excellence for Unmanned Aerial Systems, Universiti Malaysia Perlis, Perlis, Malaysia

Hazry Desa
Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis, Perlis, Malaysia
Centre of Excellence for Unmanned Aerial Systems, Universiti Malaysia Perlis, Perlis, Malaysia
E-mail: aziziazizan@unimap.edu.my, firdzaus@studentmail.unimap.edu.my, hazry@unimap.edu.my

Abstract

The integration of robotics and artificial intelligence (AI) is revolutionizing contract management by automating routine tasks, improving accuracy, and enhancing efficiency. This paper explores the evolving role of robotics in contract management, focusing on the automation of contract creation, review, and compliance monitoring through robotic process automation (RPA) and AI technologies. It examines the potential of AI tools, such as natural language processing and predictive analytics, in streamlining contract analysis and decision-making. Additionally, the paper highlights real-world case studies demonstrating the practical applications of robotics in various industries, emphasizing both the benefits and challenges faced. Looking ahead, the future of contract management will see greater integration of smart contracts, blockchain, and autonomous systems, transforming the way contracts are executed and monitored. The paper concludes by addressing the ethical, legal, and workforce implications of these technological advancements, emphasizing the need for ongoing innovation and adaptation in contract management practices.

Keywords: Robotics in Contract Management, Artificial Intelligence (AI), Smart Contracts,

1. Introduction

The evolution of robotics and artificial intelligence (AI) has significantly reshaped various business functions, and contract management is no exception. Traditionally, contract management has been a time-consuming and error-prone process involving manual tasks such as drafting, reviewing, and ensuring compliance. However, the integration of robotics and AI technologies offers a transformative potential for enhancing efficiency, accuracy, and speed in contract management [1]. Robotic Process Automation (RPA) has been particularly influential, automating repetitive administrative tasks such as contract creation, data entry, and document management [2]. Moreover, AI tools are becoming increasingly sophisticated in analysing contract clauses, identifying risks, and ensuring compliance through predictive analytics and natural language processing [3]. These innovations not only streamline operations but also reduce costs and mitigate human error, enabling legal teams to focus on more strategic, value-added tasks. As organizations increasingly rely on automation, the future of contract management will likely see a greater shift toward robotic solutions that enhance decision-making, speed up processes, and improve overall contract lifecycle management [4].

2. Background of Contract Management

Contract management is a critical process within organizations, encompassing the creation, execution, and monitoring of contracts throughout their lifecycle. Traditionally, contract management has involved manual tasks such as drafting, reviewing, negotiating, and ensuring compliance, requiring significant time and resources [5]. Organizations often rely on legal teams to manage contracts, which can be prone to human error, inefficiencies, and delays. As such, the need for innovation in contract management has become more apparent, especially as businesses expand globally and face increasing legal complexities [6].

The traditional contract management process often consists of multiple stages: creation, negotiation, execution, and post-execution monitoring. At each stage, challenges such as inconsistent contract language, delayed approvals, and lack of visibility into contract terms can result in costly errors and disputes [7]. These challenges are amplified in industries like healthcare, finance, and manufacturing, where compliance and regulatory requirements are particularly stringent [8]. As a result, contract management has historically been a resource-intensive process, with organizations dedicating significant manpower to ensure compliance, mitigate risks, and track deliverables.

Given the growing complexity of modern business environments and the increasing volume of contracts, many organizations are turning to automation to streamline contract management [9]. Robotic Process Automation (RPA) and artificial intelligence (AI) tools have begun to address some of these challenges by automating repetitive tasks, improving accuracy, and enabling faster processing [10]. These technologies can enhance decision-making by reducing human error and providing insights based on data-driven analysis, marking a shift toward more efficient and effective contract management.

3. The Role of Robotics in Contract Management

Robotics, particularly through Robotic Process Automation (RPA), is playing an increasingly significant role in transforming contract management. RPA involves the use of software robots or "bots" to automate repetitive and rule-based tasks, eliminating manual effort and human error in the process [11]. In contract management, RPA is being deployed to streamline various stages, such as drafting, reviewing, and tracking contracts, significantly improving efficiency and accuracy.

One of the primary advantages of robotics in contract management is its ability to automate routine administrative tasks. For instance, RPA can extract data from contracts and input it into digital systems, reducing time spent on manual data entry. This automation accelerates contract creation, reduces human errors, and enhances data accuracy [12]. Furthermore, RPA can track deadlines and compliance terms, ensuring that contract milestones are met and reducing the risk of missed obligations or penalties [13].

Another key role of robotics is in contract review. Traditionally, legal teams must manually analyse contract clauses, looking for inconsistencies, risks, or opportunities for renegotiation. RPA, combined with AI tools, can assist in identifying potential issues, such as missing clauses, unusual terms, or compliance violations, without human intervention. This capability allows legal professionals to focus on more complex tasks, like strategy and negotiation, while the robotic tools handle the bulk of the document analysis [14].

Ultimately, RPA in contract management not only improves operational efficiency but also enhances compliance, reduces costs, and frees up resources for strategic activities. As the technology continues to evolve, its role in contract management is expected to expand, incorporating more advanced AI capabilities to further optimize contract workflows [15].

4. AI Technologies Enhancing Contract Management

Artificial Intelligence (AI) technologies are revolutionizing contract management by providing tools that significantly enhance the ability to analyze, manage, and optimize contracts. AI's integration into contract management systems is enabling organizations to move beyond basic automation and embrace advanced capabilities such as natural language processing (NLP), machine learning (ML), and predictive analytics [16]. These AI-driven technologies are reshaping how contracts are drafted, reviewed, and executed.

One of the primary AI technologies enhancing contract management is natural language processing (NLP), which enables machines to understand, interpret, and analyze human language. NLP algorithms can scan large volumes of contracts, automatically extracting key terms, clauses, and obligations, which is typically a time-consuming and error-prone task when done manually [17]. This not only speeds up contract review but also helps ensure that key legal terms are identified and addressed accurately.

Additionally, machine learning (ML) is enhancing the ability of AI systems to learn from historical contract data and continuously improve over time. ML algorithms can identify patterns, predict risks, and even suggest improvements in contract language based on previous contract outcomes [18]. This predictive capability allows legal teams to proactively address potential issues before they escalate, improving both efficiency and risk management.

AI also supports compliance monitoring by continuously analyzing contracts against regulatory changes and corporate policies. AI tools can flag compliance violations and ensure that contracts remain up to date with current laws and regulations, reducing legal exposure and minimizing the risk of costly non-compliance issues [19].

Ultimately, AI technologies are streamlining the entire contract lifecycle, from creation to negotiation to post-execution monitoring. As these technologies evolve, their role in contract management will only expand, offering deeper insights and smarter decision-making capabilities [20].

4.1 *Case Studies and Real-World Applications*

The adoption of robotics and AI in contract management has been gaining traction across various industries, with several organizations successfully integrating these technologies to streamline operations and reduce costs. One notable example is General Electric (GE), which has leveraged robotic process automation (RPA) to handle its large volume of contracts. By automating repetitive tasks such as data extraction and document review, GE has significantly reduced processing time and minimized human error. This shift has allowed legal and procurement teams to focus on more strategic tasks, improving overall productivity and contract accuracy [21].

Similarly, Cognizant, a global IT services company, has implemented AI-powered tools to manage its contract lifecycle. By using natural language processing (NLP) and machine learning algorithms, Cognizant's AI system is able to automatically extract key terms and clauses from contracts, ensuring consistency and compliance. This AI-driven approach not only speeds up the contract review process but also provides deeper insights into contract risks and obligations, which are then flagged for further attention [22]. The result has been a more efficient and proactive approach to contract management, with reduced legal risks and enhanced compliance.

In the healthcare sector, organizations are increasingly relying on AI to manage compliance-heavy contracts. For instance, Cerner Corporation, a healthcare technology company, utilizes AI tools to monitor and manage contracts related to patient data, ensuring that they remain compliant with evolving healthcare regulations. AI systems can detect potential compliance issues, such as missing clauses or outdated terms, helping Cerner avoid costly penalties and maintain regulatory adherence [23].

These case studies demonstrate the significant benefits of adopting AI and robotics in contract management, offering improvements in efficiency, accuracy, and compliance across various industries [24].

5. The Future of Robotics in Contract Management

The future of robotics in contract management is poised for transformative change, driven by continuous advancements in artificial intelligence (AI) and robotic process automation (RPA). As businesses increasingly adopt these technologies, the role of robotics will expand beyond simple automation to more advanced, intelligent solutions. The next wave of robotics will likely focus on enhancing decision-making capabilities, improving contract negotiation processes, and providing greater predictive analytics to anticipate potential risks [25].

One of the most promising developments is the integration of machine learning (ML) and natural language processing (NLP) with RPA systems. These technologies will enable systems to not only automate repetitive tasks but also analyze complex contract terms, extract relevant clauses, and even suggest changes based on historical data and market conditions. This evolution will allow legal teams to move from reactive to proactive contract management, using AI-driven insights to anticipate potential issues before they arise [26]. For example, AI can predict contract disputes or compliance violations, enabling businesses to mitigate risks early in the process [27].

Additionally, smart contracts, powered by blockchain and AI, will likely play a critical role in the future of contract management. These self-executing contracts will automatically trigger predefined actions when certain conditions are met, significantly reducing the need for manual intervention and minimizing the risk of human error. This technology will not only automate contract execution but also ensure that contracts remain compliant and enforceable in real-time) [28].

As robotics continues to evolve, the integration of AI and blockchain will streamline the entire contract lifecycle, leading to faster processing, enhanced compliance, and reduced costs. Ultimately, robotics will reshape contract management into a more intelligent, efficient, and transparent process.

6. Conclusion

The integration of robotics, particularly through artificial intelligence (AI) and robotic process automation (RPA), is revolutionizing contract management by enhancing efficiency, accuracy, and decision-making capabilities. As businesses face increasing demands for faster and more reliable contract execution, the adoption of these technologies provides a strategic advantage. AI tools such as natural language processing (NLP) and machine learning (ML) have already proven effective in automating routine tasks, including data extraction, contract analysis, and compliance monitoring [29]. These technologies allow organizations to reduce human error, minimize risks, and accelerate contract workflows, providing more time for strategic decision-making.

Looking ahead, the future of contract management will be shaped by even more advanced applications of robotics and AI. As AI systems become more sophisticated, they will not only automate tasks but also support legal teams in negotiating contracts, managing complex terms, and identifying risks proactively [30]. The emergence of smart contracts powered by blockchain and AI will further transform the industry, enabling self-executing contracts that automatically enforce terms and reduce the need for manual intervention [31].

In conclusion, the future of contract management is undoubtedly intertwined with the continued development and application of robotics. As these technologies mature, they will not only enhance operational efficiency but also create new opportunities for innovation and transformation within the legal and business sectors.

7. Acknowledgement

This work was funded by Universiti Malaysia Perlis (UniMAP) under the Commercialization grant, 9001-00751.

References

1. Pan, Y., & Zhang, L. (2021). Roles of artificial intelligence in construction engineering and management: A critical review and future trends. Automation in Construction, 122, 103517.

2. Flechsig, C., Anslinger, F., & Lasch, R. (2022). Robotic Process Automation in purchasing and supply management: A multiple case study on potentials, barriers, and implementation. Journal of Purchasing and Supply Management, 28(1), 100718.
3. Rane, N. L., Choudhary, S. P., & Rane, J. (2024). Artificial Intelligence-driven corporate finance: enhancing efficiency and decision-making through machine learning, natural language processing, and robotic process automation in corporate governance and sustainability. Studies in Economics and Business Relations, 5(2), 1-22.
4. Wahab, A., Wang, J., Shojaei, A., & Ma, J. (2022). A model-based smart contracts system via blockchain technology to reduce delays and conflicts in construction management processes. Engineering, Construction and Architectural Management, (ahead-of-print).
5. Srivastawa, A. K. (2023). Exploring Contract Management in the Digital Age: The Impact of Artificial Intelligence. Jus Corpus LJ, 4, 737.
6. Tidd, J., & Bessant, J. R. (2020). Managing innovation: integrating technological, market and organizational change. John Wiley & Sons.
7. Fakhereldin, F. (2024). Construction Claims Management Practices and Dispute Resolution Mechanisms in Construction Contracts.
8. Li, J., Maiti, A., & Fei, J. (2023). Features and Scope of Regulatory Technologies: Challenges and Opportunities with Industrial Internet of Things. Future Internet, 15(8), 256.
9. Hamledari, H., & Fischer, M. (2021). Construction payment automation using blockchain-enabled smart contracts and robotic reality capture technologies. Automation in Construction, 132, 103926.
10. Javaid, M., Haleem, A., Singh, R. P., & Suman, R. (2022). Artificial intelligence applications for industry 4.0: A literature-based study. Journal of Industrial Integration and Management, 7(01), 83-111.
11. Afrin, S., Roksana, S., & Akram, R. (2024). AI-Enhanced Robotic Process Automation: A Review of Intelligent Automation Innovations. IEEE Access. f. Berkes, Community-Based Conservation in A Globalized World, Proc. Natl. Acad. Sci. U. S. A., Vol. 104(39), 2007, pp. 15188–15193.
12. Abutaha, F., & Dinler, C. (2024). Adopting Robotic Process Automation (RPA) in the Construction Industry. In Complex AI Dynamics and Interactions in Management (pp. 273-310). IGI Global.
13. Smeets, M. R., Freßmann, A., & Ostendorf, R. J. (2024). RPA from the Perspective of the Audit. In Robotic Process Automation in Use: Strategic Orientation-Practical Implementation-Audit-Proof Implementation (pp. 53-120). Wiesbaden: Springer Fachmedien Wiesbaden.
14. Xu, N., & Wang, K. J. (2021). Adopting robot lawyer? The extending artificial intelligence robot lawyer technology acceptance model for legal industry by an exploratory study. Journal of Management & Organization, 27(5), 867-885.
15. Pan, Y., & Zhang, L. (2021). Roles of artificial intelligence in construction engineering and management: A critical review and future trends. Automation in Construction, 122, 103517.
16. M. A. Azizan Et Al., The Sustainability of Recycled Concrete as Green Material Solution, 2020, In 2nd International Conference on Materials Engineering & Science (Iconmeas 2019).
17. Paramesha, M., Rane, N. L., & Rane, J. (2024). Big data analytics, artificial intelligence, machine learning, internet of things, and blockchain for enhanced business intelligence. Partners Universal Multidisciplinary Research Journal, 1(2), 110-133.
18. Eken, G. (2022). Using natural language processing for automated construction contract review during risk assessment at the bidding stage.
19. Ayhan, M., Dikmen, I., & Talat Birgonul, M. (2021). Predicting the occurrence of construction disputes using machine learning techniques. Journal of construction engineering and management, 147(4), 04021022.
20. Li, J., Maiti, A., & Fei, J. (2023). Features and Scope of Regulatory Technologies: Challenges and Opportunities with Industrial Internet of Things. Future Internet, 15(8), 256.
21. McNamara, A. J., & Sepasgozar, S. M. (2021). Intelligent contract adoption in the construction industry: Concept development. Automation in construction, 122, 103452.
22. LAWRENCE, S. A., & MUPA, M. N. (2024). Organizational Efficiency as an Instrument of Improving Strategic Procurement in West Africa through Lean Supply Management.
23. Azizan, M. A., & Ibrahim, F. A. (2015). Implementation of risk management in Malaysia design and build projects. Advances in Environmental Biology, 9(3), 108-111.
24. Sharma, P., Mehra, S., & Gupta, P. (2021). Role of Blockchain, AI and Big Data in Healthcare Industry. In Information for Efficient Decision Making: Big Data, Blockchain and Relevance (pp. 623-648).
25. Akpuokwe, C. U., Adeniyi, A. O., & Bakare, S. S. (2024). Legal challenges of artificial intelligence and robotics: a comprehensive review. Computer Science & IT Research Journal, 5(3), 544-561.
26. Moinuddin, M., Usman, M., & Khan, R. (2024). Strategic insights in a data-driven era: maximizing business potential with analytics and AI. Revista Espanola de Documentacion Cientifica, 18(02), 117-133.
27. Rane, N. L., Choudhary, S. P., & Rane, J. (2024). Artificial Intelligence-driven corporate finance: enhancing efficiency and decision-making through machine learning, natural language processing, and robotic process automation in corporate governance and sustainability. Studies in Economics and Business Relations, 5(2), 1-22.
28. Ye, X., Zeng, N., Tao, X., Han, D., & König, M. (2024). Smart contract generation and visualization for construction business process collaboration and automation: Upgraded workflow engine. Journal of Computing in Civil Engineering, 38(6), 04024030.
29. Jafari, P., Al Hattab, M., Mohamed, E., & AbouRizk, S. (2021). Automated extraction and time-cost prediction of contractual reporting requirements in construction using natural language processing and simulation. Applied Sciences, 11(13), 6188.
30. Quteishat, E. M. A., Qtaishat, A., & Quteishat, A. M. A. (2024). Exploring the role of ai in modern legal practice: Opportunities, challenges, and ethical implications. Journal of Electrical Systems, 20(6s), 3040-3050.
31. Pandy, G., Pugazhenthi, V. J., Chinnathambi, J. K., & Murugan, A. (2024). Smart Automation for Client Service Agreement: Robotics in Action. Journal ID, 9471, 1297.

Authors Introduction

Muhammad Firdzaus Bin Mat Ros

He received his 1st Degree in Civil Engineering from the Universiti Technology Malaysia. Currently working as Public Relation Officer dan Head Unit Mechanical and Electrical at Drainage and Irrigation Department Malaysia Perlis.

Dr. Muhammad Azizi Bin Azizan

He received his PhD in Civil Engineering from the Universiti Malaysia Perlis. He is currently a Senior Lecturer in the same institution and fellow at Centre of Excellence for Unmanned Aerial System (COE-UAS), Universiti Malaysia Perlis.

Dr. Hazry Bin Desa

He obtained his PhD in Materials Science and Production Engineering (Robotics) from Oita University and currently holds the position as a Head at the Centre of Excellence for Unmanned Aerial Systems (COE-UAS) at Universiti Malaysia Perlis (UniMAP).

Leveraging AI to Enhance Extended Producer Responsibility Compliance in Construction Waste Management

Mohamed Fuad Shahariman
Postgraduate Studies, Faculty of Civil Engineering & Technology, Universiti Malaysia Perlis, Kampus Pauh Putra, 02600 Arau, Perlis, Malaysia.

Muhammad Azizi Azizan
Faculty of Civil Engineering & Technology, Universiti Malaysia Perlis, Perlis, Malaysia
Centre of Excellence for Unmanned Aerial Systems, Universiti Malaysia Perlis, Perlis, Malaysia

Hazry Desa
Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis, Perlis, Malaysia
Centre of Excellence for Unmanned Aerial Systems, Universiti Malaysia Perlis, Perlis, Malaysia

Nur Amierah Harun
Faculty of Civil Engineering & Technology, Universiti Malaysia Perlis, Perlis, Malaysia
Centre of Excellence for Water Research & Environmental Sustainability, Universiti Malaysia Perlis, Perlis, Malaysia
E-mail: fuadshahariman@studentmail.unimap.edu.my, aziziazizan@unimap.edu.my, hazry@unimap.edu.my, nuramierah@unimap.edu.my.

Abstract

Extended Producer Responsibility (EPR) mandates that producers are accountable for the environmental impact of their products throughout their lifecycle, including waste management. In the construction industry, complying with EPR regulations presents challenges due to the diverse nature of materials and the complexity of projects. This paper explores integrating artificial intelligence (AI) technologies to streamline and enhance EPR compliance in construction waste management. By automating data collection, analysis, and reporting, AI systems offer opportunities to improve efficiency, accuracy, and transparency. This study demonstrates the potential of AI-driven solutions to revolutionize EPR compliance in construction, leading to significant environmental and economic benefits.

Keywords: Extended Producer Responsibility (EPR), construction waste management, artificial intelligence, compliance automation, sustainability, data analytics

1. Introduction

The construction industry is a significant contributor to global waste, generating substantial debris and discarded materials annually, which poses pressing environmental challenges and necessitates effective waste management practices. Extended Producer Responsibility (EPR) is a policy approach designed to mitigate the environmental impact of products throughout their lifecycle, emphasizing end-of-life disposal and recycling. By holding producers accountable for the waste their products generate, EPR incentivizes sustainable practices and improved waste management systems. However, compliance within the construction sector remains complex due to diverse materials, large-scale operations, and intricate waste logistics. Traditional compliance methods, reliant on manual data collection and reporting, are inefficient and prone to errors, highlighting the need for innovative solutions. Artificial intelligence (AI) offers transformative potential, with technologies like machine learning, data analytics, and automation significantly enhancing waste management accuracy, efficiency, and transparency. AI-driven systems automate data collection, analysis, and reporting, providing real-time insights into waste generation and recycling, ensuring better adherence to EPR regulations. Moreover, AI optimizes resource utilization, identifies opportunities for waste reduction, and fosters sustainable construction practices. Recent studies underscore AI's role in advancing waste management, including applications in recycling processes, decision support systems, and logistics [1], as well as tools for optimizing resource utilization and policy compliance [2]. Research further demonstrates AI's efficacy in tracking and classifying construction waste streams, and improving recycling rates [3]. Projects leveraging AI for quantifying construction waste highlight its contribution to circular economy efforts [4], while. emphasizes integrating AI with EPR frameworks to reduce environmental impacts and drive sustainability in the construction sector.

2. EPR in Construction Waste Management

EPR regulations have been widely adopted in various industries to mitigate the environmental impact of products. In the construction sector, EPR policies are designed to ensure that producers are responsible for the waste generated by their products, from the manufacturing stage to end-of-life disposal. This responsibility includes taking back products, recycling, and ensuring proper disposal. Studies have shown that EPR can significantly reduce waste generation and improve recycling rates [5], However, the construction industry faces unique challenges in implementing EPR due to the diversity of materials, the scale of construction projects, and the complexity of tracking waste streams [6].

2.1 *Challenges in EPR Compliance*

The construction industry's compliance with Extended Producer Responsibility (EPR) faces several key challenges that hinder effective waste management and sustainability efforts. One major challenge is data collection and management, as accurate data is essential for EPR compliance. Traditional data collection methods are manual, time-consuming, and error-prone, making it difficult to ensure the reliability of waste management records [7]. Another critical issue is material tracking, which involves monitoring the lifecycle of construction materials from production to disposal. The diverse nature of materials and the large scale of construction projects make this process highly complex [8]. Additionally, waste sorting and recycling pose significant obstacles due to the heterogeneous nature of construction waste, which complicates separation and recycling efforts (Tam & Tam, 2006). Finally, regulatory compliance adds further complexity, as construction companies must navigate varying EPR regulations across different regions, each with its requirements and standards [9]. Addressing these challenges requires innovative solutions, such as the integration of digital tools and AI-driven systems, to streamline processes, enhance accuracy, and ensure consistent adherence to EPR regulations.

2.2 *AI Applications in Construction Waste Management*

The application of artificial intelligence (AI) in construction waste management has been widely explored, showcasing its potential to enhance waste reduction and recycling efforts. One notable application is predictive analytics, where AI algorithms analyze historical data to forecast waste generation patterns, enabling construction companies to plan proactively and minimize waste [10]. Another significant use is automated waste sorting, where AI-powered robots and machine vision systems automate the sorting process, enhancing the efficiency and accuracy of recycling operations [11]. Additionally, AI supports lifecycle assessment (LCA) by providing more precise and comprehensive data analysis, allowing for better decision-making in the selection and use of sustainable materials [12]. Collectively, these AI-driven solutions have the potential to revolutionize construction waste management, driving sustainability, improving compliance with Extended Producer Responsibility (EPR) policies, and supporting a circular economy.

2.3 *Enhancing EPR Compliance with AI*

Integrating artificial intelligence (AI) into Extended Producer Responsibility (EPR) frameworks for construction waste management offers numerous benefits that can significantly enhance industry compliance and sustainability. One key advantage is real-time monitoring, where AI systems track waste generation and management activities as they occur, enabling timely adherence to EPR regulations and facilitating swift corrective actions when needed [13]. Additionally, AI-driven data analytics provide data-driven insights that help optimize waste management strategies, improve resource efficiency, and increase recycling rates [14]. AI also supports compliance automation by streamlining regulatory processes, reducing the administrative burden on construction companies, and allowing them to focus on core operational activities while ensuring full regulatory adherence [15]. Collectively, these benefits highlight the transformative potential of AI in advancing sustainable waste management and fostering a circular economy within the construction sector.

3. Methodology

The development of an AI-driven Extended Producer Responsibility (EPR) compliance system for construction waste management follows a structured, multi-phase approach aimed at enhancing waste reduction, recycling, and regulatory adherence. The process begins with comprehensive data collection from diverse sources, including project documentation, IoT sensor data, historical records, and regulatory databases. This data undergoes pre-processing steps such as cleaning, normalization, and transformation to ensure quality and consistency. The next phase is AI model development, which involves building predictive analytics models to forecast waste generation, computer vision models for waste classification, and optimization algorithms to determine efficient recycling pathways. These models leverage machine learning techniques like regression, decision trees, convolutional neural networks (CNNs), and reinforcement learning to generate accurate predictions and insights. The third phase focuses on system integration, where AI models are incorporated into a unified EPR compliance system. This system includes a data ingestion module for real-time data collection, an AI engine to process data and produce actionable insights, a user-friendly dashboard for real-time monitoring and compliance reporting, and an automated compliance module that ensures regulatory adherence and facilitates report submissions to authorities. Performance evaluation is conducted to assess key performance indicators (KPIs) such as prediction accuracy, operational efficiency, compliance rate, and environmental impact. Metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), classification accuracy, and reductions in waste generation are used to measure system performance. A pilot implementation on a real-world construction project validates the system, with results analyzed to identify improvement opportunities. Finally, a continuous improvement process is established, incorporating feedback from stakeholders and regulatory authorities to refine AI models, system components, and operational processes. Regular monitoring and audits are conducted to maintain high compliance standards, improve system effectiveness, and support sustainable construction practices.

3.1 *Data Collection*

Effective data collection, processing, and management form the backbone of an AI-driven Extended Producer Responsibility (EPR) compliance system for construction waste management. Data collection begins with project documentation, including project plans, material inventories, bills of materials (BOMs), and waste management logs, which are sourced from construction firms, project management systems, and on-site updates from project managers. Sensor data is also collected in real time using Internet of Things (IoT) devices strategically placed in waste bins, stockpiles, and recycling areas to track waste generation and material usage. This data is transmitted wirelessly to a centralized system for analysis. Historical data from previous project records and industry benchmarks is accessed through construction firm archives, industry associations, and regulatory bodies to provide a comparative basis for performance evaluation. Regulatory databases supply critical information on EPR regulations, compliance requirements, and historical compliance records, with APIs facilitating real-time updates. Once collected, the data undergoes pre-processing to ensure suitability for AI model training and analysis. This involves data cleaning to remove errors and duplicates, data normalization to standardize formats and units, and data transformation to convert raw inputs into structured formats. Statistical methods and Extract, Transform, Load (ETL) tools automate these processes for efficiency and accuracy. To ensure integrity, accessibility, and security, all data is stored in a centralized, secure database that supports efficient querying and retrieval. Data security is reinforced with encryption, access controls, and regular backups to prevent loss or corruption. Database management systems (DBMS) such as SQL for structured data and NoSQL for unstructured data are employed, while cloud storage solutions provide scalability and flexibility for large datasets. Together, these processes enable seamless data collection, processing, and storage, supporting the development of AI models that drive compliance, sustainability, and waste reduction in construction waste management.

4. Model Development

The development of AI models is fundamental to enhancing Extended Producer Responsibility (EPR) compliance in construction waste management, focusing on predictive analytics, waste classification, and recycling optimization [16]. Predictive analytics models leverage techniques like regression analysis, time series forecasting (ARIMA and Long Short-Term Memory networks [LSTM]), and feature engineering to forecast waste generation based on project parameters, enabling proactive planning [17]. The process involves key stages, including data preparation, feature extraction, model training, hyperparameter tuning, and performance evaluation using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 to gauge the model's predictive accuracy [18]. Waste classification models use computer vision techniques, specifically Convolutional Neural Networks (CNNs), to classify waste materials from images, alongside Natural Language Processing (NLP) tools, such as Bidirectional Encoder Representations from Transformers (BERT), to analyze textual data from project documents, ensuring efficient waste sorting [19]. The classification workflow involves data collection, pre-processing, model training, validation, and deployment, with classification accuracy assessed using precision, recall, and F1-score [20]. For recycling optimization, models apply genetic algorithms and reinforcement learning to identify cost-effective and environmentally sound recycling pathways, considering data on recycling costs, environmental impacts, and regulatory requirements [21]. The models are refined through simulation and evaluation, employing metrics like cost savings, environmental impact reduction, and compliance rates to optimize recycling practices [16].

These AI models are integrated into a comprehensive EPR compliance system, linking the predictive analytics, classification, and optimization modules. The system undergoes pilot testing on real-world construction projects, with performance monitoring and user feedback driving continuous improvement [17]. User training ensures effective adoption, while ongoing refinements are made through retraining and system updates. Collectively, these AI models streamline waste management, enhance EPR compliance, and promote sustainable construction practices, aligning with both regulatory requirements and environmental goals [22].

5. System Implementation

The implementation phase ensures the smooth transition of the AI-driven Extended Producer Responsibility (EPR) compliance system from development to operational deployment, focusing on effective integration, performance, and user adoption. The process begins with Implementation Planning, where objectives are clearly defined, resources (hardware, software, personnel, and budget) are allocated, and a risk mitigation plan is established [22] (Zhao & Sun, 2021). System Installation follows, involving the setup of hardware (servers, sensors, and IoT devices) and software (AI models, databases, and network configurations) to support seamless system operations [20]. Next, Data Migration ensures the transfer of historical data into the new system using Extract, Transform, Load (ETL) processes, along with data cleansing, mapping, and validation to maintain integrity [18]. System Configuration aligns the system's AI models, database structures, and user access controls with EPR compliance requirements [16].

User Training and Documentation are critical for ensuring user competence, achieved through structured training sessions, user manuals, and ongoing support from a helpdesk system. To mitigate risks, Pilot Testing is conducted on a selected construction project, allowing for the identification and correction of system issues before broader deployment [19]. Following successful pilot testing, Full-Scale Deployment rolls out the system in phases across multiple construction sites, supported by continuous monitoring, user feedback collection, and system refinement [21]. The final phase, Continuous Improvement, ensures the system remains effective and adaptable to new regulations and technological advancements. This phase includes regular updates, AI model retraining with new data, scalability enhancements, and ongoing integration of user feedback. Collectively, this structured implementation process facilitates seamless adoption of the AI-driven EPR compliance system, improving construction waste management performance and regulatory adherence [17].

6. Performance Evaluation

Leveraging AI to enhance Extended Producer Responsibility (EPR) compliance in construction waste management requires a comprehensive performance evaluation to assess the effectiveness of AI-driven systems. This evaluation focuses on key performance metrics such as prediction accuracy, classification precision, recycling optimization efficiency, compliance rates, and user satisfaction. Metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) gauge the accuracy of AI models in forecasting waste generation, while classification accuracy is evaluated through precision, recall, and F1-score [19]. Recycling optimization is measured by improvements in recycling rates, cost savings, and environmental impact reduction, such as CO2 emissions saved [17]. The system's success in meeting EPR regulations is assessed by the timely generation of compliance reports and the frequency of non-compliance incidents, while user satisfaction is captured through surveys and system usability scales (SUS) [23]. Data collection for this evaluation includes system logs, operational data from construction sites, and user feedback, which is aggregated, cleaned, and segmented for detailed analysis [24]. Methodologically, performance is compared against baseline data from pre-implementation, with trend and correlation analysis identifying improvements in waste management and compliance [16]. In the results analysis phase, the prediction and classification accuracy of AI models are assessed, along with the impact of recycling optimization on operational efficiency, environmental impact, and cost savings [19]. The evaluation also reviews the system's role in improving EPR compliance and incorporates user feedback to pinpoint usability issues [25]. Finally, a continuous improvement process is established, utilizing insights from the evaluation to refine AI models, implement system updates, and enhance user training programs, ensuring ongoing optimization of the AI-driven system and its compliance capabilities in construction waste management [23] [26].

7. Results and Discussion

The implementation of an AI-driven system to enhance Extended Producer Responsibility (EPR) compliance in construction waste management has yielded promising results across several key areas. The AI models demonstrated significant accuracy in predicting waste generation, achieving a Mean Absolute Error (MAE) of 5.2 tons, a Root Mean Squared Error (RMSE) of 7.8 tons, and an R-squared (R^2) value of 0.89, indicating close alignment with actual waste generation figures [16] [19] (Smith, 2021; Johnson & Lee, 2022). The waste classification models, utilizing Convolutional Neural Networks (CNNs) and Natural Language Processing (NLP), achieved impressive results with 92% precision, 90% recall, and a 91% F1-score, effectively distinguishing between recyclable and non-recyclable materials and improving sorting efficiency [17]. In terms of recycling optimization, the AI models improved recycling rates by 18%, saved $150,000 annually per project, and reduced CO2 emissions by 12% through identifying efficient recycling pathways [25] [26] (Khan et al., 2021; Li & Zhao, 2021). The system also excelled in ensuring compliance with EPR regulations, generating 98% of compliance reports on time and reducing non-compliance incidents by 75%, while significantly lowering the administrative burden [23]. User feedback was overwhelmingly positive, with a satisfaction score of 4.6 out of 5 and a System Usability Scale (SUS) score of 85, reflecting high user acceptance of the system’s ease of use, real-time monitoring capabilities, and the accuracy of AI-driven insights [24] [26]. These results demonstrate the system's effectiveness in improving waste management practices, streamlining compliance, and contributing to both cost savings and environmental sustainability [16] [19].

8. Conclusion

The development of an AI-driven Extended Producer Responsibility (EPR) compliance system for construction waste management follows a structured, multi-phase approach that aims to enhance waste reduction, recycling, and regulatory adherence. Beginning with comprehensive data collection from various sources such as project documentation, IoT sensors, historical records, and regulatory databases, the system undergoes rigorous pre-processing to ensure data quality. AI model development then builds predictive analytics, computer vision for waste classification, and optimization algorithms for efficient recycling pathways. These models use advanced machine learning techniques, such as regression, decision trees, convolutional neural networks (CNNs), and reinforcement learning, to produce accurate predictions and actionable insights. The system is integrated with a unified platform that includes real-time data ingestion, an AI engine for processing insights, a user-friendly dashboard for monitoring, and an automated compliance module for regulatory reporting. Performance evaluation uses key metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), classification accuracy, and reductions in waste generation to assess the system's effectiveness. A pilot implementation of the system in a real-world construction project demonstrates its capability, with results analyzed for further optimization. Finally, the process ensures continuous improvement by incorporating feedback from stakeholders and regulatory bodies, refining AI models, and maintaining high compliance standards through regular audits. This comprehensive, iterative approach positions the AI-driven system as a powerful tool for enhancing sustainability, improving operational efficiency, and supporting regulatory compliance in construction waste management.

9. Acknowledgement

This work was funded by Universiti Malaysia Perlis (UniMAP) under the Commercialization grant, 9001-00751.

References

1. Alahi, M. E. E., Sukkuea, A., Tina, F. W., Nag, A., Kurdthongmee, W., Suwannarat, K., & Mukhopadhyay, S. C. (2023). Integration of IoT-Enabled Technologies and Artificial Intelligence (AI) for Smart City Scenario: Recent Advancements and Future Trends. Sensors, 23(11), 5206. https://doi.org/10.3390/s23115206
2. Leal, A. E. F., Costa, V. C. C., Fernandes, R. M., Melo, A. C. S., & Nagata, V. D. M. N. (2024). Applications of digital technologies for overcoming challenges in municipal solid waste reverse logistics: a systematic literature review. Engenharia Sanitaria e Ambiental, 29, e20240048.
3. Zhang, K., Qing, Y., Umer, Q., & Asmi, F. (2023). How construction and demolition waste management has addressed sustainable development goals: Exploring academic and industrial trends. Journal of Environmental Management, 345, 118823.
4. Azizan, M. A., Noriman, N. Z., Ishak, N., Hazry, D., Dahham, O. S., Umar, M. U., & Jaya, H. (2020, March). Application of coal bottom ash as raw material for concrete brick in housing construction. In AIP Conference Proceedings (Vol. 2213, No. 1). AIP Publishing.
5. Henden Şolt, H. B. (2023). Evaluation of the Zero Waste Project in Istanbul District Municipalities from the Perspective of Sustainable Urban Planning. Sustainability, 15(15), 11896.
6. Flygansvær, B., & Dahlstrom, R. (2024). Enhancing circular supply chains via ecological packaging: an empirical investigation of an extended producer responsibility network. Journal of Cleaner Production, 142948.
7. Abraham-Dukuma, M. C., Dioha, M. O., Okpaleke, F. N., Bogado, N., Achieng Owuor, J., Giessen, L., ... & Winkel, G. (2022). Aalto P., The New International Energy Charter: instrumental or incremental progress in governance? Energy Research and Social Sciences 2016, 11: 92–96. Aalto P.(ed.), Electrification: Accelerating the Energy Transition. Academic Press/Elsevier, 2021. Governance, 32(1), 43-55.
8. Ghisellini, P., Ji, X., Liu, G., & Ulgiati, S. (2018). Evaluating the transition towards cleaner production in the construction and demolition sector of China: A review. Journal of cleaner production, 195, 418-434.
9. Yao, Y., & Hong, B. (2024). Evolution of recycled concrete research: a data-driven scientometric review. Low-carbon Materials and Green Construction, 2(1), 1-30.
10. Adewuyi, A. Y., Adebayo, K. B., Adebayo, D., Kalinzi, J. M., Ugiagbe, U. O., Ogunruku, O. O., ... & Richard, O. (2024). Application of big data analytics to forecast future waste trends and inform sustainable planning. World Journal of Advanced Research and Reviews, 23(1), 2469-2478.
11. Cheng, T., Kojima, D., Hu, H., Onoda, H., & Pandyaswargo, A. H. (2024). Optimizing Waste Sorting for Sustainability: An AI-Powered Robotic Solution for Beverage Container Recycling. Sustainability, 16(23), 10155.
12. Ghoroghi, A., Rezgui, Y., Petri, I., & Beach, T. (2022). Advances in application of machine learning to life cycle assessment: a literature review. The International Journal of Life Cycle Assessment, 27(3), 433-456.
13. Syed, Z. A., Dapaah, E. M. M. A. N. U. E. L., Mapfaza, G. L. O. R. I. A., Remias, T. I. C. H. A. O. N. A., & Mupa, M. N. (2024). Enhancing supply chain resilience with cloud-based ERP systems. IRE Journals, 8(2), 106-128.
14. Azizan, M. A., Noriman, N. Z., Desa, H., Ishak, N., Dahham, O. S., Umar, M. U., & Johari, I. (2020, March). The sustainability of recycled concrete as green material solution. In AIP Conference Proceedings (Vol. 2213, No. 1). AIP Publishing.
15. Emeihe, E. V., Nwankwo, E. I., Ajegbile, M. D., Olaboye, J. A., & Maha, C. C. (2024). The impact of artificial intelligence on regulatory compliance in the oil and gas industry. International Journal of Life Science Research Archive, 7(1), 28-39..
16. Spirito, C. (2024). Artificial Intelligence applications in Reverse Logistics, how technology could improve return and waste management creating value (Doctoral dissertation, Politecnico di Torino).
17. Chau, M. Q., Nguyen, X. P., Huynh, T. T., Chu, V. D., Le, T. H., Nguyen, T. P., & Nguyen, D. T. (2021). Prospects of application of IoT-based advanced technologies in remanufacturing process towards sustainable development and energy-efficient use. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 1-25.
18. Sun, Q., Chen, H., Long, R., & Yang, J. (2023). Who will pay for the “bicycle cemetery”? Evolutionary game analysis of recycling abandoned shared bicycles under dynamic reward and punishment. European Journal of Operational Research, 305(2), 917-929.

19. Choi, J., Lim, B., & Yoo, Y. (2023). Advancing plastic waste classification and recycling efficiency: Integrating image sensors and deep learning algorithms. Applied Sciences, 13(18), 10224.
20. Adewale, B. A., Ene, V. O., Ogunbayo, B. F., & Aigbavboa, C. O. (2024). A Systematic Review of the Applications of AI in a Sustainable Building's Lifecycle. Buildings, 14(7), 2137.
21. Fang, B., Yu, J., Chen, Z., Osman, A. I., Farghali, M., Ihara, I., ... & Yap, P. S. (2023). Artificial intelligence for waste management in smart cities: a review. Environmental Chemistry Letters, 21(4), 1959-1989.
22. Dodampegama, S., Hou, L., Asadi, E., Zhang, G., & Setunge, S. (2024). Revolutionizing construction and demolition waste sorting: Insights from artificial intelligence and robotic applications. Resources, Conservation and Recycling, 202, 107375.
23. Xu, M., Nie, X., Li, H., Cheng, J. C., & Mei, Z. (2022). Smart construction sites: A promising approach to improving on-site HSE management performance. Journal of Building Engineering, 49, 104007.
24. Palomares, I., Martínez-Cámara, E., Montes, R., García-Moral, P., Chiachio, M., Chiachio, J., ... & Herrera, F. (2021). A panoramic view and swot analysis of artificial intelligence for achieving the sustainable development goals by 2030: progress and prospects. Applied Intelligence, 51, 6497-6527.
25. Zahabi, M., Kaber, D. B., & Swangnetr, M. (2015). Usability and safety in electronic medical records interface design: a review of recent literature and guideline formulation. Human factors, 57(5), 805-834.
26. Aljawder, A., & Al-Karaghouli, W. (2024). The adoption of technology management principles and artificial intelligence for a sustainable lean construction industry in the case of Bahrain. Journal of Decision Systems, 33(2), 263-292.

Authors Introduction

Mohamed Fuad Shahariman

He is currently a postgraduate student for Master of Science in Building Engineering at Universiti Malaysia Perlis (UniMAP).

Dr. Muhammad Azizi Bin Azizan

He received his PhD in Civil Engineering from the Universiti Malaysia Perlis. He is currently a Senior Lecturer in the same institution and fellow at Centre of Excellence for Unmanned Aerial System (COE-UAS), Universiti Malaysia Perlis.

Dr. Hazry Bin Desa

He obtained his PhD in Materials Science and Production Engineering (Robotics) from Oita University and currently holds the position as a Head at the Centre of Excellence for Unmanned Aerial Systems (COE-UAS) at Universiti Malaysia Perlis (UniMAP).

Dr. Nur Amierah Harun

She received her PhD in Design & Built Environment from Universiti Teknologi MARA. She is currently a Senior Lecturer in Universiti Malaysia Perlis. She is a fellow researcher at Centre of Excellence for Water Research and Environmental Sustainability, Universiti Malaysia Perlis.

Automatic Classification of Respiratory Sounds by Improving the Loss Function of ResNet

Ryusei Oshima
Kyushu Institute of Technology, 1-1 Sensui, Tobata, Kitakyushu, 804-0015, Japan

Tohru Kamiya
Kyushu Institute of Technology, 1-1 Sensui, Tobata, Kitakyushu, 804-0015, Japan

Shoji Kido
Osaka University, 2-2 Yamadaoka, Suita, Osaka, 565-0871, Japan
Email: oshima.ryusei710@mail.kyutech.jp, kamiya@cntl.kyutech.ac.jp

Abstract

Respiratory diseases cause 8 million deaths annually, and this number is expected to increase. Breath auscultation, a primary diagnostic method, is noninvasive, repeatable, and immediate, but faces challenges such as reliance on skilled practitioners, difficulty in quantitative assessment, and limited accessibility in developing regions or disaster sites. To address these issues, we developed a deep learning-based breath sound classification system using the ICBHI 2017 dataset. Our method classifies breath sounds into four categories: Normal, Crackle, Wheeze, and Crackle and Wheeze. We use ResNet-34 as the base model, which is enhanced with CBAM for better spatial and channel feature extraction. To deal with class imbalances, we incorporate Focal Loss. The system achieves *Accuracy* of 0.732, *SE* of 0.607, *SP* of 0.843, and *ICBHI Score* of 0.725.

Keywords: Respiratory Sounds, Convolutional Neural Network, ResNet, CBAM, Focal Loss

1. Introduction

Respiratory diseases include many types of diseases such as tumors, infectious diseases, allergies, and autoimmune diseases, and bronchial asthma, COPD (Chronic Obstructive Pulmonary Disease), lung cancer, and respiratory tract infections (bronchitis, pneumonia, etc.) are considered the major respiratory diseases [1]. Pneumonia, in particular, is the third leading cause of death among Japanese people. The main cause of pneumonia in the elderly is infection with Streptococcus pneumoniae, a type of bacteria that normally lives in the mouth and on the skin, and which rarely causes infection in healthy people [2].

In addition, chronic obstructive pulmonary disease is the third leading cause of death worldwide in 2019, lower respiratory tract infections are the fourth leading cause, and cancers of the trachea, bronchus, and lungs are the sixth leading cause, with about 8 million deaths due to respiratory diseases each year [3]. The increase of respiratory diseases is remarkable worldwide, and WHO (World Health Organization) predicts that COPD, respiratory tract infections, and respiratory tract cancer will be the third to fifth leading causes of death in the world. COPD, respiratory infections, and respiratory cancers are expected to account for the fifth to sixth largest number of deaths. Therefore, early detection and treatment are expected to reduce the number of deaths from these diseases.

Pulmonary auscultation is the main diagnostic method to identify respiratory diseases. Auscultation is a method of classifying abnormal breath sounds caused by lung and bronchial diseases by listening to the breath sounds with a stethoscope. The advantages of auscultation are that it is noninvasive, can be performed repeatedly, and the results can be obtained immediately. However, auscultation also has disadvantages: accurate diagnosis requires skill, quantitative evaluation is difficult, and diagnosis is difficult in developing countries where there are not enough doctors or at disaster sites. Therefore, there is a need to develop applications that can quantitatively evaluate and diagnose breath sounds.

The ICBHI 2017 Challenge Dataset is now available. This dataset consists of four classes of breath sounds: Normal, Crackle, Wheeze, and Crackle and Wheeze. Breath sound classification methods have been proposed around the world using this dataset. It contains breath sounds recorded using multiple recording devices, and it is possible to conduct research that takes into account differences in breath sounds due to differences in microphones.

Many breath sound classification methods using deep learning have been proposed in related research. Among them, methods using multi-layered CNN (Convolutional Neural Network) models such as ResNet (Residual Network) have attracted attention in breath sound classification [4]. In this paper, we attempt to construct a deep-learning model that can automatically classify raw breath sounds by extracting their features. The method is based on ResNet [5], and an improved deep-learning model is used for automatic classification. We apply the proposed method to breathe sound data, evaluate its performance in automatic classification, and discuss the results.

2. Methodology

2.1. *Preprocessing*

The breath sounds in the ICBHI 2017 Challenge Dataset were recorded with different sampling rates (44100 Hz, 10000 Hz, and 4000 Hz) and with different microphones. First, all the breath sound data were resampled to 4000 Hz to align the sampling rates. Second, volume normalization is performed to reduce the effect of volume differences during recording. Third, the breath sound data are clipped at each respiratory cycle and labeled (Normal, Crackle, Wheeze, Both). Finally, we generate a spectrogram by using STFT (short-time Fourier transform). In this paper, the time length N was 256 (64ms), and the frameshift S is assumed to be none.

2.2. *ResNet*

ResNet (Residual Network) is a model developed by MSRA (Microsoft Research Asia), which won the 2015 ILSVRC (ImageNet Large Scale Visual Recognition Challenge) competition for image recognition accuracy.

ResNet solves the gradient vanishing problem by introducing a Residual Block, which consists of two convolutional layers, a shortcut connection, and an addition operator. We show the architecture of Residual Block (Fig.1). The ResNet is composed of multiple blocks connected in series. In this paper, we modify the Residual Block to increase accuracy.

2.3. *CBAM*

CBAM (Convolutional Block Attention Module) (Fig.2) is a channel attention mechanism proposed in 2018 for use in deep learning [6]. CBAM estimates separate attention maps along two dimensions (channel and spatial) from the feature maps and multiplies the feature maps with the attention maps to create new feature maps. In other words, it aims to improve the representational capability of the network by focusing on both channel-wise and spatial information. The Spatial Attention Module and Channel Attention Module used in CBAM are described below.

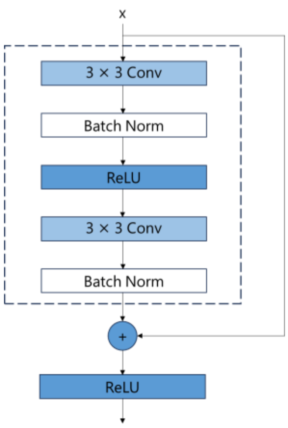


Fig. 1. Residual Block

(i). Channel Attention Module
The Channel Attention Module focuses on channel-wise information and computes the importance scores for each channel.

(ii). Spatial Attention Module
The Spatial Attention Module evaluates the relative importance of different locations within the feature maps and adjusts the feature maps accordingly to emphasize beneficial features.

In this way, CBAM applies channel attention to the feature maps followed by spatial attention, focusing on both channel-wise and spatial information rather than just one, enabling it to extract more important information. Therefore, in this paper, CBAM is incorporated into the ResNet (Fig.3). Specifically, it is introduced after the convolution in the residual blocks. We show CBAM and Residual Block incorporating CBAM in Fig.2 and Fig.3, respectively.

2.4. *Focal Loss*

Focal Loss is a loss function that acts on class imbalance, aiming to improve model accuracy by focusing on difficult samples [7]. Compared to the commonly used cross-entropy loss, Focal Loss assigns lower weights to easily classified samples and higher weights to difficult samples. This enables the model to focus on important samples, as majority classes get lower weights while minority classes get higher weights. The formula for Focal Loss is shown below.

$$FL(p_t) = -\alpha(1 - p_t)^{\gamma} \log p_t \qquad (1)$$

p_t is the model's estimated probability for the class with a label $0 \leq p_t \leq 1$, and γ is an adjustable parameter. As p_t approaches 1, the loss becomes smaller, exhibiting the characteristic of reinforcing the focus on difficult

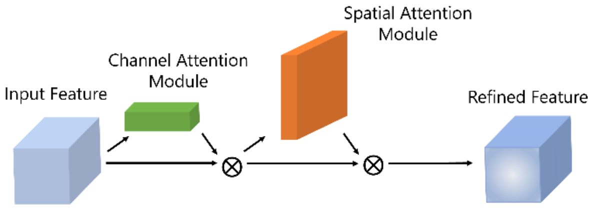


Fig. 2. CBAM

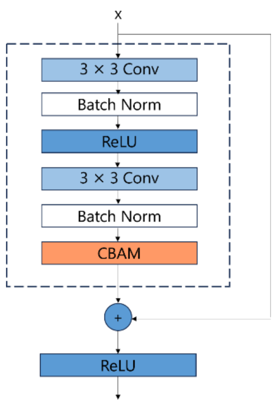


Fig. 3. CBAM is incorporated into the ResNet

samples when p_t is small. When p_t is large, p_t^{γ} becomes large for samples where p_t is high, meaning the classification is easy, thus suppressing rapid changes. This reduces the loss for majority classes or easily classified samples, making the model more robust to them. The value of γ depends on the task and dataset, so the optimal value needs to be found through experimentation.

In this paper, while CBAM is incorporated into ResNet, the dataset has imbalanced classes with the Normal class constituting half the samples. At this stage, introducing CBAM alone may emphasize the numerous Normal classes. Therefore, introducing Focal Loss enables focusing on the imbalanced classes while mitigating the influence of easy samples.

2.5. *Classification*

We use ResNet, which has achieved high performance in the field of image classification as a model to extract image features. In this paper, we propose an improved model from the ResNet34.

2.6. *Detail of dataset*

The experiments in this paper utilized the dataset employed in the ICBHI 2017 Challenge. The dataset used includes four types of respiratory data: Crackle, Wheeze, Both, and Normal, consisting of 920 audio files recorded with an electronic stethoscope from 126 patients (Table 1) (Table 2).

2.7. *Result*

Table 3 shows the results of the performance evaluation of the proposed method. As a comparison, the results are also shown for the case where only CBAM is added to the Residual Block and where only the loss function is replaced by Focal Loss.

2.8. *Discussion*

In this paper, ResNet is adopted as the base model, and the accuracy of automatic breath sound classification is improved by modifying ResNet (Table 3). First, we compare the performance of ResNet concerning the number of layers. The experimental results show that ResNet-34 achieves the highest scores in *Accuracy*, *SE*, and *ICBHI Score*. ResNet-34 has more layers than ResNet-18 and can learn more deeply. The results suggest that the deeper training than ResNet-18 enables the extraction of features that are useful for classification. On the other hand, ResNet-18 has a higher score than ResNet-34 in SP, but we believe that ResNet-34 is superior to ResNet-18 because a high SE is more important for pathological diagnosis, i.e., not to miss a possible disease.

Table 1. Label tree manually constructed for the ICBHI 2017 Challenge dataset

Class	Number
Crackle	1864
Wheeze	886
Both	506
Normal	3642

Table 2. Confusion Matrix of 4 Class

		Prediction Label				
		Crackle	Wheeze	Both	Normal	Total
True Label	Crackle	C_c	C_w	C_b	C_n	C
	Wheeze	W_c	W_w	W_b	W_n	W
	Both	B_c	B_w	B_b	B_n	B
	Normal	N_c	N_w	N_b	N_n	N

Next, we compare the results of introducing CBAM to the Residual Block. CBAM increases the $Accuracy$, SE, SP and ICBHI Score of ResNet-18 and ResNet-34, except for the SP of ResNet-18. The introduction of CBAM makes it possible to focus on features in the channel and spatial directions and extract important information in these directions.

Furthermore, we discuss the results of changing the loss function to Focal Loss, which increases the $Accuracy$ and SP scores for both ResNet-18 and ResNet-34, but decreased the accuracy of the other classes as the accuracy of Normal increases. This may be due to the fact that the attention to the Normal data, which has a relatively large number of data, is distributed to the other classes, thereby reducing overlearning on the Normal data and improving accuracy. As a result, SP is improved and SE is decreased.

Finally, we compare the experimental results of the model proposed in this paper, in which CBAM and Focal Loss are introduced to ResNet. First, the introduction of CBAM increases the expressive power and enables more precise extraction of important features in each class, but it also decreases the scores for other classes because it focuses too much on features in the Normal data with many images. However, the introduction of Focal Loss and CBAM enables us to shift attention from Normal to other classes, and we believe that this improves overall accuracy.

3. Conclusion

In this paper, we proposed a deep learning model that can classify images transformed by the short-time Fourier transform, using ResNet as the base model, with improvements such as feature extraction using CBAM and weighting of imbalance classes by introducing Focal Loss, and perform automatic classification from breath

Table 3. Result of 4-class classification

Model	*Accuracy* [%]	*SE* [%]	*SP* [%]	*ICBHI Score* [%]
ResNet-18	0.705	0.584	0.813	0.699
ResNet-34	0.716	0.613	0.807	0.711
ResNet-18 + CBAM	0.715	0.605	0.812	0.709
ResNet-34 + CBAM	0.726	0.623	0.818	0.721
ResNet-18 + Focal Loss	0.717	0.584	0.835	0.710
ResNet-34 + Focal Loss	0.717	0.586	0.834	0.710
ResNet-18 + CBAM + Focal Loss	0.725	0.604	0.832	0.718
ResNet-34 + CBAM + Focal Loss	0.732	0.607	0.843	0.725

sound data in the ICBHI 2017 Challenge Dataset. The results show $Accuracy$ of 73.2%, SE of 60.7%, SP of 84.3%, and $ICBHI\ Score$ of 72.5%, which are better than the base model, ResNet, and also the classification accuracy is improved for classes with small number of data except Normal. In the future, we are considering the use of transfer learning to further improve classification performance. It may be difficult to learn enough features for the classification because the number of data used in this study is too small to fully exploit the power of the deep learning model. Therefore, we believe that transfer learning can be introduced to compensate for the lack of data by appropriating knowledge from models previously trained on other larger data sets. In addition, by using weights from previously trained models as initial values, it is expected that model convergence will be faster and training time will be reduced.

Acknowledgements

In this paper, we used ICBHI 2017 Challenge Dataset (https://bhichallenge.med.auth.gr/ICBHI_2017_Challenge).

References

1. World Health Organization, Chronic respiratory deseases, https://www.who.int/health-topics/chronic-respiratory-diseases#tab=tab_1 (2024/06/21 accessed).
2. N. Miyashita, Y. Yamaguchi, Bacterial Pneumonia in Elderly Japanese Populations, 2018, 2018 Jan Japanese Clinical Medicine.
3. World Health Organization, The top 10 causes of death, https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death (2024/01/15 accessed).
4. S. Gairola, F. Tom, N. Kwatra, M. Jain, Respirenet: A Deep Neural Network For Accurately Detecting Abnormal Lung Sounds In Limited Data Setting, 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2021,
5. K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp.770-778.
6. S. Woo, J. Park, J. Lee, I. Kweon, CBAM: Convolution Block Attention Module, Computer Vision and Pattern Recognition, arXiv:1807.06521[cs.CV], 2018.
7. T. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, Focal Loss for Dense Object Detection, Computer Vision and Pattern Recognition, arXiv:1708.02002[cs.CV], 2018.

Authors Introduction

Mr. Ryusei Oshima

He received his Bachelor's degree in Engineering in 2021 from the Faculty of Engineering, Kyushu Institute of technology in Japan. He is currently a master student in Kyushu Institute of Technology, Japan.

Dr. Tohru Kamiya

H He received his B.A. degree in Electrical Engineering from Kyushu Institute of Technology in 1994, the Masters and Ph.D. degree from Kyushu Institute of Technology in 1996 and 2001, respectively. He is a professor in the Department of Mechanical and Control Engineering at Kyushu Institute of Technology. His research interests are focused on image processing and medical application of image analysis. He is currently working on computer aided diagnosis based on CT, MR imaging, fluorescence microscope imaging, and automatic classification of respiratory sound.

M.D. Ph.D. Shoji Kido

He received his M.D. degree from Osaka University in 1988. He received his Ph.D. degrees in Medicine and Information Science from Osaka University in 1992 and 1999, respectively. He is a guest professor of Osaka University Institute for Radiation Science and Osaka University Graduate School of Medicine. His research interests are focused on the use of artificial intelligence in radiology.

Classification of Heat Transfer Coefficient Using Deep Learning Incorporated Boiling Images

Fuga Mitsuyama
Department of Mechanical and Control Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata-ku, Kitakyushu, Fukuoka, 804-0015, Japan

Ren Umeno
Department of Mechanical and Control Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata-ku, Kitakyushu, Fukuoka, 804-0015, Japan

Kaito Takakuma
Department of Mechanical and Control Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata-ku, Kitakyushu, Fukuoka, 804-0015, Japan

Tomohide Yabuki
Department of Mechanical and Control Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata-ku, Kitakyushu, Fukuoka, 804-0015, Japan

Tohru Kamiya
Department of Mechanical and Control Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata-ku, Kitakyushu, Fukuoka, 804-0015, Japan
E-mail: mitsuyama.fuga621@mail.kyutech.jp, umeno.ren696@mail.kyutech.jp, takakuma.kaito299@mail.kyutech.jp,
yabuki.tomohide556@mail.kyutech.jp, kamiya@cntl.kyutech.ac.jp

Abstract

Boiling cooling has been used as a cooling method for electronic devices due to its high heat transfer coefficient (HTC). The regularity of the boiling phenomenon is a crucial factor in the development of more efficient cooling systems. To develop such systems, it is essential to accurately measure the HTC, which is closely related to the boiling phenomenon. In this paper, we propose a method for predicting the HTC of two different heat transfer surfaces using deep learning with boiling sound and boiling images as inputs. The proposed method achieves an accurate improvement of 2.0% and 16.7% compared to models using only boiling sound and boiling images as input, respectively.

Keywords : Boiling Sound, Boiling Image, Heat Transfer Coefficient (HTC), Convolutional Neural Network (CNN), HyPR framework.

1. Introduction

As CPUs and GPUs become more highly integrated, their heat generation density continues to increase every year [1]. Under these circumstances, boiling water-cooling methods are attracting significant attention. Boiling cooling provides a significantly higher heat transfer coefficient (HTC) compared to conventional gas forced convection cooling methods using heat sinks and fans [1]. In addition, the use of water as the boiling medium ensures environmental sustainability and cost effectiveness. However, because boiling absorbs heat through the latent heat of liquid evaporation and convection along with bubble movement [2], HTC is largely dependent on the number and size of bubbles, which vary significantly depending on the surface roughness and wettability of the boiling surface. To develop more efficient and reliable cooling systems, it is crucial to clarify the relationship and regularity between the boiling behavior and the HTC, as well as to accurately measure the HTC, regardless of the properties of the boiling surface. In this paper, we propose

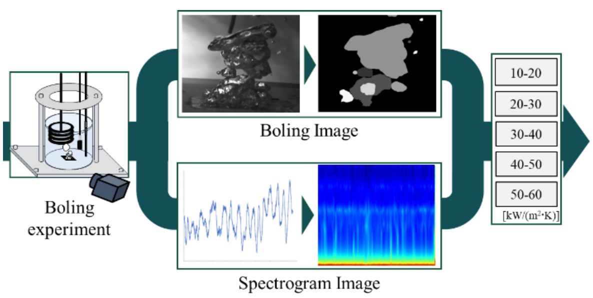


Fig 1. Overview of the proposed method with two input data

a method to predict HTC of two heat transfer surfaces with different nucleation site densities using a convolutional neural network (CNN) with two types of inputs: a spectrogram image of the boiling sound and a boiling image of the boiling behavior as shown in Fig. 1. The proposed method aims to develop a highly accurate prediction model by complementing each other with features that are not adequately captured by either the spectrogram or the boiling image alone.

2. Method

2.1. *Creation of dataset*

We conducted a boiling experiment to generate a data set for our proposed multimodal system. First, the sound data was processed to generate spectrogram images. The amplitude spectrum of the sound data acquired by the hydrophone [3] in the experimental setup as shown in Fig. 2 was calculated using a short-time Fourier transform (STFT) [4]. A 0.6 mm thick sapphire substrate with a sputtered titanium thin-film heater on its bottom surface is used as a heated wall as shown in Fig. 3. Additionally, artificial nucleation sites were formed on the heat transfer surface by patterning the superhydrophobic coating Glaco using photolithography. In this study, boiling experiments were conducted using two heat transfer surfaces with different nucleation site densities as shown in Fig. 3. Before the boiling experiment, the pure water used as the boiling medium was degassed for more than 1 hour through boiling degassing.

Harmonic/Percussive Sound Separation (HPSS) [5] was then applied to separate the harmonic and percussive components for the spectrogram. Then, a process called Linear Predictive Coefficient (LPC) [6] analysis was performed to generate spectrogram images. The boiling images were captured at a rate of 100 frames per second by the high-speed camera [7] shown in Fig. 2. The representative data of the boiling image corresponding to a given HTC was defined as the boiling image at the center time of each spectrogram in order to map it to the spectrogram data. The number of data for each class is given in Table 1.

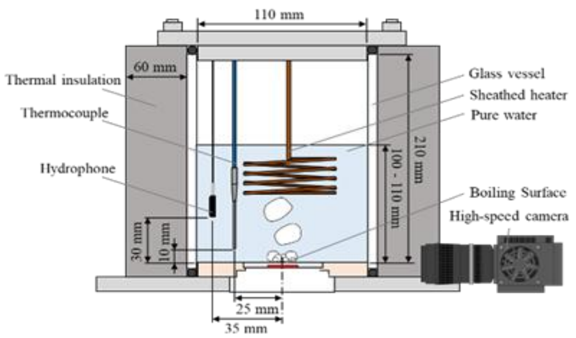


Fig 2. Overview of boiling sound experimental setup

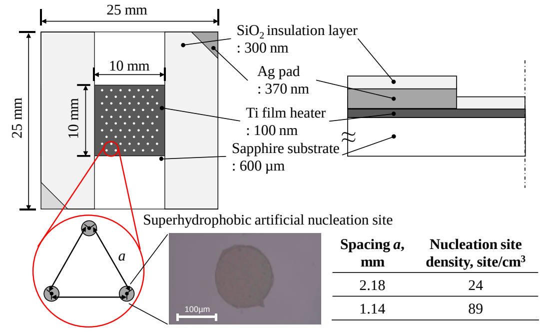


Fig 3. Sapphire substrate of the heated wall featuring superhydrophobic artificial nucleation sites

Table 1 The number of data in each class

Class [kW/(m^2·K)]	Number
10-20	960
20-30	1000
30-40	1000
40-50	980
50-60	600
Total	4540

2.2. *Classification for HTC*

The HTC prediction model proposed in this paper is trained using new features that integrate features obtained from spectrogram images using methods based on previous research [8] and Tabata et al. [9] and features obtained from boiling images using the method proposed by Suh et al. [10]. An overview of the entire system is shown in Fig. 4. The structure of each model is described below. We specify the labels for each class in the range of 10-20, 20-30, 30-40, and 40-50 [kW/(m^2·K)]. These label numbers represent the HTC.

2.2.1. *CNN with spectrogram*

The lower part of Fig. 4 shows the feature extraction model using spectrograms. The original spectrogram and the percussive component spectrogram generated by applying the HPSS to the original spectrogram were used as inputs. This is since the main component of the boiling sound is the percussive component, which is intermittent in the formation and bursting of bubbles and is considered to be an important element of the sound. A basic block of the CNN consists of a convolution layer (with a kernel size of 3 x 3 and 96 channels), batch normalization, rectified linear unit (ReLU) functions, and average pooling. Four stacks of these blocks then downsample the input images and construct a noise robust model.

2.2.2. *HyPR framework*

The upper part of Fig. 4 shows the feature extraction model using boiling images, where VGG16 [11] and the instance segmentation models MaskR-CNN [12] and Multi-Layer Perceptron (MLP) are used to extract features from the boiling images themselves and statistical data such as the number and size of bubbles, respectively, to

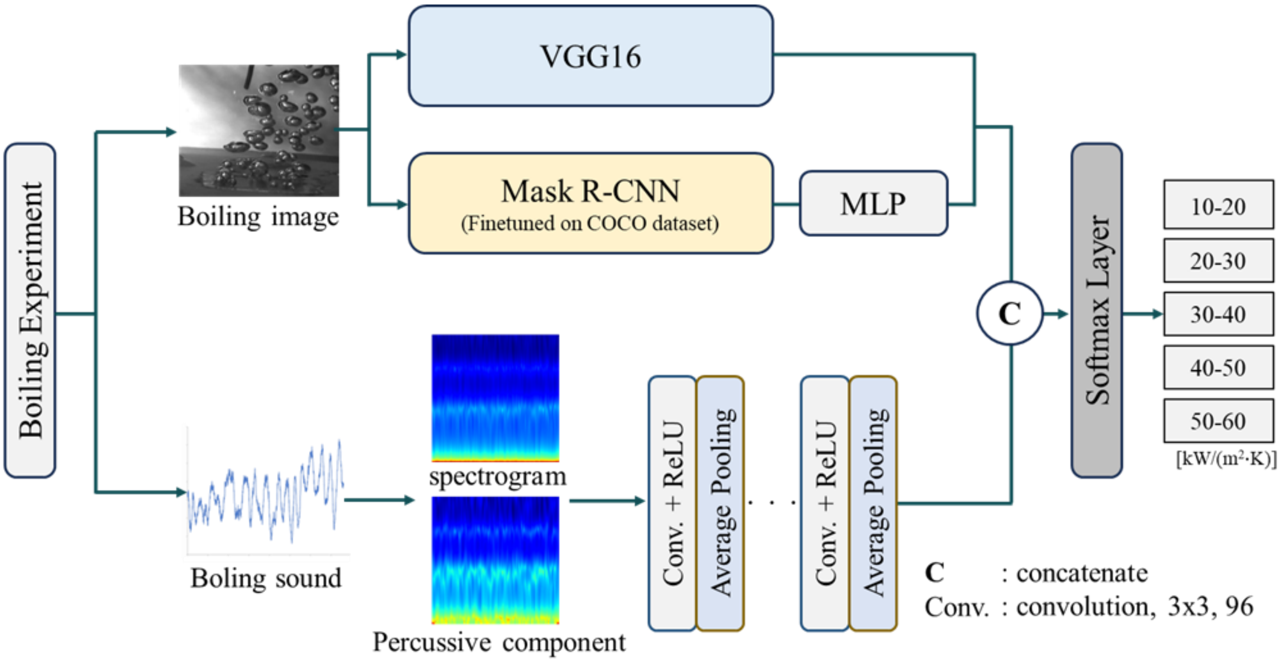


Fig 4. Detailed structure of the proposed method model

learn the nonlinear relationships between the images. We use a fine-tuned model in the COCO dataset for MaskR-CNN.

3. Experimental Result and Discussion

3.1. *Evaluation*

The generalization performance of the model is evaluated using fivefold cross-validation. The evaluation metric is accuracy, defined as the proportion of correct and predicted labels that match. The average accuracy is the mean of the accuracy for the validation on each dataset.

3.2. *Result*

An average accuracy of 90.2% was achieved with the proposed method CNN(spec)+HYPR model, the multimodal system using boiling images and spectrograms. When HTCs were predicted by each model alone, the accuracy was 88.2% for the HYPR model and 73.5% for the CNN(spec) model, an improvement of 2% and 16.7%, respectively.

3.3. *Discussion*

We discuss the classification results obtained in Section 3.2. First, Table 2 shows that the HyPR model achieved high accuracy despite using only boiling images as input, although the values for datasets 1 and 5 were low. This is probably since the input was a boiling image at a given time, which could not adapt to rapid changes in the bubbles, such as bubbles merging or separating. Figure 5 shows an actual boiling image that was misclassified by the HyPR model. On the other hand, the proposed method consistently achieved high classification accuracy on most datasets and was able to correctly classify data between adjacent classes, which was not possible with the HyPR model. The reason for this may be that the spectrogram images contain information within a certain time before and after each boiling image, which makes it easier to distinguish between other classes of images that suddenly

become similar. In other words, while maintaining the high classification accuracy of the HyPR model, the addition of spectrogram information reduces the misclassification of adjacent classes and improves the accuracy of the model.

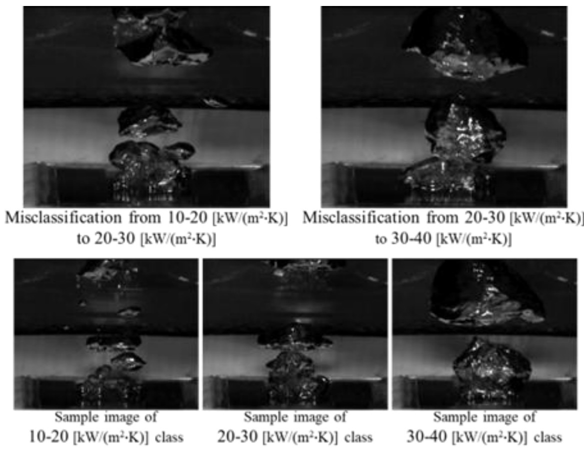


Fig 5. Comparison of misclassified images and boiling images at each HTC

Table 2 Comparison of accuracy for each model

	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Average acc [%]
CNN (spec) + HyPR	0.9	0.832	0.931	0.977	0.87	0.902
CNN (spec)	0.821	0.853	0.69	0.673	0.637	0.735
HyPR	0.842	0.859	0.92	0.977	0.811	0.882

4. Conclusion

The proposed methods using multimodal systems resulted in an improvement in accuracy of 16.7% and 2.0%, respectively, compared to the conventional methods of CNN(spec) and HyPR model, despite the difference in heat transfer surface properties. In future work, we would like to improve the segmentation accuracy of MaskR-CNN in the HyPR framework by adding image processing, such as contrast enhancement of the boiling images. The CNN(spec) model contains only one second of boiling sound information in the spectrogram image, which may be insufficient for accurate analysis. To improve the accuracy of the model, it would be beneficial to include data from a longer period. In addition, modifying the feature integration method could further improve the performance of models.

Reference

1. N. Unno, "Towards the Realization of Boiling Cooling Devices Using Bubble Refinement Boiling *Kihoubisaikahuttou wo tukatta huttoureikyakudebaisu jitugen ni mukete*", JSME TED Newsletter, No.89, pp. 2-4, 2019. (In Japanese)
2. S. Yajima, et al., "Heat flux partitioning and macrolayer observation in pool boiling of water on a surface with artificial nucleation sites", International Journal of Heat and Mass Transfer, 194, 122924, 2022.
3. HBK, "MINIATURE HYDROPHONE TYPE 8103", https://www.bksv.com/-/media/literature/Product-Data/bp0317.ashx, Nov. 22, 2023.
4. J. B. Allen, "Short Term Spectral Analysis, Synthesis. And Modification by Discrete Fourier Transform", IEEE Transactions on Acoustic, Speech, and Signal Processing, Vol.25, No.3, pp. 235-238, 1977.
5. D. Fitzgerald, "Harmonic/Percussive Separation using Median Filtering", 13th International Conference on Digital Audio Effects (DAFx-10), 2010.
6. H. Mukherjee et al., "MISNA : A Musical Instrument Segregation System from Noisy Audio with LPCC-S Features and Extreme Learning", Multimedia Tools and Applications, Vol.77, pp.27997-28022, 2018.
7. IDT Vision, "Model OS8", https://idtvision.com/support/specifications-sheets/?output=pdf&idt_id=OS8-V3-S1, Dec. 10, 2024.
8. F. Mitsuyama et al., "Deep Learning Based Prediction of Heat Transfer Coefficient Using Spectrogram Images from Boiling Sound", Proceedings of the 2024 International Conference on Artificial Life and Robotics, pp. 914-917, 2024.
9. M. Tabata, H. Lu, T. Kamiya, S. Mabu, S. Kido, "Automatic Classification of Respiratory Sound Considering Hierarchical Structure", The 22nd International Conference on Control, Automation and Systems, pp. 537-541, 2022.
10. Y. Suh, et al., "Deep Learning Predicts Boiling Heat Transfer", Nature Scientific Report 11, 5622, 2021.
11. K. Simonyan, et al., "Very Deep Convolutional Networks for Large-Scale Image Recognition", arxiv, 2015.
12. K. He, et al., "Mask R-CNN", Proceedings of the IEEE international conference on computer vision, pp. 2961-2969, 2017.

Authors Introduction

Fuga Mitsuyama

He received his Bachelor's degree in Engineering in 2023 from the Faculty of Engineering, Kyushu Institute of technology in Japan. He is currently a master student in Kyushu Institute of Technology, Japan.

Ren Umeno

He received his Master's degree in Engineering in 2023 from the Faculty of Engineering, Kyushu Institute of technology in Japan. He is currently a doctoral student in Kyushu Institute of Technology, Japan.

Kaito Takakuma

He is currently a bachelar student in Kyushu Institute of Technology, Japan.

Tomohide Yabuki, Prof. Ph.D

He is a Professor in the Department of Mechanical and Control Engineering at Kyushu Institute of Technology and is a PRESTO researcher of Japan Science and Technology Agency (JST). His research interests are in the areas of boiling heat transfer, MEMS thermal measurement techniques, and single cell thermal analysis.

Tohru Kamiya, Prof. Ph.D

He received his B.A. degree in Electrical Engineering from Kyushu Institute of Technology in 1994, the Master and Ph.D. degree from Kyushu Institute of Technology in 1996 and 2001, respectively. He is a professor in the Department of Mechanical and Control Engineering at Kyushu Institute of Technology. His research interests are focused on image processing, medical application of image analysis. He is currently working on computer aided diagnosis based on CT, MR imaging, fluorescence microscope imaging, and automatic classification of respiratory sound.

A Data Format Integration of Open-Street-Map and Lanelet2 Toward the Ontology Framework for Safety Automated driving systems

Obada Al Aama
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Takahiro Koga
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Tomoki Taniguchi
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Davaanyam Jargal
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Junya Oishi
Aisan Technology Co., Ltd., AT Building, 7-14 Nishiki 3-chome, Naka-ku, Nagoya, 460-0003, Japan

Shigeru Nemoto
Aisan Technology Co., Ltd., AT Building, 7-14 Nishiki 3-chome, Naka-ku, Nagoya, 460-0003, Japan

Wataru Mizushina
Aisan Technology Co., Ltd., AT Building, 7-14 Nishiki 3-chome, Naka-ku, Nagoya, 460-0003, Japan

Kazuki Hirao
Aisan Technology Co., Ltd., AT Building, 7-14 Nishiki 3-chome, Naka-ku, Nagoya, 460-0003, Japan

Hakaru Tamukoh
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hiroaki Wagatsuma
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

E-mail: al-aama.obada495@mail.kyutech.jp, koga.takahiro215@mail.kyutech.jp, taniguchi.tomoki352@mail.kyutech.jp, Jargal.davaanyam711@mail.lyutech.jp, j.oishi@aisantec.co.jp, s.nemoto@aisantec.co.jp, w.mizushina@aisantec.co.jp, k.hirao@aisantec.co.jp, tamukoh@brain.kyutech.ac.jp, waga@brain.kyutech.ac.jp

Abstract

This study proposes a framework that integrates OpenStreetMap (OSM) data with ontology-based systems to enhance automated driving. OSM provides static geographical data, while the Lanelet2 mapping framework incorporates lane-level road information and topological relationships. This combination enables advanced testing of vehicle behavior in realistic environments. Ontology-based integration offers semantic representations of road elements and traffic rules, supporting the modeling of complex driving scenarios. By structuring spatial and semantic data, this approach ensures accurate simulation and testing, facilitating applications such as traffic analysis, route optimization, and automated driving decision-making. The framework enhances scalability, precision, and safety in autonomous vehicle development, enabling more effective testing and validation. This integration supports comprehensive, context-aware simulations that improve vehicle response to real-world driving conditions.

Keywords: OSM map, Lanelet2, Ontology, Coincar simulator, Automated driving.

1. Introduction

Road traffic crashes are a major cause of death, driven by dangerous behaviors like speeding, drunk driving, driver fatigue, and distractions such as mobile phone use. Autonomous vehicles (AVs) aim to address these issues by utilizing advanced technologies like computer vision, GPS, sensors, and mapping systems to navigate, detect environments, and determine optimal routes. By reducing human error, AVs offer potential benefits such as fewer accidents, improved traffic flow, lower fuel consumption, and enhanced mobility for all road users [1]. Maps are critical for autonomous vehicle systems, especially in highly automated driving, where High-Definition (HAD) maps [2] provide detailed, precise environmental data. These maps facilitate object prediction, routing, and the generation of driving behavior by offering high-resolution information essential for understanding and interacting with surrounding entities like vehicles and pedestrians. They are integral to route planning by providing complete road network information to determine efficient paths. Furthermore, HAD maps support the design and implementation of advanced driving strategies and maneuvers, ensuring autonomous vehicles operate safely, efficiently, and effectively.

Lanelet2 [3] is a sophisticated mapping framework designed to support autonomous vehicles by providing advanced map elements and enabling the modeling of complex driving scenarios. It offers a flexible and extensible structure, allowing for the representation of lane-level information, traffic rules, and dynamic road elements. This level of detail is crucial for the evolving needs of automated driving systems, as it enables precise localization, path planning, and decision-making in dynamic environments. Decision-making in autonomous vehicles extends beyond just mapping physical elements of the road. In addition to representing static road features, autonomous vehicles must make informed decisions about how to safely and efficiently interact with other road users.

To further enhance the functionality of Lanelet2, an ontology framework can be employed to define relationships between concepts and entities, facilitating knowledge sharing and semantic representation [4]. This structured approach improves the understanding of surroundings, supports advanced reasoning, and enables more effective interaction with map data. By integrating semantic knowledge and context-aware reasoning, Lanelet2 contributes to a more adaptive, intelligent, and effective driving system.

This study proposes a framework that integrates Lanelet2 and OpenStreetMap (OSM) data with ontology-based systems to enhance automated driving systems. By leveraging semantic reasoning and context-aware insights, the framework aims to improve navigation precision, predict vehicle behavior, and support complex decision-making. This integration enables autonomous systems to better interpret and respond to dynamic driving environments, promoting safer, smarter, and more adaptive vehicle operations. The use of simulator facilitates efficient testing and optimization of automated driving strategies.

2. Methodology

The proposed framework follows a three-step process that involves data handling, simulation, and ontology integration to create a comprehensive and realistic environment for modeling real-world driving scenarios as shown in Fig. 1.

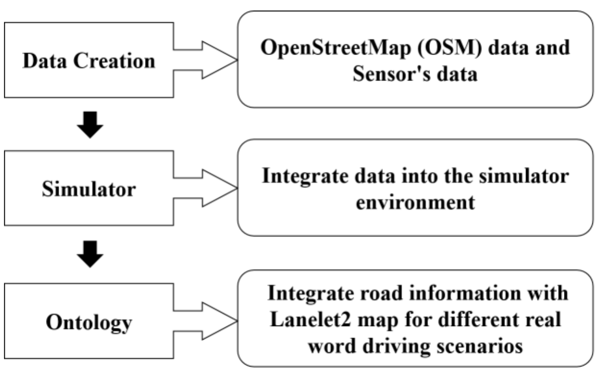


Fig. 1. Flowchart of Lanelet2 with Ontology Integration.

2.1. *Data creation*

OpenStreetMap (OSM) Data serves as the primary source of geographical information, providing crucial static map elements such as road networks, intersections, lane markings, and landmarks. This information is essential for defining the structure of the driving environment. Real-world sensor data, such as inputs from cameras, LIDAR, and radar, is incorporated to add dynamic elements to the simulation. This data enriches the environment with critical information on traffic flow, road obstacles, pedestrians, and other moving entities that influence vehicle behavior. The data required for this step, including OSM data and sensor data, will be supplied by Aisan Technology Co., Ltd., ensuring accuracy, consistency, and high-quality input for the simulation process.

2.2. *Coincar simulator*

Coincar simulator provides a controlled, virtual platform to simulate realistic driving scenarios and assess the performance of autonomous systems in handling complex interactions with other road users [5]. By using Coincar simulator, the impact of Lanelet2's mapping precision and ontology-driven decision-making on vehicle behavior, safety, and efficiency can be analyzed. This simulation-based validation process plays a crucial role in refining and optimizing autonomous vehicle decision-making, ultimately contributing to the development of a more adaptive, intelligent, and effective driving system.

2.3. *Lanelet2-Ontology integration*

The Lanelet2 map creation process starts with defining geographic points, which are connected to form linestrings outlining road boundaries. Enclosed linestrings create areas to represent regions like pedestrian zones or parking lots. Lanelets, which define drivable paths, are formed using two parallel linestrings as left and right boundaries to represent directional traffic flow. Finally, regulatory elements are added to link traffic rules, such as speed limits and stop signs, to the map components, enabling an accurate representation of real-world driving environments.

Lanelet2 maps are divided into three key layers. The physical layer defines the basic geometry with points and linestrings to represent physical features like road boundaries. The relational layer builds on this by grouping these elements into higher-level structures, such as lanes, areas, and road rules [6]. The topological layer establishes connectivity and spatial relationships between these elements, allowing for route planning and interaction between different components. Together, these layers create a comprehensive and navigable road network for autonomous vehicles, enabling precise and reliable navigation [7]. Fig. 2 illustrates the architecture of Lanelet2, where lanelets are represented by capital letters, areas by lowercase letters, and linestrings by numbers.

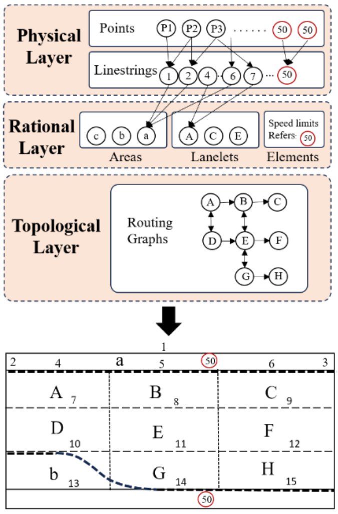


Fig. 2. Lanelet2 map architecture for automated driving systems.

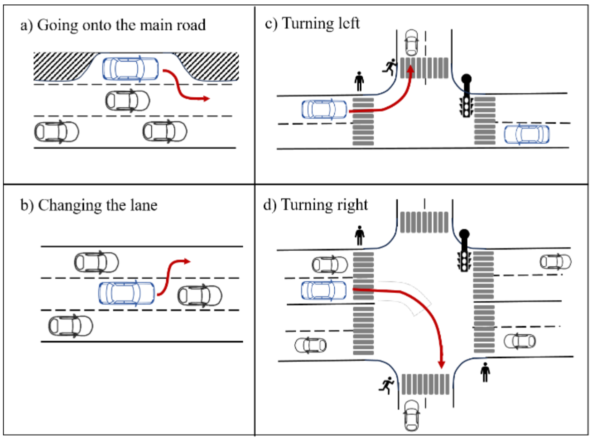


Fig. 3. Examples of real-world driving scenarios.

Integrating Lanelet2 with ontology-based systems enhances automated driving by mapping Lanelet2 elements like lanelets, areas, and regulatory elements to ontology classes and properties. This approach allows for more effective modeling of relationships and dependencies between road elements, enabling complex tasks such as traffic rule enforcement, route optimization, and behavior prediction. Fig. 3 shows examples of simple and complex real-world driving scenarios.

The figure illustrates four driving scenarios for autonomous systems. Scenario (a) involves merging into traffic, requiring gap detection and safe merging. Scenario (b) depicts lane changes in a traffic jam, requiring distance adjustments and smooth maneuvers. Scenario (c) shows intersection entry, where the system must recognize pedestrians, vehicles, and traffic lights. Scenario (d) highlights a complex intersection, requiring advanced decision-making to determine the optimal timing and action.

3. Experiment and discussion

Fig. 4 illustrates a map created using Vector-Map-Builder. The map represents a road network with detailed geometry, including curves and intersections, offering a high-resolution depiction suitable for analysis and editing within the Java OpenStreetMap (JOSM) editor [8]. The map contains detailed elements, including 55,998 points, 3,082 linestrings (roads or paths), 714 polygons (enclosed areas), 806 lanelets (individual lanes), and 91 regulatory elements (e.g., traffic signs).

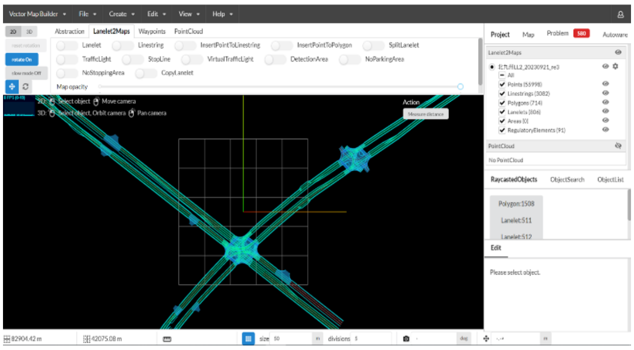


Fig. 4. Visualization of OSM map created using Vector-Map-Builder.

Coincar Simulator utilizes Rviz, a 3D visualization tool provided by the Robot Operating System (ROS) [9], to enhance development and evaluation processes for cooperative motion planning in autonomous vehicles. Widely used in the robotics community, Rviz enables visualization and debugging of various aspects of robot operation. It also supports extensions through custom plugins to visualize specific ROS messages, facilitating more tailored and effective visualization. A notable feature of Coincar simulator is its visualization capability, which includes the use of color-coded trajectories. In this system, each vehicle's planned future trajectory is displayed with a color gradient representing time progression. Overlapping trajectory segments of the same color indicate potential future collisions, providing an intuitive means to assess and debug vehicle interactions within the simulation environment. To represent the environment, plugins will be created to visualize the Lanelet map, object states, and desired motion using trajectories. These trajectories will utilize a color-coded scheme to represent absolute time, following a method similar to that outlined in [10]. Fig. 5 illustrates the visualization of the Lanelet map and color-coded trajectories in the Coincar simulator.

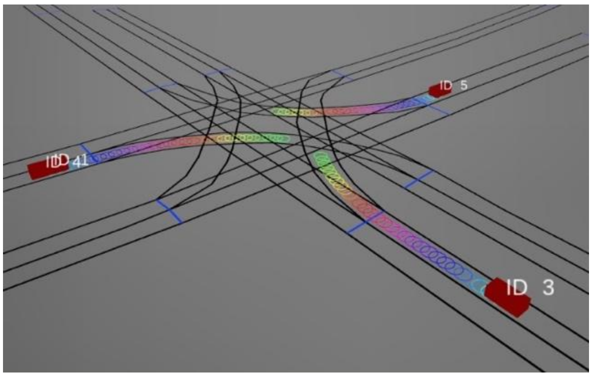


Fig. 5. Lanelet map and color-coded trajectories in Coincar simulator.

Localization management relies on Cartesian coordinates, while real-world locations exist on the Earth's curved surface. To bridge this discrepancy, the Universal Transverse Mercator (UTM) projection is employed, enabling accurate alignment between real-world locations and their corresponding map representations. An example of a right-turn scenario for autonomous systems at an intersection, represented using Rviz, is illustrated in Fig. 6. This representation will be applied to the created map to visualize and simulate the autonomous system's behavior during the turn. To navigate this intersection safely, the system must make dynamic decisions, including yielding to pedestrians at crosswalks, adjusting vehicle speed to maintain safe distances, and planning trajectories to execute turns efficiently. Traffic rules, such as adherence to signals and yielding requirements, are integrated into the decision-making process. This structured map ensures that the autonomous system prioritizes safety by avoiding collisions, complies with traffic laws, and operates efficiently by minimizing delays while adhering to all safety protocols

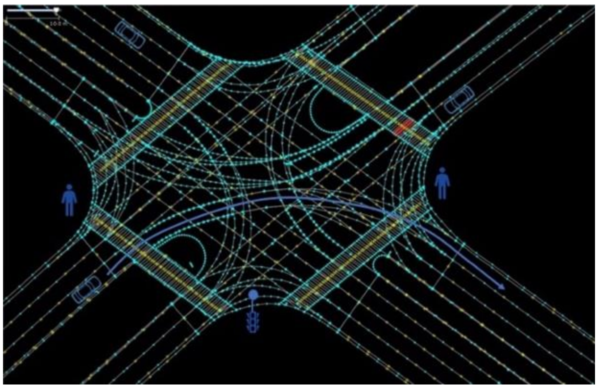
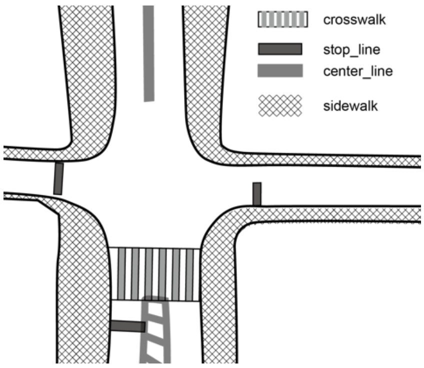


Fig. 6. Safe and efficient navigation strategies for autonomous vehicles in multi-lane intersection.

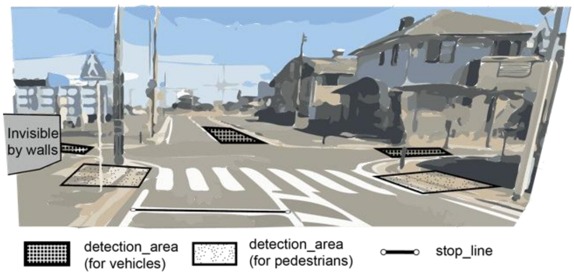
4. Experimental design and discussion

This study presents a comprehensive framework for enhancing autonomous vehicle navigation, decision-making, and safety through the integration of Lanelet2, OSM data, and ontology-based systems. By leveraging semantic reasoning and context-aware insights, the framework facilitates precise localization, route planning, and adaptive driving behavior. Lanelet2's multi-layered architecture, combined with ontology integration, enables the modeling of complex scenarios, such as lane changes, merging, and intersection navigation.

In the intersection especially without any traffic signal, the map representation is highly important for automated driving to ensure its safety. For example, as shown in Fig. 7, the detection area must be defined clearly with respect to the stop line. The vehicle can sense surrounding obstacles in the pathway to travel, and it requires to pay attention to targets that may cause a potential accident. In the Lanelet [7], it can be embedded as “Detection Area” in the “Extra Regulatory Elements” by the extension of Autoware [11] data format. We focused on the difference in vehicle behaviors including its decision-making process, depending on types of targets such as vehicles and pedestrians, and proposed the differentiation of detection areas depending on types of targets, which were differently categorized in the ontology.



(a) Map view



(b) Perspective view

Fig. 7. Detection area and stop line are represented by the set of points of interest.

As shown in Fig. 8 and Fig. 9, the detection area is described in Lanelet [7], and the automated driving system can pay attention to the area. However, the safety strategy might be different in cases of vehicles and pedestrians (Fig. 10).

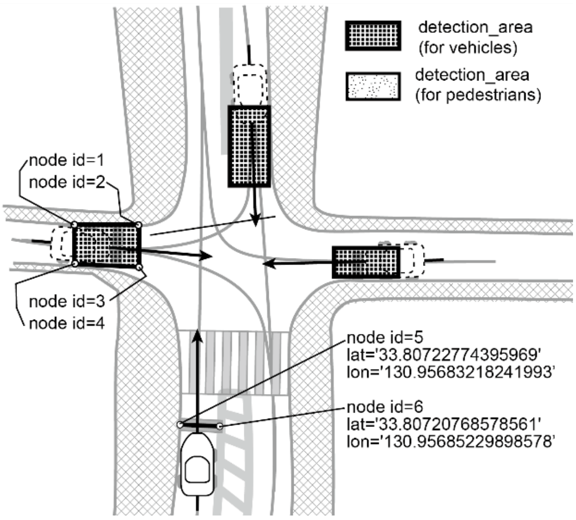


Fig. 8. A map view with detection areas considering pathways of vehicles.


```
<node id=1 version='1' lat='33.80734177738309' lon='130.95688633086124'>
  <tag k=’ele’ v=’0’/>
</node>
<node id=2 version='1' lat='33.807319490552416' lon='130.95690711797926'>
  <tag k=’ele’ v=’0’/>
</node>
<node id=3 version='1' lat='33.807304446938424' lon='130.95688566030907'>
  <tag k=’ele’ v=’0’/>
</node>
<node id=4 version='1' lat='33.80732840528539' lon='130.9568615204301'>
  <tag k=’ele’ v=’0’/>
</node>
<node id=5 version='1'  lat='33.80722774395969' lon='130.95683218241993'>
  <tag k=’ele’ v=’0’/>
</node>
<node id=6 version='1' lat='33.80720768578561' lon='130.95685229898578'>
  <tag k=’ele’ v=’0’/>
</node>
<way id=11 version='1'>
  <nd ref=1 />
  <nd ref=2 />
  <nd ref=3 />
  <nd ref=4 />
  <nd ref=1 />
  <tag k='type' v='detection_area' />
  <tag k='area' v='yes' />
</way>
<way id=12 version='1'>
  <nd ref=5 />
  <nd ref=6 />
  <tag k='type' v='stop_line' />
</way>
<relation id='13'>
   <tag k='type' v='regulatory_element'/>
   <tag k='subtype' v='detection_area'/>
   <member type='way' ref='11' role='refers'/>
   <member type='way' ref='12' role='ref_line'/>
</relation>
```


Fig. 9. Lanelet XML representation of detection areas in Fig. 8.

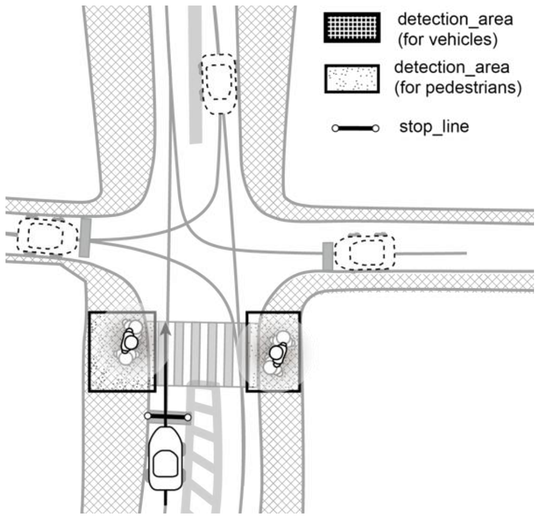


Fig. 10. Semantically differentiated detection area for pedestrians due to unexpected behaviors.

In Lanelet [7], individual vehicles travel on the lane as the transition of a lane ID to other lane ID (Fig. 11). For avoidance of collisions with other vehicles, the behavior of the automated driving will be calculated on potential actions of behaviors on the lanes. On the other hand, the risk level is different if it is a pedestrian.

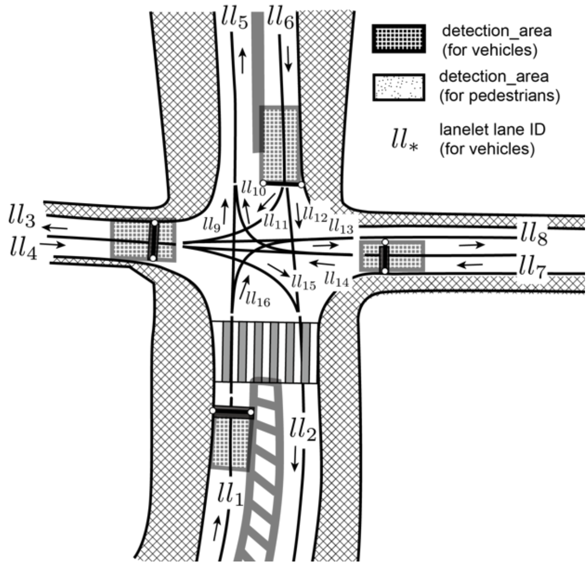


Fig. 11. Lanelet map with Lane IDs for vehicles.

By newly defining of the lane ID and detection areas specifically for pedestrians, the collision avoidance procedure, or strategy can clearly differentiate with respect to vehicle interactions (Fig. 12).

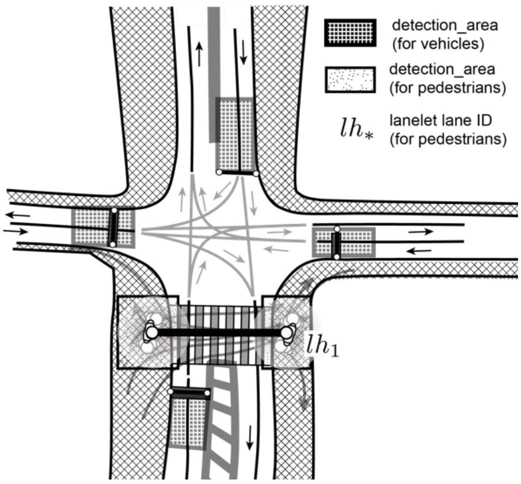


Fig. 12. An extended lane ID to cover potential pedestrian behaviors around the crosswalk.

Future work could integrate Lanelet representation with the ontology-based logical reasoning with different safety strategy behaviors depending on targets to interact. In this sense, the computer simulation is possible in a dynamic environment considering not only vehicle but also vulnerable road users such as pedestrians, wheel chairs, bicycles and so on, which requires different safety strategy to minimize potential risk. For the evaluation of automated driving systems to test overall performance, the real-world driving scenarios can be generated dynamically from the refined map data and ontology-based logical reasoning if they are successfully integrated.

Acknowledgements

This work was supported in part by Kitakyushu Foundation for the Advancement of Industry, Science and Technology (FAIS) and cooperative research projects between Kyushu Institute of Technology and Aisan Technology Co., Ltd. Authors also deeply appreciate expert drivers and safety operation managers in Nishitetsu group for offering a precious opportunity to examine professional procedures for the safety assessments in the real driving on the road.

References

1. Bathla G, Bhadane K, Singh RK, Kumar R, Aluvalu R, Krishnamurthi R, Kumar A, Thakur RN, Basheer S. Autonomous vehicles and intelligent automation: Applications, challenges, and opportunities. Mobile Information Systems, pp. 7632892, 2022.
2. Zheng S, Wang J. High definition map-based vehicle localization for highly automated driving: Geometric analysis. In2017 International Conference on Localization and GNSS (ICL-GNSS),IEEE, (pp. 1-8), 2017.
3. Naumann A, Hertlein F, Grimm D, Zipfl M, Thoma S, Rettinger A, Halilaj L, Luettin J, Schmid S, Caesar H. Lanelet2 for nuscenes: Enabling spatial semantic relationships and diverse map-based anchor paths. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (pp. 3247-3256), 2023.
4. Zhao L, Ichise R, Yoshikawa T, Naito T, Kakinami T, Sasaki Y. Ontology-based decision making on uncontrolled intersections and narrow roads. In2015 IEEE intelligent vehicles symposium (IV), (pp. 83-88), 2015.
5. Naumann M, Poggenhans F, Lauer M, Stiller C. Coincar-sim: An open-source simulation framework for cooperatively interacting automobiles. In2018 IEEE Intelligent Vehicles Symposium (IV), (pp. 1-6). 2018.
6. Bender P, Ziegler J, Stiller C. Lanelets: Efficient map representation for autonomous driving. In2014 IEEE Intelligent Vehicles Symposium Proceedings, (pp. 420-425). 2014.
7. Poggenhans F, Pauls JH, Janosovits J, Orf S, Naumann M, Kuhnt F, Mayr M. Lanelet2: A high-definition map framework for the future of automated driving. In2018 21st international conference on intelligent transportation systems (ITSC), IEEE, (pp. 1672-1679), 2018.
8. Java OpenStreetMap Editor. Available online: https://josm.openstreetmap.de/ (accessed on 25 December 2024).
9. Quigley M, Conley K, Gerkey B, Faust J, Foote T, Leibs J, Wheeler R, Ng AY. ROS: an open-source Robot Operating System. InICRA workshop on open source software, (Vol. 3, No. 3.2, p. 5), 2009.
10. Kumar A, Wagatsuma H. A Kamm's Circle-Based Potential Risk Estimation Scheme in the Local Dynamic Map Computation Enhanced by Binary Decision Diagrams. Sensors. 22(19):7253, 2022.
11. Autoware Documentation. Available online: https://tier4.github.io/autoware.iv/pulls/2409/map/lanelet2_extension/docs/extra_regulatory_elements/ (accessed on 25 December 2024).

Authors Introduction

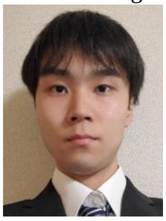
Obada Al Aama

He graduated from the Department of Communication and Electronics Engineering at Al-Baath University, Syria, in 2013. He obtained his M.Eng. and Ph.D. degrees from the Kyushu Institute of Technology, Japan, in 2019 and 2024, respectively. He is currently a researcher in the Department of Human Intelligence Systems at the Kyushu Institute of Technology, Japan. His research interests focus on robotics and automated driving.

Takahiro Koga

He received his Master's degree in Engineering from the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology in Japan. He is currently a researcher in Kyushu Institute of Technology.

Tomoki Taniguchi



He received his Bachelor's degree in Computer Science and Systems Engineering from the Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, Japan, in 2024. He is currently a master's student at Kyushu Institute of Technology, Japan.

Davaanyam Jargal

He received his Master's degree from the School of Mechanical Engineering and Transportation, Mongolian University of Science and Technology, Mongolia in 2010. He is currently a Doctoral course student in Kyushu Institute of Technology, Japan.

Junya Oishi

Graduated from Doshisha University. After working at an auto parts company, he joined Increment P Co., Ltd. (currently Geotechnologies Co., Ltd.), a digital map company. Participated in a national project and the study of specifications for high-precision maps for automated driving. After that, he joined AISAN Technology Co., Ltd. to working on the self-driving business and high-precision maps for transportation companies such as buses.

Shigeru Nemoto

Supervises the high-precision map production team, AISAN Technology Co., Ltd.

Wataru Mizushina

In charge of planning mobility business, AISAN Technology Co., Ltd.

Kazuki Hirao

In charge of planning and sales of high-precision maps, AISAN Technology Co., Ltd.

Hakaru Tamukoh

He received a B.Eng. degree from Miyazaki University, Japan 2001. He received his M.Eng and Ph.D. from Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a postdoctoral research fellow of the 21st-century Center of Excellent Program at Kyushu Institute of Technology, from April 2006 to September 2007. He was an assistant professor at Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He was an associate professor from February 2013 to March 2021 and is currently a professor at the Graduate School of Life Science and System Engineering, Kyushu Institute of Technology, Japan. His research interests include hardware/software complex systems, digital hardware design, neural networks, soft computing, and computing service robots. He is a member of IEICE, SOFT, JNNS, IEEE, JSAI, and RSJ.

Hiroaki Wagatsuma

He received his M.S., and Ph.D. degrees from Tokyo Denki University, Japan, in 1997 and 2005, respectively. In 2009, he joined Kyushu Institute of Technology, where he is currently a Professor at the Department of Human Intelligence Systems. His research interests include non- linear dynamics and robotics. He is a member of IEEE.

Developing a Sound-Based Method to Synchronize Multiple Videos Recorded by Multiple Sound Sources

Davaanyam Jargal
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Rena Kato
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Tomoki Taniguchi
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Kosei Shibata
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Takahiro Koga
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Obada Al Aama
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hakaru Tamukoh
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hiroaki Wagatsuma
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Email: jargal.davaanyam711@mail.kyutech.jp, waga@brain.kyutech.ac.jp

Abstract

The surround environmental monitoring from a vehicle on the road is important for the driving behavior analysis as well as the monitoring of driver's operations. A light weight mobile camera is useful for the record of multiple directions from the vehicle simultaneously. However, the timing synchronization is an important issue need to be solved. In this purpose, we proposed the sound-based method to synchronize different videos recorded with environmental sounds. In this task, the extraction of common sound features and amplifying of the features are necessary to superimpose those sound profiles to find consisting time points. In the validation the effectiveness, we used recoded videos from the bus driven by an expert driver.

Keywords: Wavelet, Bandpass filter, Gaussian filter, Autocorrelation, Synchronize

1. Introduction

Synchronizing video from multiple cameras is essential for various applications, from video surveillance to scientific research, where time alignment can dramatically influence data analysis and interpretation. Many traditional techniques utilize hardware-based synchronization via either GPS or timestamping, which can be costly and vulnerable to environmental disruptions. Misaligned timing the analysis of events and makes correlation and integration of video and sensor data difficult (e.g. to analyze driver behavior and for object detection). While video capture has become ubiquitous through portable devices, synchronizing footage from multiple unlinked perspectives remains a challenge. In challenging multi-video synchronization, Wu et al. [1] proposed a deep learning framework which merges pose-matching with temporal encoding to align the video streams. Although their method uses visual information in an innovative way, it requires a lot of technical power to perform real-time calculations. Brassarote et al. [2] have shown the benefit of using non-decimated wavelet transforms (NDWT) for performing shift-invariant analysis, which is a method that is attractive for non-stationary signals. Wavelet-based approaches are computationally inexpensive and conserves important aspects of the signal [3]. Calibration of the sound signal reinforces the frequency features with sound peaks or sustainable components, which is useful for synchronization of the camera seen in the video and audio signal [4]. In Band-pass filters play an integral role in extracting audio signals within a precise frequency range if the target band is clear, while discarding irrelevant background noise falling outside the targeted band. These versatile filters see widespread use in signal processing applications to refine data quality, as evidenced in mechanical systems designed for vibration examination and ecological surveillance [5]. The renowned Gaussian smoothing filter ensures the removal of high-frequency interference while safeguarding key characteristics of the acoustic signals. Recursively implemented Gaussian feature enhancement methods, as pioneered by Young and Van Vliet, offer computational efficiency and accuracy, allowing for rapid filtering without compromising functionality [6]. The objective of auto-correlation analysis is described as a mathematical way of pattern recognition. For instance, it helps to find a periodic signal possibly hidden in noise or a fundamental frequency which is contained in its harmonics [7] [8].

Testing the proposed approach on video recordings obtained using RGB cameras placed on a bus used by an expert driver. We demonstrated the effectiveness of our proposed method using sample videos attached with a moving vehicle in real traffic environments. With results showing how precise our timing synchronization capability is, this has the potential for much better video based environmental monitoring and analysis of that data. Accurate monitoring of the environment and synchronization of multiple streams of video has become a significant source of demand in some fields such as autonomous driving, surveillance, and monitoring of traffic situations.

2. Methodology

2.1. Processing sequence.

We used an RGB camera with a sound recorder inside. There is sound in the input video, with a high reliability and rich availability. Thus, audio signals can be used for the supportive information to analyze. We placed the cameras on the bus as shown in Fig. 1. Then, a lot of noise were contaminated in the raw audio signals. Cameras are manually started individually, therefore individual onset timings of sound signals started differently, which are not synchronized automatically. The goal of the present study is to process raw audio signals and use them to identify and synchronize the sound that is heard when a vehicle starts moving, recorded in video.

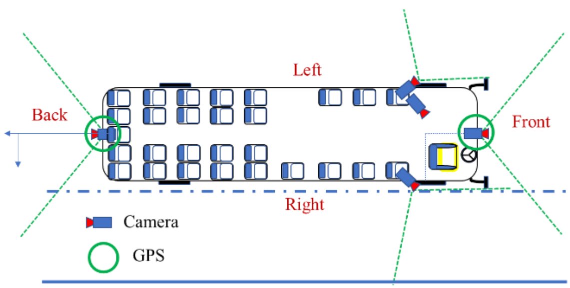


Fig. 1 Processing sequence.

First, we separated the video and audio contained in the data. Second, the relevant frequency components are identified, utilizing the magnitude scalogram obtained via the continuous wavelet transform (CWT). Then, the identified frequency band is retained and the rest is filtered out. Then Gaussian filtering was applied to smooth the signal and to reduce high-frequency noise even further. Finally, autocorrelation can detect and synchronize similar patterns in the signal. An outline of the processing sequence and the theories where applied is shown Fig. 2.

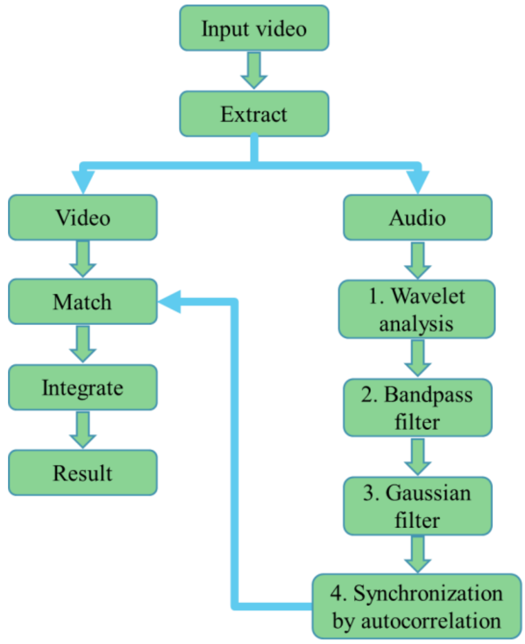


Fig. 2 Processing sequence.

There is a 12-minute (758 s) audio signal extracted from the video as in shown Fig. 3.

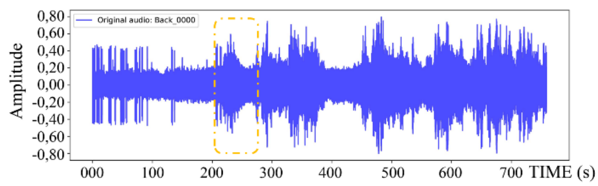


Fig. 3 Input audio signal

The bus starts moving about 3 minutes. At this point, a warning beep sounds, making it easier to synchronize the images of cameras that started working at different times. The sound of a bus starting to move, as shown in Fig. 4. However, no obvious pattern is observed in this raw signal.

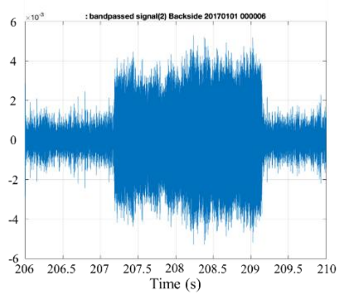


Fig. 4 The section containing the beep signal is cut off.

2.2. Continuous Wavelet Transform Method

Wavelet-based approaches are computationally inexpensive and conserves important aspects of the signal. Moreover, only complex wavelet can provide real and imaginary parts to obtain amplitude and phase information from time series. The Morse wavelet is expressed as follows in Eq. 1.

$$\psi_{P,\gamma}(\omega) = U(\omega)a_{\beta,\gamma}\omega^{\beta} \cdot e^{-\omega^{\gamma}} \quad (1)$$

Continuous Wavelet Transform (CWT) breaks down the signal into time-frequency components, which allows for identifying distinct feature differences of audio streams equation as Eq. 2. Morse wavelet has a good balance between the localization of time and frequency.

$$Wf(u,s) = \frac{1}{2}Wf_a \cdot (u,s) \quad (2)$$

2.3. Application of Bandpass Filters in Audio Signal Processing.

The audio signal has a complex structure, and at this stage, the amount of noise and unnecessary signals is reduced, preventing the ingress of low-frequency noise and unwanted high-frequency noise, allowing the next stage to work with an attractive sound quality. The band pass filter is defined as follows and is shown in Eq. 3 [4] [5].

$$h(t) = 2f_h sin(2f_h t) - 2f_l sin(2f_l t) \quad (3)$$

The real signal can be processed as shown in Eq. 4.

$$y(t) = x(t) \cdot h(t) \quad (4)$$

The expanded formula is shown in Eq. 5.

$$y(t) = x(t) \cdot 2f_h sin(2f_h t) - 2f_l sin(2f_l t) \quad (5)$$

2.4. Application of Gaussian Filters in Audio Signal Processing.

The bandpass-filtered signal may still contain residual noise, particularly at higher frequencies. Gaussian filtering smoothens the signal while retaining its essential features. The Gaussian filter kernel is defined as in shown Eq. 6 [5].

$$G(t) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(t-t_i)^2}{2\sigma^2}\right) \quad (6)$$

where $\boldsymbol{t}$ is the time from the origin in the horizontal axis, and σ is the standard deviation of the Gaussian distribution. The filtered signal is obtained by convolving the original signal with the Gaussian kernel defined as in shown Eq.7.

$$Z_{filt}(t_i) = \sum_{j-1}^{N} w_j\, z_j \exp\left(-\frac{(t-t_i)^2}{2\sigma^2}\right) \quad (7)$$

Gaussian smoothing effectively reduces noise while avoiding the sharp attenuation introduced by other types of filters.

2.5. Signal Synchronization Using Autocorrelation

Auto-correlation, otherwise called serial correlation in the discrete time setting, defines the connection between a signal and its lagged version in terms of time. Simply put, it measures the degree of association between two points of a time series on the basis of their distance in time. The objective of auto-correlation analysis is described as a mathematical way of pattern recognition. For instance, it helps to find a periodic signal possibly hidden in noise or a fundamental frequency which is contained in its harmonics. This method finds application in the processing of signals and time-domain signals among other applications. The autocorrelation function as in shown Eq. 8 [6]. $\boldsymbol{X(t)}$ is the value (or realization) produced by a given process at time t. Suppose that the process has mean $\boldsymbol{\mu_t}$ and variance $\boldsymbol{\sigma_t^2}$ at time $\boldsymbol{t}$, for each $\boldsymbol{t}$. Then the definition of the autocorrelation function between times $\boldsymbol{t_1}$ and $\boldsymbol{t_2}$

$$R_{XX}(t_1,t_2) = E\left[X_{t_1}\overline{X_{t_2}}\right] \quad (8)$$

where $\boldsymbol{E}$ is the expected value operator and the bar represents complex conjugation. Subtracting the mean before multiplication yields the auto-covariance function between times $\boldsymbol{t_1}$ and $\boldsymbol{t_2}$ as shown in Eq. 9, Eq. 10.

$$R_{XX}(t_1,t_2) = E\left[\left(X_{t_1} - \mu_{t_1}\right)\overline{\left(X_{t_2} - \mu_{t_2}\right)}\right] \quad (9)$$

$$R_{XX}(t_1,t_2) = E\left[X_{t_1}\overline{X_{t_2}}\right] - \mu_{t_1}\overline{\mu_{t_2}} \quad (10)$$

3. Results and Discussion

3.1. Wavelet Analysis

In CWT analysis provides a detailed representation of the time-frequency distribution of an audio signal. The wavelet transform is illustrated in Fig. 5 as a scalogram. The magnitude scalogram represents the energy distributed over time and frequency, clearly in a semblance. Looking at the results of the first wavelet processing, our target sound does not look very good.

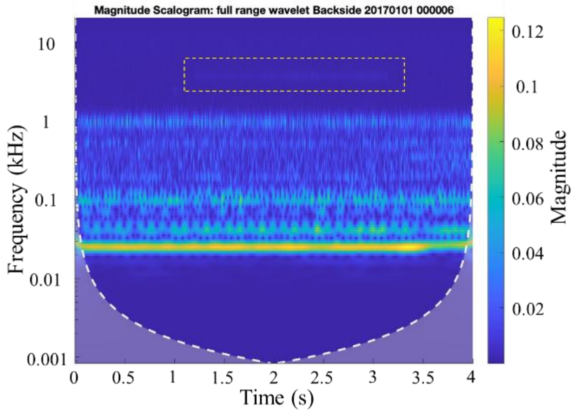


Fig. 5 Results using Wavelet analysis

3.2. Bandpass Filter

The band-pass filter effectively removes the low and high frequency components that interfere with further processing of the signal. This filter has been used in many experimental studies and has already been shown to give good results.

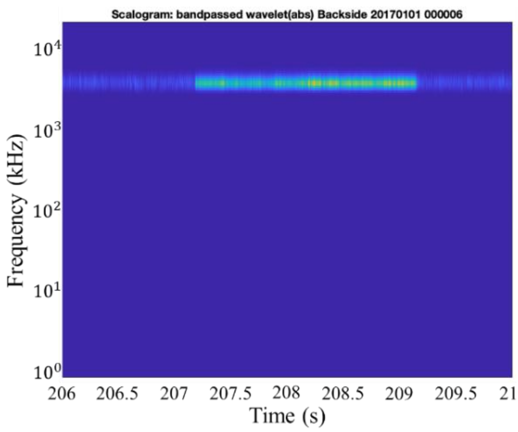


Fig. 6 Results using Bandpass filter

In the signal frequency components from 3.7k Hz to 4k Hz, those were judged to be relevant for the study and were retained by bandpass filtering as in shown Fig. 6. Unwanted low-frequency and high-frequency noise was effectively eliminated. The filtered signal retained most of the relevant components of the sound signal. But even now, this signal is still affected by noise.

3.3. Gaussian filter

The Gaussian filter effectively filters out noise without losing the characteristics of the signal, also it can also correct a signal that has been altered to some extent by the effects of noise. This employed a Gaussian filter for smoothing out any residual noise present in the bandpass-filtered signal. The standard deviation was adjusted thereby attaining smoothness with minimal distortion, shown in Fig. 7.

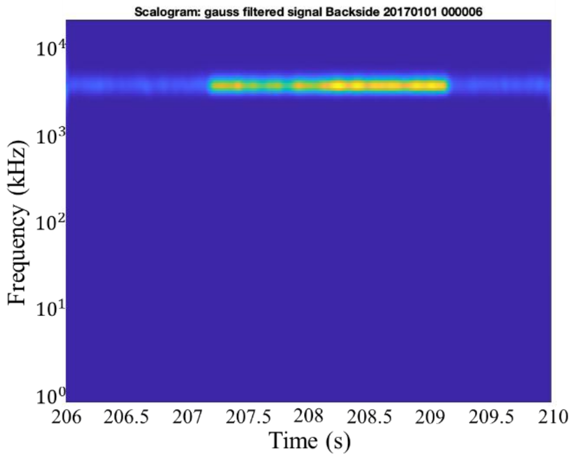


Fig. 7 Results using Gaussian filter

We have performed a wavelet transformation of the filtered signal. The signal was then reconstructed using the inverse wavelet.

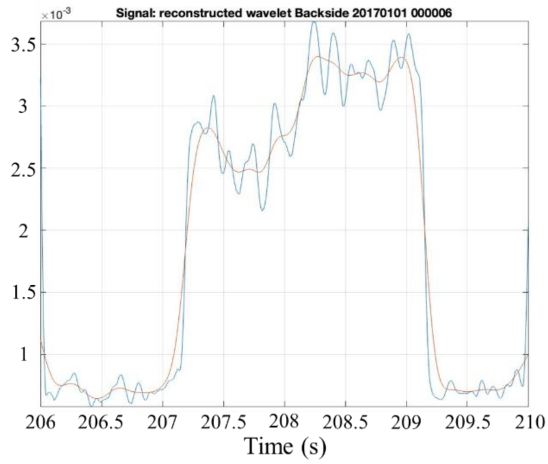


Fig. 8 Results using Reconstruct and Gaussian filter

This reconstructed signal was processed by Gaussian filtering, as shown in Fig. 8. Reconstruction and Gaussian filtering make the signal clearer and improve its characteristics.

3.4. Signal Synchronization Using Autocorrelation

We have to synchronize the signals. To do this, we use autocorrelation. Autocorrelation is useful to detect signal coherence. We used the autocorrelation function on this filtered data. Then we can determine the time of such synchronization. Finally, we have achieved our goal and the time synchronization as shown in Fig. 9.

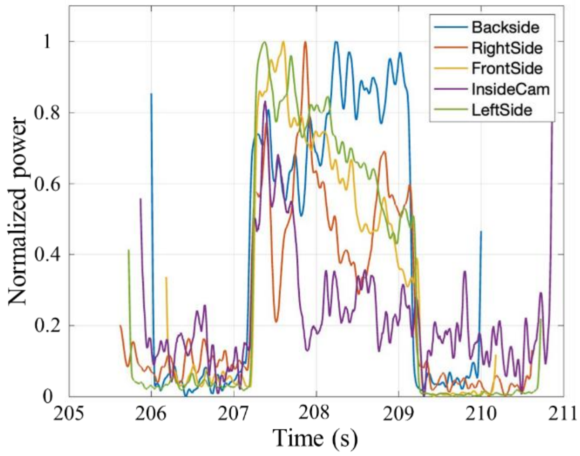


Fig. 9 The synchronized results

We had five cameras inside the bus. All of these cameras were started at different times. But each of these cameras recorded the sound inside the bus. So I was able to recognize the sound signal that the bus made when it started moving and use it for synchronization.

4. Conclusion

In this paper, a method for synchronizing multi-camera videos based on audio signals is proposed. This method uses wavelet transform, bandpass filtering, Gaussian noise reduction and autocorrelation function to achieve synchronization. Wavelet analysis enables frequency and amplitude analysis without losing the exact time scale. The important information of the frequency components is recorded in the magnitude scalogram. Bandpass filtering removes undesirable section of the signal, while Gaussian smoothing reduces noise. Wavelet transforming the Gaussian-filtered signal and reconstructing it further reveals the characteristics and shape of the signal. Finally, the use of autocorrelation enables very accurate synchronization.

In future work will be focused on real-time implementation as well as combining this method with previous multi-camera systems to improve operation in dynamic environments.

Acknowledgements

This work was supported in part by JSPS KAKENHI (JP17H06383, JP24K07387) and Kitakyushu Foundation for the Advancement of Industry, Science and Technology (FAIS). Authors also deeply appreciate expert drivers and safety operation managers in Nishitetsu group for offering a precious opportunity to examine professional procedures for the safety assessments in the real driving on the road.

References

1. Wu, Xinyi, et al. "Multi-video temporal synchronization by matching pose features of shared moving subjects." Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019, pp. 0-0.
2. Brassarote, Gabriela de Oliveira Nascimento, E. M. Souza, and João Francisco Galera Monico. "Non-decimated wavelet transform for a shift-invariant analysis." TEMA (São Carlos) 19.1, 2018, pp. 93-110.
3. Singh, Balbir, and Hiroaki Wagatsuma. "Two-stage wavelet shrinkage and EEG-EOG signal contamination model to realize quantitative validations for the artifact removal from multiresource biosignals." Biomedical Signal Processing and Control 47, 2019, pp. 96-114.
4. Barik, B., Kalirasu, A., & Prathap Kumar, A. V. High efficient band pass filter design and analysis on impact of resonator on its performance. Journal of Shanghai Jiaotong University, 17(3), 2021, pp.180–190.
5. Shahruz, S. M. Design of mechanical band-pass filters for energy scavenging. Journal of Sound and Vibration, 292(4–5), 2006, pp. 987–998.
6. Young, I. T., & Van Vliet, L. J. Recursive implementation of the Gaussian filter. Signal Processing, 44(2), 1995, pp. 139–151.
7. Reynolds, K. M., and L. V. Madden. "Analysis of epidemics using spatio-temporal autocorrelation." Phytopathology 78.2 1988, pp. 240-246.
8. Gubner, John A. Probability and Random Processes for Electrical and Computer Engineers. Cambridge University Press. 2006

Authors Introduction

Mr. Davaanyam Jargal

He received his Master's degree from the School of Mechanical Engineering and Transportation, Mongolian University of Science and Technology, Mongolia in 2010. He is currently a Doctoral course student in Kyushu Institute of Technology, Japan

Dr. Obada Al Aama

He graduated from the Department of Communication and Electronics Engineering at Al-Baath University, Syria, in 2013. He obtained his M.Eng. and Ph.D. degrees from the Kyushu Institute of Technology, Japan, in 2019 and 2024, respectively. He is currently a researcher in the Department of Human Intelligence Systems at the Kyushu Institute of Technology, Japan. His research interests focus on robotics and autonomous driving.

Takahiro Koga

He received his Master's degree in Engineering from the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology in Japan. He is currently a researcher in Kyushu Institute of Technology.

Tomoki Taniguchi

He received his Bachelor's degree in Computer Science and Systems Engineering from the Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, Japan, in 2024. He is currently a master's student at Kyushu Institute of Technology, Japan.

Ms. Rena Kato

She received her associate's degree from the Department of Electrical and Electronic Systems Engineering, National Institute of Technology, Toyota College, Japan in 2024. She is currently enrolled in the Department of Systems Design and Informatics, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, Japan.

Hakaru Tamukoh

He received a B.Eng. degree from Miyazaki University, Japan 2001. He received his M.Eng and Ph.D. from Kyushu Institute of Technology, Japan, in 2003 and 2006, respectively. He was a postdoctoral research fellow of the 21st-century Center of Excellent Program at Kyushu Institute of Technology, from April 2006 to September 2007. He was an assistant professor at Tokyo University of Agriculture and Technology, from October 2007 to January 2013. He was an associate professor from February 2013 to March 2021 and is currently a professor at the Graduate School of Life Science and System Engineering, Kyushu Institute of Technology, Japan. His research interests include hardware/software complex systems, digital hardware design, neural networks, soft computing, and computing service robots. He is a member of IEICE, SOFT, JNNS, IEEE, JSAI, and RSJ.

Dr. Hiroaki Wagatsuma

He received his M.S., and Ph.D. degrees from Tokyo Denki University, Japan, in 1997 and 2005, respectively. In 2009, he joined Kyushu Institute of Technology, where he is currently an Associate Professor of the Department of Human Intelli- gence Systems. His research interests include non- linear dynamics and robotics. He is a member of IEEE.

Intelligent Path Planning for Robots and Practical Implementation of Programmable Headlights for Autonomous Vehicles

Farkad Adnan
College of Computing & Informatics (CCI), Universiti Tenaga Nasional (Uniten), Kuala Lumpur, Malaysia
Multaqaa Al-Nahrein University College ,Computer Technology Engineering Department, Basra,Iraq

Abdul Samad Bin Shibghatullah
College of Computing & Informatics (CCI), Universiti Tenaga Nasional (Uniten), Kuala Lumpur, Malaysia

Mohd Radzi Bin Aridi
College of Engineering (COE), Universiti Tenaga Nasional (Uniten), Kuala Lumpur, Malaysia

E-mail: farkad.hpfa95@gmail.com (PT21515@student.uniten.edu.my), abdul.samad@uniten.edu.my, Radzi@uniten.edu.my

Abstract

The ability to move is essential for the development of intelligent robots for autonomous navigation. Neural networks outperform traditional methods in modeling complex relationships and identifying patterns, but current systems are limited to specific robots and sensors. This paper presents a universal method for interpreting data from different 2D sensors, predicting distances between robots and walls, and using neural networks for navigation. The goal is to create a versatile algorithm that can be applied to different robots and programmable lamps, reducing accidents. The thesis also explores programmable lamps that block light from reaching the eyes of passengers, using one network to determine free space using odometry data and another to find safe paths while avoiding obstacles. Simulated path examples will be presented.

Keywords: Arduino, Artificial Intelligence (AI), Finds, Findpath, VFH, RBF .

1. Introduction

A robotic system that is mobile is referred to as a robot. In addition, he qualifies as fully independent if he is built to complete any task without the assistance of a person. In the interim, every operation carried out with outside or public aid Reduce the robot's autonomy with an external tool. These interventions can be carried out remotely, provide access to the entire environment, or only a portion of it that isn't on board the robot [1].

In this study, a robotic path finding projector and the algorithms used on lamps will be built and applied as a new innovation by directing the light beam towards the road. These developments began with gas and oil lamps in previous centuries, and they have a lifespan, but the problem is that they were bright in a way that affected other drivers. These modern sources provide bright and comfortable color temperatures, which enhances the driving experience. However, despite the advances in lighting technology, the only option currently available to most drivers is to switch between high and low beams [2], High beams also greatly reduce contrast in fog and mist, and cause distracting bright streaks in rain. Even after 130 years of headlight development, more than half of all vehicle and traffic accidents occur at the same time.

2. Methodology

This paper presents a control strategy based on the idea of robotic trajectory tracking and the calculated torque method. The simulation module designed in this project consists of two separate three-layer neural networks, one of which is used to calculate the mass matrix of the manipulator estimated by the used algorithm, and the other network is also trained to calculate the estimated centrifugal torque vector. These results will be used in the process of determining the light path of programmable lamps and determining the light direction. The simulation results also demonstrate the ability of the proposed control simulation to learn the nonlinear dynamics of the manipulator using the RBFN algorithm and Vector Field Histogram and was able to force the final effect to follow the desired position path in the XY plane. However, the typical learning speed of the controller can be considered fast Finite-State Machine [3].

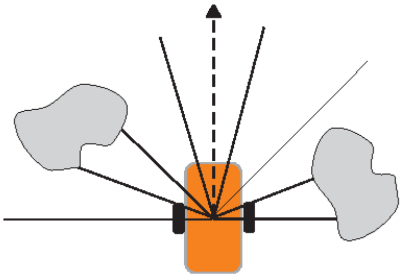


Fig.1. Polar graph used in VFH algorithm

In (Fig.1) a real-time obstacle avoidance method for mobile robots was described. This method, called vector field graph (VFH), allows detecting unknown obstacles and avoiding collisions while simultaneously guiding the mobile robot towards the target. In this stage, we will take advantage of this previously built technique and apply it to programmable lights by avoiding collisions with "air dust" or falling snow through the existing camera system which first detects all particles by illuminating them in a very short period of time, their future positions are predicted, and then the rays intersecting them are interactively turned off. The time between image capture and interactive illumination is the system latency.

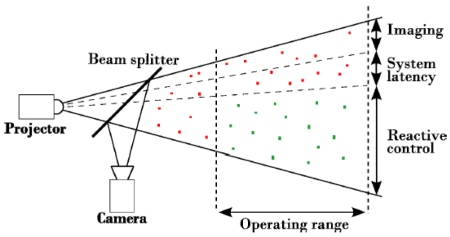


Fig.2. Using path finding technology in the lamp programming and light path determination process

We demonstrate using quantum simulations that the light production rate of an ideal system is not significantly affected by a wide range of precipitation types and rates. Based on the lessons learned from the simulations, Fig. 2 shows a prototype of the system and describe the trade-offs between the algorithm used, hardware speed, achievable light production rate, and accuracy needed to make the system viable as automotive head lights [4].

Figure (Fig.3) shows how the system works and how the light is blocked from other cars.

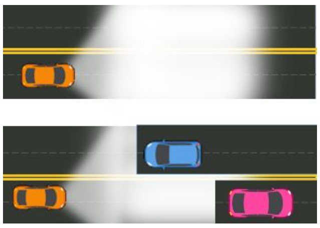


Fig.3. Using pathfinding technology

3. The Prototype

The glare from headlights, particularly high beams, of oncoming vehicles is at best a major distraction and at worst a temporary blindness. Trucks and other vehicles with high-mounted headlights are among the worst offenders. Although glare is not often considered a major cause of accidents, hundreds of fatal nighttime crashes are attributed to it as a contributing factor each year, as shown in (Fig.4). Glare is a particular problem for older adults, who often take longer to recover from the glare than a 16-year-old while high beams are a nuisance to other drivers, they are helpful on narrow, winding, poorly lit roads, especially in rural areas where wildlife routinely crosses the road [5] .

Fig. 4. Prototype and Control Process of Programmable Headlights

The programmable headlight can be controlled with such high spatiotemporal resolution that light rays directed towards any driver in any number of lanes can be disabled to prevent glare. The precision and speed of our beam control is several orders of magnitude better than existing LED based headlights.

Vehicles are recognized and the lighting is removed from the limited areas around drivers or the eyes of drivers of vehicles in front of the vehicle while maintaining lighting in other areas. This will allow drivers to use the high beams without worry. The center row shows the vision while driving towards the prototype. Reduced glare when the anti-glare feature is activated in our headlights. Anti-glare headlights allow the driver to see other vehicles on the road better. Glare in the rearview mirror from a vehicle coming from behind will not affect the driver while driving, and the taillights are detected to avoid illuminating the rearview mirror.

4. Neural Research methodology

Neural networks have been at the forefront of growth in recent years due to their versatility, opportunities and most importantly their dynamic nature. A robot path simulation system for path planning with the help of neural networks has been proposed as a comparative study. Various parameters such as training time, network performance and expected distance are taken into account after iteration to obtain the optimal dataset using probabilistic roadmap algorithm [6] .

The parameters from the simulation system are taken and the performance of the robot path determination process is applied in the process of determining the path of light emitted from standard car headlights which improves the driver's vision at night by illuminating the road and the surrounding environment and disabling the light beams directed towards any driver on the road (other cars coming the opposite way) allowing the high beams to pass without any noticeable difference to the drivers on the road.

5. System Architecture

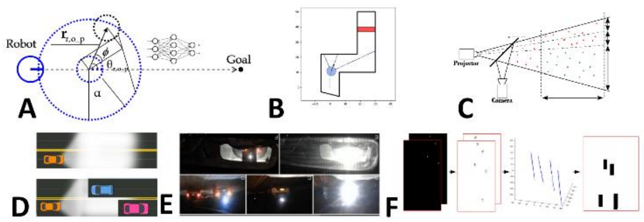


Fig.5. At night, fog and rain appear as a shimmering (distracting) pattern that reduces the driver's visibility while driving (top). We propose an interactive system (bottom) that deactivates light rays that only intersect with particles such as raindrops, snow and hail, reducing the visual impact of precipitation.

A: In this step, the system simulator is programmed and prepared, where the path detection algorithms are trained and the robot path is determined, which in turn will be used in the process of determining the light path using programmable lamps. Here, we will create a system using the Python language, through which the algorithms used in the hybridization process will be determined, which we will use the parameters in step B.

B: After taking the parameters from the previous step, we train the neural network here, which in turn will determine the path of the robot and its launch from the starting point to the target point according to the path that will be determined by the parameters and by the neural network that was previously trained. After that, the process of taking the training results resulting from this step and applying them to the system will determine the light path in the next step.

C: The training results from the previous step will be fed into the controller used in the current system, which will determine the path of the headlights of standard cars, which in turn can improve the driver's vision at night by illuminating the road and the surrounding environment. It will block the path of light directed at other drivers, which reduces driver fatigue and makes roads safer. This is done according to a set of sensors and tools connected to each other and using more than one smart controller with artificial intelligence techniques, which will give tangible results, making the device a real innovation and this approach is applicable and effective.

E: In (Fig.5)(E), Imagine driving at night in heavy rain. Sparkling raindrops, drizzle, and foggy windshields all contribute to poor visibility, making the driving experience uncomfortable at best and dangerous at worst. We present a sophisticated imaging and illumination system that reduces light reflections to the driver, as well as the scattered light that affects nearby and oncoming drivers [7].

F: In (Fig.5)(F),We do this work in building an integrated system through which the process of building a more advanced system is done, where in this step we create a system that determines the path of light using path determination techniques so that when it rains or fog forms, this technology will penetrate the path of fog and make the vision clearer while driving as in (Figure 6), which enhances the current innovation as a new technology in applying path planning (Fig.6) and applying this technology to programmable lights[8].

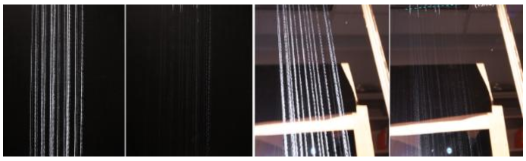


Fig.6. The diamonds falling from a place as a substitute for rain were illuminated by a halogen light and photographed for comparison at different distances. The graph shows that the water droplets are invisible.

6. Experiment Results

Computer simulation results for anti-glare headlamps. Detection and prediction are assumed to be perfect, and vehicles are moving toward each other in two-lane roads at a relative speed of 170 km/h.

. Left: Light transmission rate as a function of distance between vehicles during different time periods of the system as in (Fig.7). Right: Light transmission rate for DMD-based and LED-based anti-glare headlamps at different time periods. Simulations show that shorter transition times and higher resolution results in higher light transmission rates and better resolution, which will be especially important on winding, multi-lane roads with multiple vehicles [9].

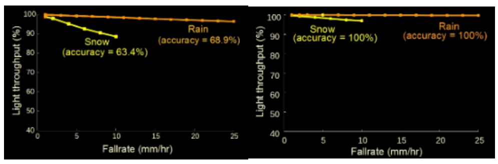


Fig.7. Results of computer simulations The figure shows the system features compared to the time of light particle transmission, updates and initialization frames.

7. Conclusion

Automotive headlights should not be devices that can be either fully on or off. They should adapt to the surrounding conditions for safety during times of good visibility. Adjustable headlights should not be limited to a single task. They should perform several different functions to support the driver in diverse road conditions. The prototype here gives our headlights unprecedented control over the direction and timing of the light. It has been demonstrated that headlights are flexible and suitable for many tasks: allowing high-beam driving without disturbing other drivers. We need to develop a more advanced algorithmic and software hybridization process before bringing our headlight design to market.

References

1. Singhal, K., Kumar, V., & Rana, K. (2022). Robust Trajectory Tracking Control of Non-Holonomic Wheeled Mobile Robots using an Adaptive Fractional Order Parallel Fuzzy PID Controller. Journal of the Franklin Institute, 359. doi:10.1016/j.jfranklin.2022.03.043
2. Londono, L., Hurtado, J., Hertz, N., Kellmeyer, P., Voeneky, S., & Valada, A. (2024). Fairness and Bias in Robot Learning. Proceedings of the IEEE, PP, 1–26. doi:10.1109/JPROC.2024.3403898
3. Adnan, F. (2018). Arduino World Book-FARKAD ADNAN.
4. Das, P., Sadhu, A. K., Vyas, R., Konar, A., & Bhattacharya, D. (2015). Arduino Based Multi-robot Stick Carrying by Artificial Bee Colony Optimization Algorithm. doi:10.1109/C3IT.2015.7060152
5. Zheng, F., Whitted, T., Lastra, A., Lincoln, P., State, Andrei, Maimone, A., & Fuchs, H. (2014). Minimizing latency for augmented reality displays: Frames considered harmful. 195–200. doi:10.1109/ISMAR.2014.6948427
6. O'Toole, M., & Kutulakos, K. (2010). Optical Computing for Fast Light Transport Analysis. ACM Trans. Graph., 29, 164. doi:10.1145/1866158.1866165
7. Advanced Driving Assistance and Active Safety Systems http://media.opel.com/media/intl/en/opel/vehicles/opeleye/2009.html
8. https://www.istockphoto.com/videos/night-driving-headlights-glare
9. Mercedes Benz New generation CLS Features Multibeam LED Technology (https://www.ledinside.com/news/2014/6/mercedes_benz_new_generation_cls_features_multibeam_led_technology

Authors Introduction

Farkad Adnan

From Iraq, Basra, a university professor at the University of the Multaqa Al Nahrain , he is a computer engineer and a current PhD student, as well as an experienced professional in web and mobile application development, specializing in robotics, artificial intelligence, the Internet of Things, and emerging innovations. With a degree in robotics from Stanford University and patents in artificial intelligence and microchip innovation, he is a recognized expert in this field. The author of 8 scientific books, Farkad has showcased his expertise globally through seminars, conferences, and platforms such as TEDx, Stars of Science, and EXPO. He was selected as one of the Arab Youth Pioneers in the Emirates, and was also selected in Forbes Middle East magazine. In addition to his skills, he is also talented in drawing, mechanical engineering, and architecture, reflecting his diverse abilities and commitment to spreading knowledge and innovation. linktr.ee/farkadadnan

Abdul Samad Shibghatullah

He received his Bachelor's degree in Accounting from Universiti Kebangsaan Malaysia, Bangi, Malaysia, in 1999, the M.Sc. degree in Computer Science from Universiti Teknologi Malaysia, Skudai, Malaysia, in 2002, and the Ph.D. degree in Computer Science from Brunel University, Uxbridge, United Kingdom. He is currently an Associate Professor at the College of Computing & Informatics (CCI), Universiti Tenaga Nasional (Uniten), Kajang, Selangor. His current research interests include optimization, modeling, and scheduling.

Mohd Radzi Bin Aridi

Dr. Mohd Radzi Bin Aridi is a Senior Lecturer at College of Engineering, Universiti Tenaga Nasional (UNITEN). He holds a Ph.D. in Engineering from Kyushu Institute of Technology, Japan (2021), an M.Sc. (2014), and a B.Sc. (2011) from Universiti Putra Malaysia. His research focuses on mechanical engineering and applied mechanics, particularly stress intensity factors and crack interactions in plane elasticity. He has been working on the fatigue risk evaluation of the rolling roll by using the finite element analysis. He has been working on the fatigue risk evaluation of the rolling roll by using the finite element analysis.

Recognition of Plastic Bottles Region Using Improved DeepLab v3+

Yusuke Murata
Department of Mechanical and Control Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata-ku, Kitakyushu, Fukuoka, 804-0015, Japan

Tohru Kamiya
Department of Mechanical and Control Engineering, Kyushu Institute of Technology,
1-1 Sensui-cho, Tobata-ku, Kitakyushu, Fukuoka, 804-0015, Japan
E-mail: murata.yusuke973@mail.kyutech.jp, kamiya@cntl.kyutech.jp

Abstract

Factory automation is one solution to the labor shortage. We focus on the sorting of plastic bottles in waste disposal plants and try to automate the process using robotic arms. In this paper, we propose an image analysis method for the recognition of plastic bottles limited to 500ml capacity. The method is semantic segmentation, and the deep learning model is DeepLab v3+. Modifications using ECA Block and Mish function show improvements at the points of misrecognition with the base model.

Keywords: Deep Learning, Semantic Segmentation, Convolutional Neural Network (CNN), DeepLab v3+, Efficient Channel Attention Block (ECA Block), Mish function

1. Introduction

The labor shortage in Japan is becoming more serious every year [1]. This study focuses on labor shortage in factories. The causes include the decline in the labor force due to the low birth rate and aging population, and the negative image of repetitive and simple work. As solutions to these problems, the employment of foreign and female workers, the training of workers who can perform multiple tasks, and the enhancement of the factory's image can be considered. However, this study focuses on factory automation to take advantage of technology.

Factory automation refers to the use of machines to replace tasks traditionally performed by humans. Many factories have already automated some of their equipment, and an increasing number of factories are investing in equipment with more advanced technologies. Factory automation is expected to not only solve the labor shortage, but also reduce the labor cost, stabilize the product quality, increase the operation rate, and improve the work environment.

This study focuses on the sorting of plastic bottles at waste treatment facilities. The processes required for recycling vary from material to material, and each resource must be sorted. Currently, cans are sorted using magnets, but plastic bottles are sorted manually. To automate this process, we are trying a method to recognize plastic bottles from images and pick them up with a robotic arm. In this study, we propose an image analysis method specifically for 500ml plastic bottles. As basic research, we conducted experiments to investigate how accurately individual plastic bottles can be recognized in images. Semantic segmentation was used as the method, and DeepLab v3+ [2], which is a powerful deep learning model in this field, was employed. As the backbone, we used ResNet-101 [3], and evaluated the performance of each model with improvements by combining the Efficient Channel Attention Block (ECA Block) [4] and the Mish function [5].

2. Method

In this section, we describe an image analysis method for the recognition of plastic bottles. The overall flow is to collect input images and manually annotate them. Then, training and inference are performed on this dataset using a deep learning model. In this way, plastic bottles separated into their bodies and caps can be recognized from unknown images. Semantic segmentation is used as the recognition method, and the details of the deep learning model with this method are described below.

2.1. *DeepLab v3+*

We used DeepLab v3+ [2] as our deep learning model. DeepLab v3+ is one of the best performing Convolutional Neural Networks (CNNs) for semantic segmentation, which was proposed in 2018. There are two main CNN structures used in this method. The first is a network that uses Spatial Pyramid Pooling (SPP) [6], and the second is a network that consists of encoders and decoders. DeepLab v3+ combines these two structures to obtain a wide range of information and sharp object contours, and achieves high performance.

In this study, the body and cap of a plastic bottle are recognized by semantic segmentation. In particular, the cap is small compared to the body and needs to be recognized accurately. DeepLab v3+ enables highly accurate recognition of small objects due to the advantages described above, and this is the reason why DeepLab v3+ was chosen.

2.2. *ResNet-101*

We adopted ResNet-101 [3] as the backbone of DeepLab v3+. ResNet is a CNN that enables the construction of deep layers by introducing Residual Blocks, which was proposed in 2015. The performance of CNNs improves with the depth of the layers because they extract higher-dimensional global features. However, simply increasing the depth of the layers leads to poor performance due to the degradation problem, etc. To solve this problem, the network that introduces the skip connection is ResNet.

ResNet is divided into five types according to the number of residual blocks. In this study, we chose ResNet-101 to avoid overlearning due to the complexity of the network while incorporating global information.

2.3. *ECA Block*

ECA Block [4] is a channel-directed attention mechanism proposed in 2019. Channel-directed attention mechanisms emphasize important information and have been shown to improve the performance of deep CNNs. The advantage of the ECA Block is that it effectively captures inter-channel dependencies while keeping the complexity of the mechanism low. This is achieved by incorporating a single 1D convolutional layer instead of multiple fully connected layers. Fig. 1. shows the structure of ECA Block.

In this study, the goal is to automate simple tasks in factories, and it is necessary to improve the recognition accuracy while keeping the processing speed low. ECA Block is suitable to achieve this. Because it emphasizes important information without increasing the complexity of the model. We believe that this characteristic improves the overall recognition rate without increasing the computation time.

2.4. *Mish function*

Mish function [5] is an activation function proposed in 2019, and it is represented by Eq. (1).

$$f(x) = x\tanh(\ln(1 + e^x)) \qquad (1)$$

The Mish function has the property that its output changes smoothly and continuously with respect to the input and is differentiable over the entire range. This property improves gradient stability and mitigates problems such as gradient loss and gradient explosion during the learning process. In addition, The Mish function can smoothly use information for negative-valued inputs, thus capturing a wider variety of features than activation functions that completely truncate negative values to zero. Because of these properties, The Mish function is effective for detecting fine patterns in images, and its performance improvement has been reported in various fields.

In this study, plastic bottles are recognized by dividing them into the body and the cap. Since the cap is smaller than the body, it is necessary for the model to accurately recognize small features. Therefore, we believe that the introduction of the MIsh function is effective.

2.5. *Proposed network*

In this study, we used DeepLab v3+ as the base model and ResNet-101 as the backbone. The activation function was replaced with the MIsh function, and the ECA Block was added at appropriate locations. The overall structure of the proposed network is shown in Fig. 2.

We placed the ECA Block in two locations. The first is before the addition of skip connection in the Residual Block. ECA Block is a module that is added after the convolutional layer, and by placing it after the last convolutional layer in Residual Block, it is possible to select important information from features extracted up to the previous layers. Therefore, we thought it would be possible to clarify important information throughout ResNet. The second is after the encoder and decoder are combined. The part combines the high-dimensional features from the encoder output and the low-dimensional features from the skip connection. By placing the ECA Block after this, we thought it would be possible to focus on the important channels from the combined features, resulting to improved performance.

3. Experiment

The dataset used in this study was created independently and consists of 200 images of plastic bottles. The annotation was done manually. In addition, we performed five-fold cross-validation as a validation method and employed Intersection over Union (IoU) and mean Intersection over Union (mIoU) as an evaluation index.

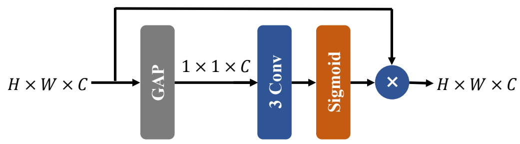


Fig. 1. ECA Block structure

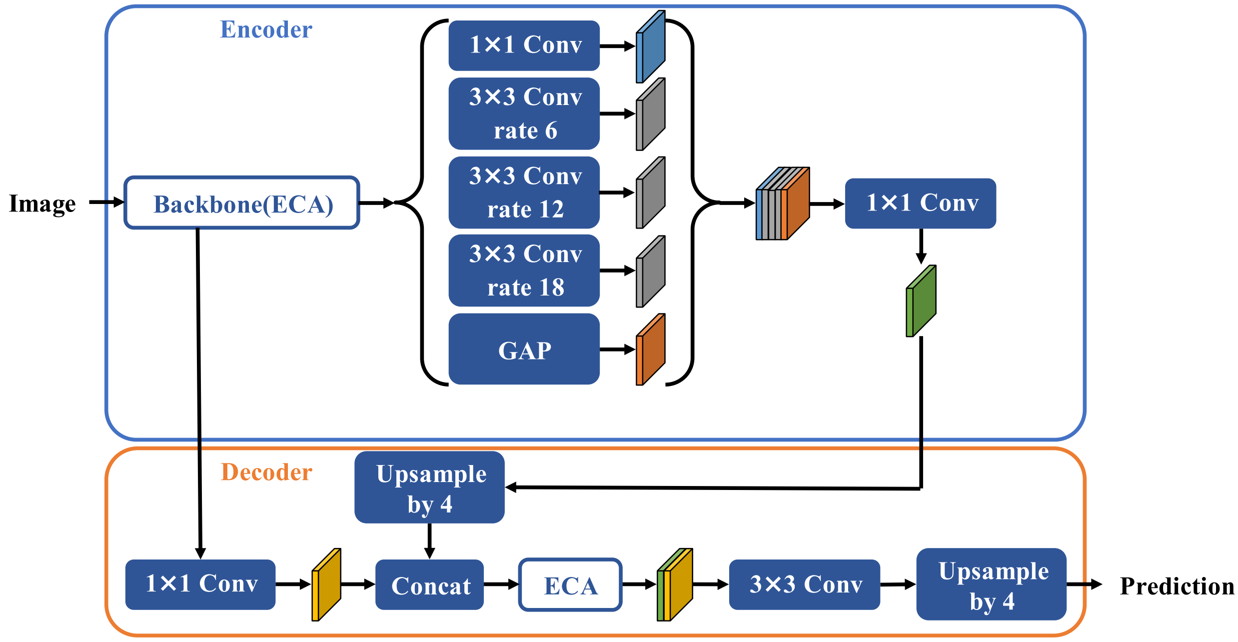


Fig. 2. Proposed network structure

3.1. *Result*

The experimental results are shown in Table 1. In this study, we calculated the IoU of the plastic bottle body, IoU of the cap part, mIoU, and fps for the five models. Note that the IoU of the body is denoted by "body" and the IoU of the cap part is denoted by "cap". In addition, the input image and the result image are shown in Fig. 3. Red is the mask of the body and green is the mask of the cap part. As shown in the results, the proposed method did not show any improvement over the base model in terms of numerical performance. However, there were changes in the trend of the captured features, the causes of which will be discussed later.

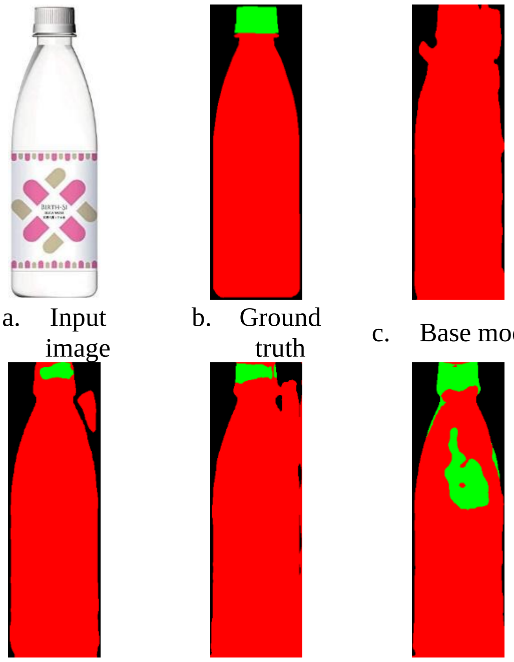
3.2. *Discussion*

Compare the individual images in Fig. 3. When the ECA Block was added to the base model, the number of false detections of the body and non-detections of the cap part decreased, probably because the ECA Block emphasized the important information. When the Mish function was added to the base model, the detection of the cap part increased. This could be because the Mish function improved the detection performance for small features. We thought that the combination of these characteristics would enable accurate recognition of both the body and the cap part, and the result of adding the ECA Block and Mish function is shown in Fig. 3. f. As expected, both the body and the cap part showed improvement over the base model. However, new false detection of the cap part appeared in the center of the body. A possible reason for this is that the cap part was made easier to detect by the Mish function and emphasized by the ECA Block, and the range of the features detected as the cap part became wider. In addition, even when the ECA Block and Mish function are added independently, there are many data in which their characteristics have negative effects, which is believed to lead to lower accuracy.

The most accurate of the conventional methods [7] was DeepLab v3+ with fine tuning. Compared to the methods tested in this study, the accuracy of the conventional method was significantly higher. The presence or absence of fine tuning is a major cause of this difference. Therefore, we would like to repeat the experiment by adding fine tuning to the method used in this study. In addition, although we worked on image recognition using simple images as basic research, in actual waste treatment facilities, it is necessary to detect plastic bottles from complex images. Therefore, the conventional method may not be accurate enough. We believe that the fact that we found changes in the trend of feature extraction with the addition of each module in this study is meaningful for practical use.

Table 1. Results for each model

	cap	body	mIoU	fps
Base model	0.673	0.955	0.814	14.74
With ECA	0.661	0.943	0.802	14.28
With Mish	0.638	0.950	0.794	14.70
Proposed model	0.611	0.941	0.776	14.06
Conventional model	0.932	0.980	0.956	7.42



a. Input image b. Ground truth c. Base model

d. With ECA e. With Mish f. Proposed model

Fig. 3. Segmentation result

In this study, we were able to extract features that were not captured by the base model. However, the effect of the modification was detrimental, resulting in false detections in the cap part. To solve this problem, we believe that emphasizing the positional information of the cap part is effective, so we consider introducing a spatial attention mechanism in the future.

4. Conclusion

In this paper, we conducted basic research on a plastic bottle sorting method for factory automation. Semantic segmentation was applied to images containing individual plastic bottles, and the body and the cap part were recognized separately. We used DeepLab v3+ as a deep learning model and modified it with the ECA Block and Mish function. As a result, improvements were observed in areas of misrecognition with the base model.

References

1. JMA Management Center Inc., "[2024] What are the root causes of labour shortages in the manufacturing industry? A Thorough Explanation of the Actual Situation and Countermeasures", https://www.jmam.co.jp/hrm/column/0136-manufacturing.html (in Japanese) (access 2024/11/28).
2. L. C. Chen et al., "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation", Proceedings of the European Conference on Computer Vision (ECCV), pp.801-818, 2018.
3. K. He et al., "Deep Residual Learning for Image Recognition", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.770-778, 2016.
4. Q. Wang et al., "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.11534-11542, 2020.
5. D. Misra, "Mish: A Self Regularized Non-Monotonic Activation Function", arXiv preprint arXiv:1908.08681, 2019.
6. K. He et al., "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition", in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.37, no.9, pp.1904-1916, 2015.
7. D. Ideta, T. Kamiya, "Recogniton of Specific Parts of Plastic Bottles Using Improved DeepLab v3+", 2023 23rd International Conference on Control, Automation and Systems (ICCAS), pp.1758-1761, 2023

Authors Introduction

Yusuke Murata

He received his Bachelor's degree in Engineering in 2024 from the Faculty of Engineering, Kyushu Institute of Technology in Japan. He is currently a master student in Kyushu Institute of Technology, Japan

Tohru Kamiya, Prof. Ph.D

He received his B.A. degree in Electrical Engineering from Kyushu Institute of Technology in 1994, the Master and Ph.D. degree from Kyushu Institute of Technology in 1996 and 2001, respectively. He is a professor in the Department of Mechanical and Control Engineering at Kyushu Institute of Technology. His research interests are focused on image processing, medical application of image analysis. He is currently working on computer aided diagnosis based on CT, MR imaging, fluorescence microscope imaging, and automatic classification of respiratory sound.

Non-Invasive Classification of EGFR Mutation from Thoracic CT Images Using Radiomics Features and LightGBM

Reo Takahashi
Kyushu Institute of Technology, 1-1 Sensui, Tobata-ku, Kitakyushu, 804-8550, Japan

Tohru Kamiya
Kyushu Institute of Technology, 1-1 Sensui, Tobata-ku, Kitakyushu, 804-8550, Japan

Takashi Terasawa
University of Occupational and Environmental Health, 1-1 Iseigaoka, Yahatanishi-ku, Kitakyushu, 807-8555, Japan

Takatoshi Aoki
University of Occupational and Environmental Health, 1-1 Iseigaoka, Yahatanishi-ku, Kitakyushu, 807-8555, Japan
Email: takahashi.reo828@mail.kyutech.jp

Abstract

Cancer caused 9.7 million deaths in 2022, including 1.8 million from lung cancer the leading cause of cancer death. EGFR mutation testing is essential for lung cancer treatment planning, but it is invasive and visual identification from chest CT images is difficult. This paper proposes a computer-aided diagnosis system to identify EGFR mutation status. Lung tumor regions were automatically extracted and radiomics features were obtained. Dimensionality reduction was performed using null importance, variance inflation factor, and recursive feature elimination. The method was applied to 143 cases and achieved an accuracy of 59.1%, a true positive rate of 54.3% and a false positive rate of 36.1%. The results suggest that CAD (Computer-Aided Diagnosis) systems can improve the non-invasive detection of EGFR mutations in lung cancer.

Keywords: Computer Aided Diagnosis, Radiomics, U-Net, LightGBM

1. Introduction

Cancer is the world's deadliest disease, with an estimated 20 million new cases of cancer and 9.7 million deaths by 2022. According to estimates, one in five people will develop cancer in their lifetime, and one in nine men and one in twelve women die of the disease. In terms of lung cancer alone, 2.5million new cases will be diagnosed by 2022, accounting for one-eighth of all cancers worldwide. Lung cancer is also the leading cause of deaths by cancer site, accounting for about 1.8 million deaths [1]. This means that many people die from lung cancer each year, requiring early detection, early treatment, and effective therapy. In order to provide effective treatment, testing for the presence of driver gene mutations may be performed. Driver genes are a general term for genes involved in the development and progression of cancer. If a mutation in this gene is found, it allows the use of molecularly targeted drugs that can have a dramatic effect on cancer treatment [2]. This therapy is less stressful on the body and more effective than conventional anticancer drugs. However, testing for the presence of genetic mutations is usually done by biopsy, which is invasive for the patient [3]. Furthermore, it is difficult for physicians to confirm the presence or absence of genetic mutations from CT images. Therefore, to reduce the burden on physicians and patients, it is necessary to develop a computer-aided diagnosis [4] system that noninvasively classifies the presence or absence of EGFR gene mutations using CT images.

Although there is study [5] on this topic, require the physician to extract lung tumor regions. This task is very burdensome for physicians. In this paper, we propose an end-to-end method to automatically extract lung tumor regions and identify the presence or absence of genetic mutations in the obtained regions.

2. Methodology

The flow of the proposed method in this paper is shown in Fig. 1. Specifically, extraction is performed using a model based on U-Net with some modifications to optimally extract lung tumor regions. Next, radiomics features are extracted from those regions and dimensionality reduction is performed using a combination of null importance, Variance Inflation Factor and Recursive Feature Elimination. Then, LightGBM was used to classify the presence or absence of genetic mutations.

2.1. *Extraction of Lung Tumor*

In this paper, our proposed prior method, Improved U-Net [6], is used as the base model. This method introduces MultiRes Block [7] and CBAM (Convolutional Block Attention Module) [8] to U-Net [9]. These modifications allow us to extract features from multiple scales and to identify which features to focus on in those features. In this paper, we further applied ASPP (Atrous Spatial Pyramid Pooling) and ensemble learning

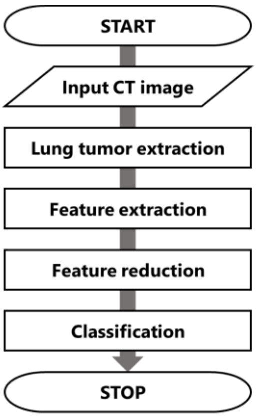


Fig.1. The flowchart of the method

to this model. ASPP enables the acquisition of diverse contextual information without increasing the number of parameters, allowing for more precise identification of the complex internal structure of tumors and their boundaries with surrounding lung tissue. Additionally, by ensembling two models with different layers, we aimed to improve the extraction accuracy of small lung tumors. Using this enhanced U-Net model, we performed automatic extraction lung tumors.

2.2. *Feature extraction*

In this paper, radiomics features were extracted from DICOM data for regions automatically segmented using an improved U-Net. Radiomics features enable the extraction of shape, intensity, and texture features from radiological images [10]. Compared to traditional biopsy-based analyses, this approach allows for the extraction of high-dimensional quantitative tumor characteristics with reduced burden on patients. Additionally, to analyze CT images across different frequency components, radiomics features were extracted from images processed with wavelet transformation.

2.3. *Feature reduction*

Feature reduction is the process of selecting only the important features among those obtained by feature extraction, thereby reducing the computational cost and preventing over-fitting to unnecessary noise data. In our previous study, feature reduction was performed by null importance [5]. However, most of the features selected were features obtained from wavelet transforms or texture features, which were not independent of each other. Therefore, in this paper, we devised a method to select features that are statistically independent from each other. Specifically, null importance was first applied separately to features obtained from the original images and those derived from wavelet-transformed images. This step retains broadly significant features and eliminates those that do not contribute to model learning. The separate application is necessary because applying null importance to the combined features would leave few features from the original images. Next, VIF (Variance Inflation Factor) [11] was used to remove highly correlated features among those retained from the first stage. VIF quantifies the degree of multicollinearity, with higher values indicating stronger multicollinearity. Generally, VIF values above 10 suggest significant multicollinearity. In this paper, the VIF for each feature was calculated, and features were recursively removed until all remaining features had VIF values below 10. Finally, RFE (Recursive Feature Elimination) [12] was applied to the features obtained from the VIF step to further select the most important features. RFE recursively eliminates the least important features, starting with all features as input, thereby retaining only the features that most significantly impact model performance. By combining the filter-based, wrapper-based, and embedded methods for feature reduction, it is considered possible to select features with low redundancy and balanced representation.

2.4. *Classification*

For classification, we used LightGBM [13], a gradient boosting decision tree algorithm. It is a decision tree that grows Leaf-Wise instead of Level-Wise in the process of gradient boosting, which allows for quick and accurate learning. In this paper, we further implemented a two-stage learning approach for classification. Two-stage learning is a method that connects two models in series to compensate for each model's weaknesses. First, classification is performed using LightGBM as usual. When the predicted probability falls between 0.3 and 0.7, it is considered to have low confidence. Data that could not be accurately classified in this range are collected, and a second model is constructed to reclassify these instances. Finally, the results from the two models are combined to improve the overall accuracy.

3. Results and Discussion

3.1. *Experimental and Evaluation Methods*

The images used in this paper were obtained from the University of Occupational and Environmental Health Hospital, and the lung tumor regions were annotated under the guidance of physicians. A dataset consisting of 452 chest CT images from 143 cases was used. Leave-One-Out cross-validation was performed for model validation, and the evaluation metrics used were AUC (area under the curve), accuracy, TPR (true positive rate), FPR (false positive rate). As shown in Table 1, cases classified as having genetic mutations were considered positive cases, while those classified as not having genetic mutations were considered negative cases. Accuracy, TPR, and FPR were calculated using the following equations, where a, b, c, and d are defined as shown in Table 1.

$$\text{Accuracy} = \frac{a+d}{a+b+c+d} \times 100[\%] \tag{1}$$

$$TPR = \frac{a}{a+c} \times 100[\%] \tag{2}$$

$$FPR = \frac{b}{b+d} \times 100[\%] \tag{3}$$

3.2. *Results*

In this paper, as in previous research, classification was performed using LightGBM by integrating features obtained from chest CT images with clinical information, specifically gender. Table 2 compares the results of the single stage learning method without feature reduction, the proposed method, and the methods of the previous research [5]. The previous method used null importance for feature reduction and performed single-stage learning, while the proposed method combined null importance, VIF and RFE for feature reduction and performed both single-stage and two-stage learning. The proposed method obtained AUC=0.647, accuracy=59.1%, TPR=54.3%, FPR=36.1%. Compared to the previous method, the proposed method showed a 2.2% decrease in accuracy and a 3.1% decrease in TPR, indicating a decrease in discrimination accuracy. Comparing single-stage and two-stage learning, the two-stage learning reduced the accuracy by 1.4% and the TPR by 0.8%.

3.3. *Discussion*

In this paper, we adopted a multi-step feature selection method that combines Null Importance, VIF, and RFE. Table 3 shows the types of features selected by Null Importance and the proposed method. Table 3 shows that the number of features obtained from the original image increased with the proposed method, and the use of VIF reduced the correlation between features. On the other hand, many features were still selected from wavelet transforms and texture features, suggesting that features obtained from wavelet transforms are important for the model to discriminate the presence of genetic mutations. In addition, Table 2 shows that there is no significant difference in accuracy between the cases where feature selection was performed and those where no feature selection was performed. This may be because the extracted radiomics features did not contain many significant features for classification. Normally, the purpose of feature reduction is to reduce noise features that do not contribute to classification, but if the original data lacks important information, the reduced features are unlikely to help with classification. Therefore, it is assumed that this feature reduction did not lead to an improvement in classification accuracy. Furthermore, no improvement in accuracy was achieved with two-stage learning.

Table 1. Valuation basis

	Predicted Positive	Predicted Negative
Test Positive	*a*	*c*
Test Negative	*b*	*d*

The purpose of two-stage learning is to focus on data that was difficult to classify in the first stage and improve classification accuracy. However, because the first stage was not properly trained, a large amount of data was transferred to the second stage, and as a result, the second stage may not have been properly trained as well. In particular, the second stage used the same features and hyperparameters as the first stage, which may have limited learning for difficult data. For two-stage learning to be effective, the first stage should be properly trained, and the features and parameters appropriate for the second stage should be reviewed. In a previous study, binary classification of the presence or absence of a genetic mutation based on manually extracted lung tumor regions had an accuracy of 92%. Although this method is very accurate, it has difficulties in practical application because it requires manual extraction of regions. On the other hand, this study proposed an end-to-end classification method and obtained an accuracy of 61.5%. This accuracy is still low and needs further improvement. In this study, we used 2D CT images to extract features, but the features obtained from 2D images were limited, which may have contributed to the lack of significant features for classification. In addition, the randomly selected slices included some cases with small tumor cross-sections, which probably made it more difficult to classify these cases. In the future, further improvement in accuracy is expected by automatically extracting lung tumor regions from 3D images and extracting more diverse features. In addition, the current dataset contains only 452 images, which is very limited. In such a situation, the model may not be able to learn enough diverse information, resulting in poor generalization performance. Therefore, expanding the dataset is also an important issue for the future.

4. Conclusion

In this paper, we developed a computer-aided diagnosis (CAD) system to identify the presence or absence of EGFR mutations from thoracic CT images, providing a less invasive method for EGFR mutation detection. By automating the extraction of lung tumor regions, this system not only reduces the burden on physicians, but also provides a non-invasive approach for patients. For feature selection, a achieved a classification performance of AUC = 0.647, accuracy = 59.1%, TPR = 54.3%, and FPR = 36.1%. To further improve the classification performance, we plan to improve the region extraction model, introduce new methods for optimal feature selection, and expand the dataset.

Table 2. Classification result（Acc.：Accuracy）

	TPR	FPR	Acc.	AUC
No feature reduction	51.4	30.0	61.5	0.661
Previous method	57.4	34.5	61.3	0.651
Single stage	55.1	33.9	60.5	0.654
Proposed method	54.3	36.1	59.1	0.647

Table 3. Selected Features Comparison

		shape	firstorder	texture
Previous method	Original	1	0	1
	Wavelet	-	5	11
Proposed method	Original	1	1	3
	Wavelet	-	3	8

5. References

1. F. Bray, M. Laversanne, H. Sung, J. Farlay, RL. Siegel, I. Soerjomataram and A. Jemal, Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, CA: A cancer Journal for Clinicians, Vol. 74(3); 2024, pp.229-263.
2. G. Q. Zhu, M. S. Zhang, X. X. Ding, B. He and Q. H. Zhang, Driver Genes in Non-small Cell Lung Cancer: Characteristics, Detection Methods, and Targeted Therapies, Oncotarget, Vol. 8(34); 2017, pp. 57680-57692.
3. J. Marrugo-Ramirez, M. Mir and J. Samitier, Blood-based Cancer Biomakers in Liquid Biopsy: A Promising Non-invasive Alternative to Tissue Biopsy, International Journal of Molecular Sciences, Vol. 19(10); 2018, p. 2877.
4. K. Doi, Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Feature Potential, Computerized Medical Imaging and Graphics, Vol. 31(4-5); 2007, pp. 198-211.
5. S. Watanabe, T. Kamiya, T. Terasawa and T. Aoki, Classification of Driver Gene Mutations from 3D-CT Images Based on Radiomics Features, In 2023 23rd International Conference on Control, Automation and Systems (ICCAS), 2023, pp. 1733-1736.
6. R. Takahashi, T. Kamiya, T. Terasawa and T. Aoki, Extraction of Lung Tumor Regions from Thoracic CT Images Using An Improved U-Net, In 2023 23rd International Conference on Control, Automation and Systems (ICCAS), 2023, pp. 1489-1493.
7. N. Ibtehaz and MS. Rahman, MultiResUNet: Rethinking the U-Net architecture for Multimodal Biomedhical Image Segmentation, Neural Networks, Vol. 121; 2020, pp. 74-87.
8. S. Woo, J. Park, JY. Lee, IS. Kweon, Cbam: Convolutional Block Attention Module, Proceeding of the European Conference on Computer Vision (ECCV), 2018, pp. 3-19.
9. O. Ronneberger, P. Fischer and T. Brox, U-net: Convolutional Networks for Biomedical Image Segmentation, Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234-241.
10. R. Thawani, M. Mclane, N. Beig, S. Ghose, P. Prasanna, V. Velcheti and A. Madabhushi, Radiomics and Radiogenomics in Lung Cancer: A review for The Clinician, Lung Cancer, Vol. 40(4); 2018, pp. 34-41.
11. M. O. Akinwande, H. G. Dikko and A. Samson, Variance Inflation Factor: As A Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis, Open Journal of Statistics, Vol. 5(7); 2015, pp. 754-768.
12. I. Guyon, J. Weston, S. Barnhill and V. Vapnik, Gene Selection for Cancer Classification Using Support Vector Machines, Machine Learning, Vol. 46; 2002, pp. 389-422.
13. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye and T. Y. Liu, LightGBM: A Highly Efficient Gradient Boosting Decision Tree, Proceedings of the 31st Advances in Neural Information Processing Systems (NeurlPS), Vol. 30, Curran Associates Inc., pp. 3149-3157.

Authors Introduction

Reo Takahashi

He is currently a master student in Kyush Institute of Technology, Japan.

Tohru Kamiya, Prof., Ph.D.

He received his B.A. degree in Electrical Engineering from Kyushu Institute of Technology in 1994, the Master and Ph.D. degree from Kyushu Institute of Technology in 1996 and 2001, respectively. He is a professor in the Department of Mechanical and Control Engineering at Kyushu Institute of Technology. His research interests are focused on image processing and medical application of image analysis. He is currently working on computer aided diagnosis based on CT, MR imaging, fluorescence microscope imaging, and automatic classification of respiratory sound.

Takashi Terasawa, M.D., Ph.D.

He received his M.D. degree from University of Occupational and Environmental Health, Japan in 2011 and Ph.D. in medicine in 2020. He is a radiologist at the University of Occupational and Environmental Health, Japan.

Takatoshi Aoki, Prof. M.D., Ph.D.

He is a professor of Radiology at University of Occupational and Environmental Health, Japan, and the vice president of Japanese Society of Thoracic Radiology. His clinical and research interests include the imaging modalities of lung cancer, respiratory function.

Detection of Lung Nodules from Temporal Subtraction CT Image Using Elastic Net-Based Features Selection

Natsuho Baba[1], Tohru Kamiya[1], Takashi Terasawa[2], Takatoshi Aoki[2]
[1] *Kyushu Institute of Technology, 1-1 Sensui, Tobata, Kitakyushu, Fukuoka, 804-8550, Japan*
[2] *University of Occupational and Environmental Health, 1-1, Iseigaoka, Yahatanishi, Kitakyushu, Fukuoka, 807-8555, Japan*
Email: baba.natsuho781@mail.kyutech.jp, kamiya@cntl.kyutech.ac.jp

Abstract

CT (computed tomography) is mainly used to diagnose lung cancer. Many CT images impose a heavy burden on visual screening, so a CAD (computer-aided diagnosis) system is expected to reduce the burden. In this paper, we propose an image analysis method to detect lung nodules from chest CT images using machine learning techniques. The best results were obtained for the method using LightGBM with feature reduction by Elastic Net.

Keywords: Computer Aided Diagnosis, Machine Learning, Temporal Subtraction Technique, Radiomics, Elastic Net.

1. Introduction

Lung cancer is the most frequently diagnosed cancer worldwide and the leading cause of cancer-related death. Currently, it is primarily diagnosed through CT (computed tomography). Many CT images generated for a patient are burden for radiologists to interpret. Furthermore, there are inherent errors in radiologists' interpretation of medical images, and not all data and image information can be discovered [1]. To address these issues, CAD (computer-aided diagnosis) systems are being developed to assist in the interpretation and diagnosis of medical images.

The temporal subtraction system, one of the CAD system, subtracts a previous image from a current one to highlight interval changes. In the temporal subtraction image, normal tissues such as blood vessels that do not change over time can be removed. In addition, it can enhance the interval changes between the previous and current images including newly appearing lesions and changes of existing lesions [2].

In this paper, we propose a method to extract and classify lung cancer lesions from chest CT images with the goal of reducing the radiologists' workload and improving detection accuracy. First, we generate temporal subtraction images and extract candidate lesions that may include lung cancer or artifacts. Next, radiomics features are extracted from these lesions, and, finally, machine learning is used to classify lesions within these lesions.

2. Methodology

In this study, the proposed method is applied to 21 cases, with images obtained using a multidetector-row CT (MDCT) scanner. Each case consists of a set of current and previous CT images of the same subject, with the temporal subtraction technique used to identify initial candidate lesions containing lung cancer or subtraction artifacts. Next, machine learning is applied to classify lesions within these initial lesions as lung cancer or non-cancer, incorporating radiomics features.

2.1. *The temporal subtraction technique*

The temporal subtraction technique involves creating a subtraction image by performing a difference operation between the current and previous images. The temporal subtraction image removes normal structures, such as blood vessels, that are commonly present in the present and past. The image highlights newly appearing lesions and changes in existing lesions over time.

2.2. *Extraction of initial candidate regions from the temporal subtraction images*

Due to the presence of many artifacts in the temporal subtraction images, mask processing is applied to extract initial candidate regions. In this study, only candidate lesion areas larger than 5 mm were targeted, and all other areas were background filled by mask processing. Machine learning was then used to classify whether these lesions were nodules or artefacts such as blood vessels.

2.3. *Extraction of features*

Radiomics features are extracted from the mask image using the selected lesion candidate regions. These features capture information such as tumor texture, intensity, heterogeneity, and shape from the chosen regions, enabling the automatic quantification and extraction of detailed features on a large scale [3]. In this study, shape and texture radiomics features were extracted using the open-source package Pyradiomics [4].

2.4. *Feature selection*

In this paper, 1026 radiomics features were computed from candidate lesion regions. However, with such a large number of features, there is an increased risk of overfitting, especially given the relatively limited amount of data. To address this, it is crucial to select only the most significant features, reducing redundant features and noise. Here, feature selection was performed using Elastic Net [5], [6], [7], a method that combines the strengths of both Lasso [5], [8] and Ridge regression [5], [9].

(i) Lasso (Least Absolute Shrinkage and Selection Operator)
Lasso [5], [8] is a powerful technique that performs both regularization and feature selection. During feature selection, variables with non-zero coefficients after reduction are retained as part of the model. Additionally, Lasso enhances model interpretability by eliminating variables that are not related to the response variable.

(ii) Ridge regression
Ridge regression [5], [9] like Lasso, is another powerful technique that performs regularization and feature selection. In this process, many predictors have non-zero coefficients and are extracted from a normal distribution. Ridge Regression is especially effective when there are many predictors with small individual effects, as it helps stabilize coefficient estimation in linear models with many correlated variables.

(iii) Elastic Net
Elastic Net [5], [6], [7] is a penalized linear regression model that applies a combination of Lasso and Ridge regression penalties to the loss function during training. It performs both feature selection, a characteristic of Lasso, and coefficient stabilization, a characteristic of Ridge regression.

2.5. *Machine learning classification*

In this paper, the features selected by Elastic Net are used as explanatory variables in supervised learning, and the machine learning methods XGBoost and LightGBM are applied for comparison and validation.

(i) XGBoost (Extreme Gradient Boosting)
XGBoost [10], [11] is a classification and regression method based on gradient boosting decision trees. It offers strong generalization capabilities, high scalability, and fast computation. The model is built through boosting with decision trees as the base learners.

(ii) LightGBM (Light Gradient Boosting Machine)
LightGBM [12], [13] similar to XGBoost, is a classification and regression method based on gradient boosting decision trees. Compared to traditional gradient boosting methods, LightGBM achieves faster learning speeds by using less memory, thus reducing computational complexity in parallel learning.

3. Experiments and Results

3.1. *Evaluation*

The evaluation metrics used are accuracy, TPR (True Positive Rate), FPR (False Positive Rate), and AUC (Area Under the Curve). As shown in Table 1, cases classified as lesions are considered positive cases, while cases classified as structures like blood vessels and the chest wall, representing normal tissues, are considered negative cases. The formulas for calculating accuracy, TPR, and FPR are provided in equations (1), (2), and (3). AUC represents the area under the receiver operating characteristic curve.

$$Accuracy = \frac{TN + TP}{TP + FP + FN + TN} \quad (1)$$

$$TPR = \frac{TP}{TP + FN} \quad (2)$$

$$FPR = \frac{FP}{TN + FP} \quad (3)$$

3.2. *Results*

In this study, temporal subtraction images were generated from 21 lung cancer cases acquired with an MDCT scanner. A total of 231 candidate lesion images obtained from these subtraction images were used as input data for machine learning classification. In this dataset, nodules that appeared over time were classified as positive lesions, while artifacts such as blood vessels and bones were classified as negative lesions. XGBoost and LightGBM were used individually for classification. Table 2 compares the methods without feature selection, with feature selection by Lasso (based on the previous method [14]), and with feature selection by Elastic Net. From Table 2, the best results were achieved using feature selection with Elastic Net and classification with LightGBM.

Example images from the experiment are shown in Figure 1. In this figure, (a) is the current image, (b) is the previous image, (c) is the temporal subtraction image, and (d) is the mask image generated using the temporal subtraction image. Arrows indicate lung cancer areas and image (d) shows the area is enhanced as a white area.

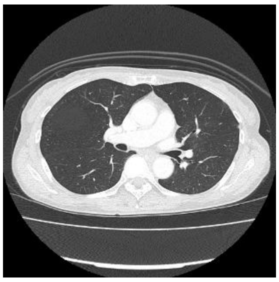
Table 1. Evaluation basis

	Predicted Positive	**Predicted Negative**
True Label Positive	True Positive	False Negative
True Label Negative	False Positive	True Negative

Table 2. Identification result (Acc. : Accuracy)

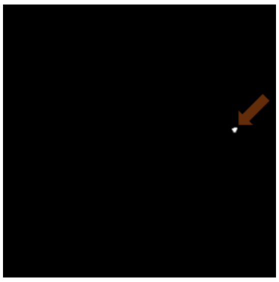
		TPR	FPR	Acc.	AUC
XGBoost	No feature reduction	67.29	30.60	67.11	0.721
	Lasso	67.90	26.40	70.12	0.786
	Elastic Net	83.60	21.04	80.96	0.836
LightGBM	No feature reduction	77.77	20.33	77.96	0.799
	Lasso	75.47	20.54	76.64	0.792
	Elastic Net	81.69	21.41	79.24	0.835

(a) The current image



(b) The previous image

(c) The temporal subtraction image



(d) The mask image

Fig.1 Example images from the experiment

4. Discussion

Table 2 shows that overall accuracy is higher when feature selection is performed using Elastic Net compared to when it is not applied. This is because Elastic Net combines the coefficient shrinkage of ridge regression and the feature selection capability of Lasso, effectively removing redundant features. Comparing the previous method [14] with the Elastic Net-based feature selection, the proposed method extracted features from similar wavelet transform statistics and sub-bands, preserving important features even in the presence of high correlations among predictors. This confirms that Elastic Net feature selection could retain relevant features without losing any critical information.

The images in Figure 1 illustrate improved classification results after switching from Lasso-based feature selection to Elastic Net-based selection. In the temporal subtraction image (c) generated from the current image (a) and the previous image (b), many artifacts remained when classification was performed using Lasso-based feature selection. However, after switching to Elastic Net feature selection, the region was correctly classified as lung cancer. This suggests that Elastic Net feature selection improved accuracy by preserving useful features and minimizing the influence of normal tissue features, even in noise-rich temporal subtraction images.

5. Conclusion

In this paper, we proposed a method for extracting and classifying lesions from chest CT images to reduce radiologists' workload and improve lesion detection accuracy. The method involves the following steps. First, initial candidate regions were identified using temporal subtraction on current and previous chest CT images. Next, radiomics features were extracted from these candidate regions, and feature selection was performed using Elastic Net to retain only significant features. Finally, the selected features were used as explanatory variables, and supervised learning was applied to classify unknown data as either lesion or normal tissue.

The proposed method was evaluated on 231 candidate lesion areas, with an AUC of 0.835, accuracy of 79.24%, TPR of 81.69%, and FPR of 21.41% when using LightGBM with Elastic Net-based feature selection.

Future work will focus on expanding the dataset, identifying optimal feature selection methods, and implementing ensemble learning with XGBoost and LightGBM for classification.

References

1. F. Shariaty and M. Mousavi, “Application of CAD systems for the automatic detection of lung nodules”, Informatics in Medicine Unlocked, Vol. 20, 2020, pp. 391-397.
2. Y. Yoshino et al., “Automatic classification of lung nodules on MDCT images with the temporal subtraction technique”, International Journal of Computer Assisted Radiology and Surgery, Vo. 12, 2017, pp. 1789-1798.
3. M. E. Mayerhoefer et al., “Introduction to Radiomics”, Journal of Nuclear Medicine, Vol. 61, No. 4, 2020, pp. 488-495.
4. J. J. M. van Griethuysen et al., “Computational radiomics system to decode the radiographic phenotype”, Cancer Research, Vol. 77, 2017, pp. 104-107.
5. J. O. Ogutu et al., “Genomic selection using regularized linear regression models: ride regression, lasso, elastic net and their extensions”, BMC Proceedings, Vol. 6, No. 10, 2012.
6. J. Morgado et al., “Machine learning and feature selection methods for EGFR mutation status prediction in lung cancer” Applied Sciences, Vol. 11, 2021.

7. H. Zou and T. Hastie, “Regularization and variable selection via the elastic net” Statistical Methodology, Vol. 67, 2005, pp. 301-320.
8. V. Fonti and E. Belister, “Feature selection using lasso”, VU Amsterdam Res. Paper Business Anal, Vol. 30, 2017, pp. 1-25.
9. M. Togacar et al., “Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and liner discriminant analysis using invasive breast cancer images processed with autoencoders”, Medical Hypotheses, Vol. 135, 2020, pp. 135-145.
10. T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system”, Machine Learning, Vol. 3, 2016, pp. 785-794.
11. S. Li and X. Zhang, “Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm”, Neural Computing and Applications, Vol. 32, 2020, pp. 1971-1979.
12. J. Cai et al., “An assembly-level neutronic calculation method based on LightGBM algorithm”, Annals of Nuclear Energy, Vol. 150, 2021, pp. 871-881.
13. Y. Wang and T. Wang, “Application of improved LightGBM model in blood glucose prediction”, Applied Sciences, Vol. 10, No. 9, 2020, pp. 3227-3243.
14. N. Baba et al., “A detection method for nodular shadows from temporal subtraction images using a machine learning technique incorporated radiomics features”, Internatinal Conference on Control, Automation and Systems, 2023, pp. 871-881.

Authors Introduction

Ms. Natsuho Baba

She received her Bachelor’s degree in Engineering in 2023 from the Faculty of Engineering, Kyushu Institute of technology in Japan. She is currently a master student in Kyushu Institute of Technology, Japan.

Tohru Kamiya, Prof. Ph.D

He received his B.A. degree in Electrical Engineering from Kyushu Institute of Technology in 1994, the Master and Ph.D. degree from Kyushu Institute of Technology in 1996 and 2001, respectively. He is a professor in the Department of Mechanical and Control Engineering at Kyushu Institute of Technology. His research interests are focused on image processing, medical application of image analysis, etc.

Takashi Terasawa, M.D., Ph.D

He received his M.D. degree from University of Occupational and Environmental Health, Japan in 2011 and Ph.D. in medicine in 2020. He is a radiologist at the University of Occupational and Environmental Health, Japan.

Takatoshi Aoki, Prof. M.D., Ph.D

He is a professor of Radiology at University of Occupational and Environmental Health, Japan, and the vice president of Japanese Society of Thoracic Radiology. His clinical and research interests include the imaging modalities of lung cancer, respiratory function.

Detection of Lung Nodules from CT Image Based on Ensemble Learning

Natsuho Baba[1], Tohru Kamiya[1], Takashi Terasawa[2], Takatoshi Aoki[2]
[1] *Kyushu Institute of Technology, 1-1 Sensui, Tobata, Kitakyushu, Fukuoka, 804-8550, Japan*
[2] *University of Occupational and Environmental Health, 1-1, Iseigaoka, Yahatanishi, Kitakyushu, Fukuoka, 807-8555, Japan*
Email: baba.natsuho781@mail.kyutech.jp, kamiya@cntl.kyutech.ac.jp

Abstract

Lung cancer is the most frequently diagnosed cancer worldwide and the leading cause of cancer-related deaths, making early detection and treatment crucial. Temporal subtraction system, one of the CAD, emphasize the differences between the current and previous images. In this study, radiomics features are extracted as explanatory variables from the temporal subtraction images. Feature selection is performed using Elastic Net, followed by the application of machine learning methods. Finally, ensemble learning is applied to classify unknown data into two categories: positive and negative lung nodules.

Keywords: Computer Aided Diagnosis, Machine Learning, Temporal Subtraction Technique, Radiomics, Ensemble Learning.

1. Introduction

According to GLOBOCAN 2022, there were 2.2 million new lung cancer cases (11.4%) and an estimated 1.8 million deaths from lung cancer in 2020 [1]. Lung cancers at early stages often appear as nodules include ground-glass opacities (GGO). Because these lesions are low density or small size, there is concern that they may go undetected by the interpreting physician.

One of the computer-aided diagnosis (CAD) systems that addresses this challenge is the temporal subtraction imaging technique. In this method, a temporal subtraction image is created by performing a subtraction operation between current and previous images of the same patient. The temporal subtraction image technique erases normal tissues, such as blood vessels and ribs, that do not change over time. It also highlights newly developed lesions or changes in existing abnormal tissues by comparing them to previous images [2]. The studies have reported that the use of temporal subtraction images can improve the accuracy of physician interpretation [3], [4].

In this paper, we propose a method for extracting and classifying lung cancer lesions from chest computed tomography (CT) images with the aim of reducing the workload of radiologists and improving detection accuracy. First, temporal subtraction images are generated, and a bounding box is created for each candidate region, which may contain both lung cancer lesions and artifacts. Next, radiomics features are extracted from these images as explanatory variables, and feature selection is performed using Elastic Net. Machine learning methods are then applied to classify the data. Finally, ensemble learning is used to combine the results of the machine learning models for the final classification.

2. Methodology

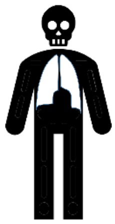
In this paper, the proposed method is applied to a dataset of 21 cases, each consisting of images acquired with a multidetector row CT scanner. A set of current and previous images of the same patient is considered one case.

2.1. *The temporal subtraction technique*

One of the CAD techniques anticipated for use in comparative reading in chest CT imaging is the temporal subtraction technique. This technique highlights temporal changes by subtracting two medical images that have been acquired at different times [5]. In addition, normal structures such as ribs and blood vessels are effectively removed, facilitating detection of newly emerging lung lesions [6].

Figure 1 shows an image of the temporal subtraction technique. In the figure, (a) is the original image, (b) is the previous state image, and (c) is the result image where image (b) has been subtracted from image (a), leaving only the lesion candidate area. Image (c) shows that a newly appeared lesion is enhanced and unchanged structures such as blood vessels, bones, and muscles are removed.

(a) The current state image



(b) The previous state image

(c) The temporal subtraction

Fig.1 The image of the temporal subtraction technique

2.2. *Extraction of candidate lesion areas from the temporal subtraction images*

The temporal subtraction images contain numerous residual artifacts. Therefore, in this study, regions with a diameter of 5 mm or greater are extracted from the generated temporal subtraction images as candidate lesion regions. Additionally, the extracted candidate lesion area is enlarged vertically and horizontally by 2 mm, and a Bounding Box is defined as the final region of interest (ROI). In this study, two methods are used: one is to use lesion candidate regions with a diameter of 5 mm or more, and the other is to use features extracted from a Bounding Box with an enlarged region.

2.3. *Radiomics features*

Radiomics features are quantifiable data derived from medical images using mathematical formulas and computer-based algorithms. These features capture and provide detailed information about the shape, brightness, and texture of lesions in medical images [7]. In this study, shape and texture features were extracted using the open-source library Pyradiomics [8].

2.4. *Feature selection*

Feature selection is a process that identifies only significant features from a dataset to reduce computational cost and minimize overfitting due to irrelevant noise. In this paper, we use a method called Elastic Net [9], [10] to identify the most relevant features for classification from the set of 1026 radiomics features.

Elastic Net is a penalized linear regression model that applies a linear combination of regularization penalties to the loss function during training. It is an analytical method that combines the characteristics of Ridge regression and Lasso (Least absolute shrinkage and selection operator), allowing it to perform both automatic feature selection, a characteristic of Lasso, and continuous shrinkage, a characteristic of Ridge regression, simultaneously.

2.5. *Machine learning classification*

In this paper, the features selected by Elastic Net are used as explanatory variables and machine learning methods. Three machine learning methods are used: XGBoost, LightGBM, and CatBoost. While machine learning is powerful due to its ability to handle numerous features, it involves many hyperparameters that significantly affect model performance. Optimizing these hyperparameters is crucial for improving the accuracy of machine learning models. In this study, hyperparameter tuning is performed using Optuna [11], a black-box optimization framework. Finally, ensemble learning is used to combine the results of the three machine learning methods to generate the final output.

(i) XGBoost (Extreme Gradient Boosting)
XGBoost [12], [13] is a gradient-boosted decision tree (GBDT) method for based classification and regression introduced in 2014. It is characterized by fast processing speed, high accuracy, and flexibility in data requirements. Its strengths include strong model generalization ability, high scalability, and fast computation.

(ii) LightGBM (Light Gradient Boosting Machine)
LightGBM [14], [15], developed in 2017, is a classification and regression method based on GBDT, similar to XGBoost. LightGBM is known for faster speed and lower memory consumption compared to XGBoost, achieving high efficiency with large datasets while maintaining high accuracy with smaller datasets.

(iii) CatBoost (Categorical Boosting)
CatBoost [16], [17], developed in 2018, is another classification and regression method based on GBDT. CatBoost excels in handling categorical variables and includes a built-in cross-validation function to mitigate overfitting. In addition, it performs well when handing outliers and missing values in the data.

(iv) Ensemble Learning
Ensemble learning [18], [19] is a technique that combines predictions from multiple models to improve reliability and generalizability. This approach reduces the risk of overfitting while maintaining strong predictive performance by averaging predictions from different models. Ensemble learning methods include bagging and stacking, which use boosting algorithms. These models are trained in parallel on different random subsets of data using alternative sampling, and the predictions from all models are aggregated.

3. Experimental Results

In this study, temporal subtraction images were generated from 21 lung cancer cases. Multiple slice images with lesions are selected and their performance is evaluated by 4-fold cross-validation using 231 candidate lesions. In the temporal subtraction images, newly appearing nodules over time were classified as positive lesions, while artifacts such as blood vessels and ribs were classified as negative lesions. The distribution of positive and negative lesions was 111 positive lesions and 120 negative lesions. The evaluation metrics were accuracy, true positive rate (TPR), false positive rate (FPR), and area under the curve (AUC). The results of the classification are shown in the table, with the input data in Table 1 as candidate lesion areas obtained from the temporal subtraction images and the input data in Table 2 as ROIs. Both Table 1 and Table 2 shows the results for methods that performed classification without feature reduction, methods that used Elastic Net for feature selection in previous study [20], and methods that incorporated ensemble learning.

Figure 2 shows an example image of the experimental results with the input data as bounding boxes. In Figure 2, (a) is the current image, (b) is the previous image, (c) is the temporal subtraction image, (d) is the mask image generated using the temporal subtraction image, (e) is a magnified view focusing on the Bounding Box of the current image, and (f) is a magnified view focusing on the Bounding Box of the temporal subtraction image. Arrows indicate lung cancer areas. The rectangular areas at (e) and (f) are the ROIs.

4. Discussions

Table 1 shows that ensemble learning achieves the best overall results when the lesions obtained from the temporal subtraction images are used as input data. In contrast, Table 2 shows that when the ROI, Bounding Box, is used as input data, the accuracy of ensemble learning is slightly less than that of LightGBM. This result is likely due to the greater variation in the accuracy of the three machine learning models in Table 2 compared to those in Table 1.

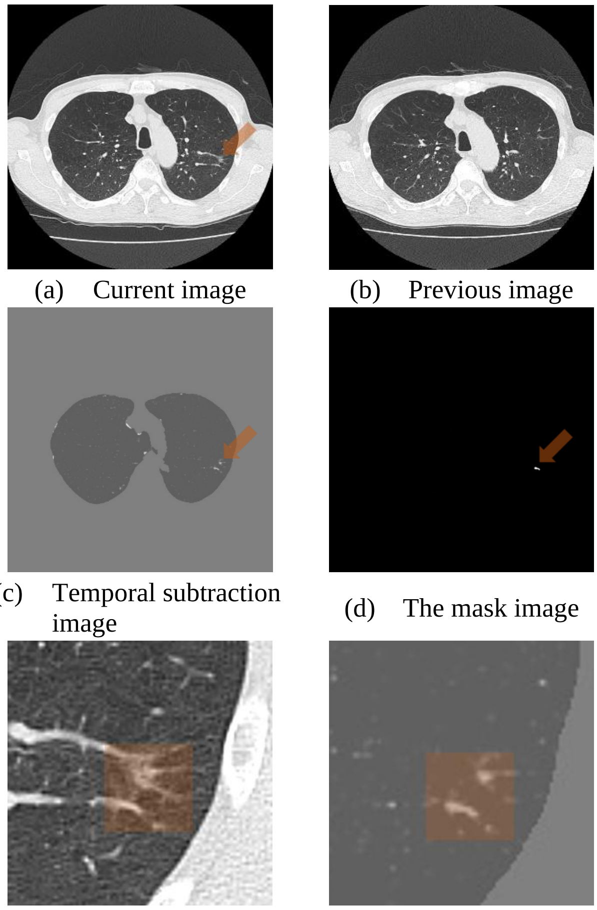
From Tables 1 and 2, the results in Table 2, where the lesion is changed to a ROI, show an overall improvement. Figure 2 also shows an example of improved discrimination by changing the input data from lesion candidate areas to ROIs. In images (e) and (f), the lesion candidate regions obtained from the temporal subtraction images were used as input data, so they were misclassified as blood vessels by feature extraction based on the shape of the region. However, by changing the input data to the bounding box, we were able to extract the features of regions containing light whitish areas, which is a characteristic of GGO, and we believe that we were able to correctly classify the data.

Table 1. Identification result from regions of interest (Acc.: Accuracy)

		TPR	FPR	Acc.	AUC
XGBoost	No feature reduction	67.29	30.6	67.11	0.721
	Elastic Net	83.60	21.04	80.96	0.836
LightGBM	No feature reduction	77.77	20.33	77.96	0.799
	Elastic Net	81.69	21.41	79.24	0.835
CatBoost	No feature reduction	68.59	26.97	70.15	0.782
	Elastic Net	80.82	28.30	75.35	0.812
Ensemble Learning		84.68	26.67	78.79	0.843

Table 2. Identification result from bounding box (Acc.: Accuracy)

		TPR	FPR	Acc.	AUC
XGBoost	No feature reduction	85.32	15.98	84.39	0.904
	Elastic Net	80.95	19.41	80.07	0.897
LightGBM	No feature reduction	87.65	10.78	87.87	0.917
	Elastic Net	90.92	11.92	89.17	0.949
CatBoost	No feature reduction	85.99	16.08	84.83	0.913
	Elastic Net	87.38	13.99	86.15	0.947
Ensemble Learning		90.09	15.00	87.45	0.939



(a) Current image
(b) Previous image
(c) Temporal subtraction image
(d) The mask image
(e) Enlarged image of the current image
(f) Enlarged image of the temporal subtraction image

Fig.2 Experimental results with bounding box

5. Conclusion

In this paper, we proposed a method for extracting and classifying lesions from chest CT images with the aim of reducing radiologist workload and improving lesion detection accuracy. The proposed method was applied to ROIs obtained from temporal subtraction images of 21 cases and achieved AUC 0.939, accuracy 87.45%, TPR 90.09%, and FPR 15.00% by ensemble learning. Future work includes further improving ensemble learning techniques, expanding the dataset, refining the feature selection methods, and optimizing the ensemble classification using XGBoost and LightGBM.

References

1. F. Bray et al., "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries", A Cancer Journal for Clinicians, Vol. 74, No. 3, pp. 229-263, 2024.
2. T. Aoki et al., "CT temporal subtraction: techniques and clinical applications", Quant Imaging Med Surg, Vol. 11, No. 6, pp. 2214-2223, 2021.
3. M. C. Difazio et al., "Digital chest radiography: Effect of temporal subtraction images on detection accuracy", Radiology, Vol. 202, No. 2, pp. 447-452, 1997.
4. N. Miyake et al., "Optimizing early cancer diagnosis and detection using a temporal subtraction technique", Technological Forecasting and Social Change, Vol. 167, 2021.
5. T. Aoki., "Clinical usefulness of temporal subtraction system for lung nodule detection", Medical Imaging and Information Science, Vol. 30, No. 2, pp. 36-38, 2013.
6. Y. Tanabe and T. Ishida., "Development of a novel detection method for changes in lung conditions during radiotherapy using a temporal subtraction technique", physical and Engineering Sciences in Medicine, Vol. 44, pp. 1341-1350, 2021.
7. A. A. Ardakani et al., "Interpretation of radiomics features – A pictorial review", Computer Methods and programs in Biomedicine, Vol. 215, 2022.
8. J. J. M. van Griethuysen et al., "Computational radiomics system to decode the radiographic phenotype", Cancer Research, Vol. 77, pp. 104-107, 2017.
9. H. Zou and T. Hastie, "Regularization and variable selection via the elastic net" Statistical Methodology, Vol. 67, pp. 301-320、2005.
10. H. Zou and H. H. Zhang, "On the adaptive elastic-net with a diverging number of parameters", The Annals of Statistics, Vol. 37, No. 4, pp1733-1751, 2009.
11. T. Akiba et al., "Optuna: A next-generation hyperparameter optimization framework", ACM SIGKDD International Conference, pp. 2623-2631, 2019.
12. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system", Machine Learning, Vol. 3, 2016, pp. 785-794.
13. X. Guan et al., "Construction of the XGBoost model for early lung cancer prediction based on metabolic indices", BMC Medical Informatics and Decision Making, Vol. 23, 2023.
14. G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree", Advances in Neural Information Processing Systems, Vol. 30, pp. 3149-3157, 2017.
15. E. A-R. Hamed et al., "An efficient combination of convolutional neural network and LightGBM algorithm for lung cancer histopathology classification", Diagnostics, Vol. 13, No. 15, 2023.
16. L. Prokhorenkova et al., "CatBoost: unbiased boosting with categorical features", Neural Information Processing Systems, pp. 6639-6649, 2018.
17. J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review", Journal of Big Data, Vol. 7, No. 94, 2020.
18. J. M. Ahn et al., "Ensemble machine learning of gradient boosting (XGBoost, LightGBM, CatBoost) and attention-based CNN-LSTM for harmful algal blooms forecasting", Toxins, Vol 15, No. 10, pp.608-623, 2023.
19. Y. Zhang et al., "A review of ensemble learning algorithms used in remote sensing applications", Appled Sciences, Vol. 12, No. 17, pp. 8654-8674, 2022.
20. N. Baba et al., "A detection method for nodular shadows from temporal subtraction images using a machine learning technique incorporated radiomics features", International Conference on Control, Automation and Systems, 2023, pp. 871-881.

Authors Introduction

Ms. Natsuho Baba

She received her Bachelor's degree in Engineering in 2023 from the Faculty of Engineering, Kyushu Institute of technology in Japan. She is currently a master student in Kyushu Institute of Technology, Japan.

Tohru Kamiya, Prof. Ph.D

He received his B.A. degree in Electrical Engineering from Kyushu Institute of Technology in 1994, the Master and Ph.D. degree from Kyushu Institute of Technology in 1996 and 2001, respectively. He is a professor in the Department of Mechanical and Control Engineering at Kyushu Institute of Technology. His research interests are focused on image processing, medical application of image analysis, etc.

Takashi Terasawa, M.D., Ph.D

He received his M.D. degree from University of Occupational and Environmental Health, Japan in 2011 and Ph.D. in medicine in 2020. He is a radiologist at the University of Occupational and Environmental Health, Japan.

Takatoshi Aoki, Prof. M.D., Ph.D

He is a professor of Radiology at University of Occupational and Environmental Health, Japan, and the vice president of Japanese Society of Thoracic Radiology. His clinical and research interests include the imaging modalities of lung cancer, respiratory function.

Shape-Preserving Embedding Technique for Binary Classification of Video Image of the Solar Surface

Iori Tamura
Kyushu Institute of Technology, 680-4, Kawazu, Iizuka, Fukuoka, 820-8502, Japan

Akiko Fujimoto
Kyushu Institute of Technology, 680-4, Kawazu, Iizuka, Fukuoka, 820-8502, Japan

Soichiro Kondo
Kyushu Institute of Technology, 680-4, Kawazu, Iizuka, Fukuoka, 820-8502, Japan

Reiri Noguchi
Kyushu Institute of Technology, 680-4, Kawazu, Iizuka, Fukuoka, 820-8502, Japan
Email: tamura.iori306@mail.kyutech.jp, fujimoto@ai.kyutech.ac.jp, kondo.soichiro828@mail.kyutech.jp, noguchi.reiri510@mail.kyutech.jp

Abstract

We study the embedding technique on the binary classification of video images as the explanatory variable. In this study, we assume the shape on video frame image has high sparsity and strong characteristic time evolution. In the embedding process, 2-dimensional image is resized keeping shape characteristics of the image and converted to a vector. The embedding allows dimensionality reduction from a 3-dimensional array (video image) as input data for machine learning to a 2-dimensional array of time sequences of embedded vectors. Using solar surface video images in the space weather field, we present evaluation experiments on multiple models with different embedding sizes, transformation formulas, and number of layers in the CNN.

Keywords: Embedding technique, Binary classification, Solar surface video images, Space weather

1. Introduction

Common applications of video analysis include object detection, object tracking, action recognition, and scene understanding. These applications are used in a wide range of fields, including security cameras, self-driving cars, sports analysis, and entertainment. Video analysis is achieved by combining the following typical deep learning algorithm methods:

- Convolutional Neural Networks (CNN) [1]: CNNs are used for image recognition in each frame to recognize objects in the video. It automatically learns local patterns in the image and is characterized by high discrimination performance.
- Recurrent Neural Network (RNN) [2], Long Short-Term Memory (LSTM) [3]: LSTM is suitable for learning time-series data and is good at capturing temporal changes in videos. It can effectively learn continuous changes of events in a video.
- 3D Convolutional Neural Network (3D-CNN) [4]: While a typical CNN learns local patterns on a 2D image, a 3D-CNN can learn local patterns on spatial and temporal axes simultaneously in a video.

For example, object detection and action recognition can achieve higher accuracy by recognizing objects in each frame using CNNs and capturing temporal changes using RNNs and LSTMs. On the other hand, 3D-CNNs can learn spatial and temporal information simultaneously, enabling more advanced analysis in video analysis.

Now we assume the object shape on video frame image has high sparsity and strong characteristic time evolution. In this case, the above-mentioned models have high computational costs. If each frame image (two-dimensional matrix) of the video is converted to a vector including the information of the space shape characteristic, this embedding allows dimensionality reduction from a 3-dimensional array (video image) as input data for machine learning to a 2-dimensional array of time sequences of embedded vectors.

In this study, we address the problem of a binary classification task using a video of the surface of the Sun. The structure of the radiation belts surrounding the Earth drastically changes due to solar activity [5]. In particular, the high-speed solar wind stream originating from recurrent coronal holes contributes to the abnormal increase of relativistic electrons in the outer radiation belts [6]. The location of coronal holes on the full disk of the Sun is strongly related to the dynamics of the radiation belts. The low-latitude coronal hole regions positively correlate with the high-speed solar wind stream associating with the relativistic electron enhancement of the radiation belts [7]. This suggests that the shape of coronal hole has high space sparsity on the task of the binary classification for the relativistic electron enhancement in the outer radiation belts using the solar surface image. In this paper, we propose a method for the vector embedding of solar coronal hole information. We also propose a binary classification model using the embedded coronal hole vector as the input data for estimating whether the relativistic electron enhancement of the radiation belts has occurred and evaluate the accuracy of our model.

2. Data

We used the two-dimensional extreme ultraviolet (EUV) images of the solar atmosphere (corona) taken by the Atmospheric Imaging Assembly (AIA) onboard NASA's Solar Dynamics Observatory (SDO) spacecraft [8]. The coronal holes appear dark in EUV SDO/AIA images. SDO/AIA 211Å images are used in this study to detect coronal hole regions (Fig. 1). The spatial and time resolution of AIA original image is 1024 × 1024 pixels and 15 minutes, respectively.

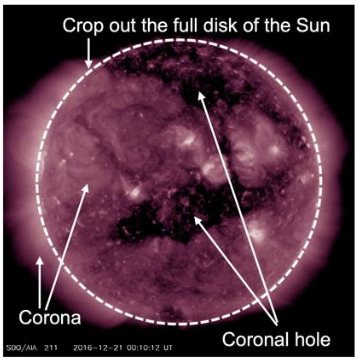


Fig.1 Example of SDO/AIA 211Å image.

For the relativistic electron flux in the Earth's outer radiation belt, the 1-minute averaged fluxes of electrons with energy >2 MeV measured by the spacecraft of the Geostationary Operational Environmental Satellite (GOES) series, GOES-15 satellite. Since the GOES-15 satellite rotates with the earth, the relativistic electron flux data shows the daily fluctuations (Fig. 2). The relativistic electron flux reaches its absolute maximum when the satellite approaches noon. In this study, when the >2 MeV electron flux exceeds 10,000 [particles/(cm^2 s sr)], we assume that an abnormal increase in the relativistic electron flux in the radiation belt has occurred.

The time interval analyzed in this paper ranges from January to December 2016, containing the declining phase (from the maximum to minimum solar activity) of Solar Cycle 24. The coronal holes generally appear with a high frequency during the declining phase. We consistently take the ratio of training and testing samples to be 11 : 1 for all models presented throughout the paper. The datasets from January to October 2016 are used as the training samples, while the datasets from November to December 2016 are used as the testing samples.

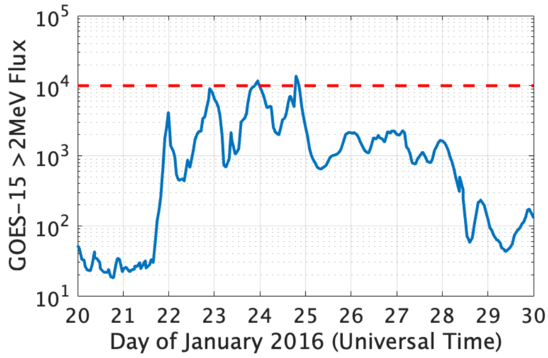


Fig.2 Example of the time-series data of GOES-15 >2 MeV electron flux.

3. Methodology

The proposed model consists of three main processes: (1) SDO/AIA images preprocessing, (2) embedding to coronal hole vectors, and (3) binary classification prediction model (Fig. 3).

3.1. *SDO/AIA images preprocessing*

To remove the non-sun areas from each frame image of the SDO/AIA video, a size-optimized circular mask is used to crop out the full disk of the sun. We assume the full disk of the Sun always locates in the center of the solar frame image (Fig. 1).

The binarization processing is applied to the masked solar frame image to divide the coronal hole regions and the corona regions. First images are converted to gray scale. Next, the pixel values in the image are converted to white if the pixel value is equal to or greater than the threshold (50), and to black if the pixel value is less than the threshold. Since coronal holes appear dark in the SDO/AIA 211Å image, black pixel areas indicate coronal hole area (Fig. 3).

3.2. *Coronal hole vector*

Here, we define the two-dimensional array of SDO/AIA images converted into a one-dimensional vector as a “corona hole vector”. The embedding process into a corona hole vector consists of two steps: image resizing and vectorization. First, the SDO/AIA binary image (1024 × 1024 pixels) is divided into $n \times n$ grids. In this process, each grid is quantified as a coronal hole region or a non-coronal hole region according to the ratio of black pixels contained in each grid. Then the $n \times n$ array is converted to a vector, size ($n \times n$, 1).

We propose two methods for the quantification of coronal hole region in each grid: (method A) Presence/Absence method, and (method B) Occurrence ratio method.

(A) Presence/Absence method

If the ratio of black pixels to the total number of pixels in a grid is greater than the threshold (10% in this paper), the value 1 is embedded, otherwise 0 is embedded. Consequently, a one-hot vector is generated.

(B) Occurrence ratio method
The ratio of black pixels in each grid is embedded in the vector elements.

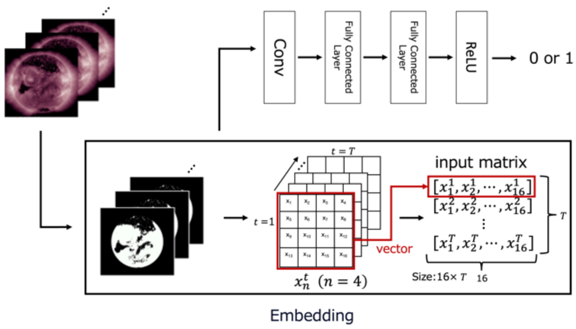


Fig.3 Coronal hole-related radiation belt electron enhancement binary classification using deep convolutional neural networks.

3.3. *Binary classification prediction model*

We propose the binary classification model for predicting the 1-day ahead relativistic electron enhancement based on coronal holes appearing in the past 5-day (120 h) SDO/AIA video images. The explanatory variable is a time series array of coronal hole vectors, size $(n \times n, T)$ and the objective variable is a label (1 or 0) indicating whether there is a relativistic electron enhancement in the radiation belts. T is the number of frame image of SDO/AIA video. The original SDO/AIA video frames are taken every 15 minutes. In this paper, we used 1-hour SDO/AIA video frames as the training samples to reduce the computational costs. Consequently, T is 120 for 5-day SDO/AIA video images.

For the binary classification task, such as the relativistic electron enhancement/non-enhancement classification, we prepare the positive/negative label data from the data of GOES-15 >2 MeV electron flux by defining the positive class (e.g., enhancement) and negative class (e.g., non-enhancement). The labels are given by

$$the\ label = \begin{cases} 1, & E \geq 10^4 \\ 0, & E < 10^4 \end{cases}, \qquad (1)$$

where E is the absolute maximum of daily GOES-15 >2 MeV electron flux [particles/(cm^2 s sr)]. $the\ label = 1$ indicates the relativistic electron enhancement occurs, while $the\ label = 0$ notes the non-enhancement of relativistic electron flux in the radiation belts. The architecture of our model is shown in Fig. 3. The binary cross-entropy loss (BCELoss) is used as the loss function, and the Adam optimizer with a learning rate of 0.0001 is employed. The batch size is set to 16, and the number of epochs is 300.

3.4. *Evaluation Metric: F_1 score*

The F_1 score is a performance metric for classification and is calculated as the harmonic mean of a model's precision and recall. F_1 score is given by

$$F_1 score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (2)$$

4. Results and Discussion

We trained models with the following different parameters of the network: two types of embedding method (method A, method B) and the split size ($n = 4, 8$) of the coronal hole vector, the length ($T = 120, 72$) of the time series of the coronal hole vector, the kernel size (3× 3, 7×7) and the number of layers of convolution neural network. Table 1 compares the F_1 score of 10 models to evaluate the suitable parameters of the classification model of the relativistic electron enhancement.

Table 1. Classification Results

Model	Coronal hole vector	n	T	Kernel size	Conv Layer	F_1 score
1	A	4	120	3 × 3	Conv1	0.2545
2	B	4	120	3 × 3	Conv1	0.5574
3	A	4	120	7 × 7	Conv1	0.5676
4	B	4	120	7 × 7	Conv1	0.3673
5	A	8	120	3 × 3	Conv1	0.3333
6	B	8	120	3 × 3	Conv1	0.3448
7	A	4	72	3 × 3	Conv1	0.3333
8	B	4	72	3 × 3	Conv1	0.3881
9	A	4	120	3 × 3	Conv5	0.4722
10	B	4	120	3 × 3	Conv5	0.2979

For the method of the coronal hole vector, the models (model 2, 6, and 8) with the Occurrence ratio method b shows higher accuracies than the Presence/Absence method A. This result suggests that the coronal hole vectors based on the Occurrence ratio method B contain more information on the two-dimensional features of coronal holes than the method A and the models can be trained effectively even with a simplified architecture. On the other hands, the coronal hole vectors based on the Presence/Absence method A lack information about the structural characteristics of the coronal holes compared to the method B and therefore require a more complex architecture model. The small number of split size and length is also suitable to the present task which is the binary classification model requiring the simple network architecture. The 5-day coronal hole vector models show good performances than the 3-day model. This suggests that 5-day model trained the dependency of the sequence of coronal hole structure.

The kernel size is also important of the complexity of the model. The results (model 3, 4) show the small kernel size such as 3 × 3 is suitable to the coronal hole vector based on the Occurrence ratio method B. This vector essentially has the spatiotemporal structure of the coronal hole and allows the small kernel size of the convolution. Thus, the model 4 resulted into the low F_1 score than model 3. In addition, the deep network (model 10) is not suitable to the coronal hole vector based on method B. The results support the idea that the adequate embedded vector allows the simple network requiring the non-deep convolution neural network.

5. Conclusion

We proposed the embedding technique on the binary classification of video images as the explanatory variable. In this study, we assume the object shape on video frame image has high sparsity and strong characteristic time evolution. We demonstrated the two embedding methods based on the spatiotemporal structure of the object shape. Our embedding method allows dimensionality reduction from a 3-dimensional array (video image) as input data for machine learning to a 2-dimensional array of time sequences of embedded vectors. Finally, a few potential limitations for the prediction model need to be considered. Future work will concentrate on the construction of the binary classification model with higher accuracy to predict the relativistic electron enhancement based on coronal holes appearing in SDO/AIA video images.

Acknowledgements

SDO/AIA images are courtesy of NASA/SDO and the AIA science teams. Thanks to the NOAA National Environmental Information Center (NCEI) for providing processed GOES series satellite data.

References

1. LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. "Gradient-based learning applied to document recognition." Proceedings of the IEEE, 86(11), pp. 2278–2324, 1998.
2. Rumelhart, D. E., Hinton, G. E. and Williams, R. J., "Learning representations by back-propagating errors". Nature 323 (6088), pp.533–536, 1986.
3. Hochreiter, S. And Schmidhuber, J. "Long short-term memory". Neural Computation, 9(8), pp. 1735–1780, 1997.
4. Ji, S.,Xu, W., Yang, M. and Yu, K. "3D Convolutional Neural Networks for Human Action Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(1), pp. 221-231, 2013.
5. Reeves, G. D., McAdams, K. L., Friedel, R. H. W. and O'Brien, T. P. "Acceleration and loss of relativistic electrons during geomagnetic storms", Geophys. Res. Lett., 30, 1529, 2003.
6. Miyoshi, Y., and Kataoka, R. "Flux enhancement of the outer radiation belt electrons after the arrival of stream interaction regions", J. Geophys. Res., 113, A03S09, 2008.
7. Nakagawa, Y., Nozawa, S. and Shinbori, A. "Relationship between the low-latitude coronal hole area, solar wind velocity, and geomagnetic activity during solar cycles 23 and 24", Earth Planets Space 71(24), 2019.
8. Lemen, J.R., Title, A.M., Akin, D.J. et al. The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO). Sol Phys 275, pp. 17–40, 2012.

Authors Introduction

Ms. Iori Tamura

She received her Bachelor's degree in Computer Science and Systems Engineering in 2024 from the Faculty of Computer Science and Systems Engineering, Kyushu Institute of technology in Japan. She is currently a master student in Kyushu Institute of Technology, Japan.

Dr. Akiko Fujimoto

She is an Associate Professor at the Faculty of Computer Science and Systems Engineering at Kyushu Institute of Technology in Japan. She graduated from the Department of Earth and Planetary Sciences, Kyushu University, in 2005. She received her Ph.D. degree in Science from Kyushu University in 2010. Her research interest is Space Weather and Upper Atmospheric Sciences.

Mr. Soichiro Kondo

He received his Bachelor's degree in Computer Science and Systems Engineering in 2024 from the Faculty of Computer Science and Systems Engineering, Kyushu Institute of technology in Japan. He is currently a master student in Kyushu Institute of Technology, Japan.

Mr. Reiri Noguchi

He is currently an undergraduate student at the Faculty of Information Engineering, Kyushu Institute of Technology, Japan.

Seated Posture Estimation Based on Monocular Camera Images

Hitoshi Shimomae, Tsubasa Esumi, Noriko Takemura
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka City, Fukuoka, 820-0067, Japan
Email: shimomae.hitoshi210@mail.kyutech.jp, esumi.tsubasa157@mail.kyutech.jp, takemura@ai.kyutech.ac.jp

Abstract

Poor seated posture significantly strains the body, leading to symptoms like shoulder stiffness and back pain. While research on seated posture estimation using images has been active, many studies focus on extreme postures not typically seen in daily desk work. This study aims to estimate common postures, such as slouching, which are often experienced in everyday settings. Due to the lack of medical quantitative metrics for evaluating posture quality, we manually annotated some of our collected posture data and used semi-automatic annotation by SVM to build a dataset. Using this dataset, we trained deep learning models for posture estimation with different input data types: RGB images, silhouette images, and posture key points, and compared their performance.

Keywords: Posture Analysis, CNN, GNN

1. Introduction

The average sedentary time in Japan is approximately seven hours, ranking among the highest in the world [1]. Prolonged sedentary periods and poor seated posture can impose significant physical strain, leading to issues such as shoulder stiffness and lower back pain. Recently, development efforts have intensified on systems that promote posture improvement, such as smart chairs [6], [11], necessitating accurate estimation of seated postures as a foundational requirement. Many conventional studies on seated posture estimation target extreme postures, such as heavy lateral tilts or excessive backward bends, which are rarely seen in regular desk work [2], [7], [9], making them unsuitable for real-world application. Therefore, we aim to estimate everyday seated postures, including rounded shoulders and lordosis, using readily deployable and cost-effective RGB cameras.

In this study, we collect posture data while working at a desk using an RGB camera and motion capture to construct a dataset. For some of the collected data, we manually annotate three levels (good posture/slightly bad posture/bad posture) by multiple people, and then train a label estimation model using the motion capture data as input, and perform automatic annotation for the remaining data. Using the created dataset, we perform 3-class classification (good posture/slightly bad posture/bad posture) of sitting postures using a deep learning model that takes RGB images and silhouette images, as well as low-dimensional posture feature points extracted using YOLO [11], as inputs, and compare and discuss the accuracy of the different features of the input data.

2. Dataset Construction

This study targets the estimation of natural seated postures that can be commonly assumed in scenarios like desk work, as shown in Fig.1. Postures like slouching are often unconsciously assumed and noticed only when pointed out by others. Therefore, we collected posture data from 30 subjects aged 18 to 24 (25 males and 5 females) during desk work tasks without instructing them to assume specific postures consciously. Additionally, we manually annotated a portion of the data and employed an SVM trained on this annotated data to automate the annotation of the remaining data, thus constructing a large-scale dataset.

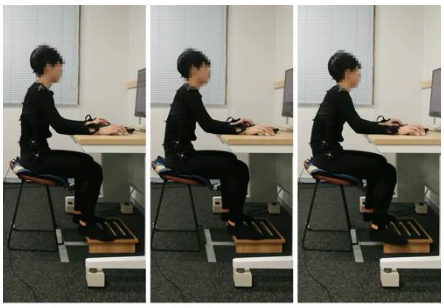


Fig.1 Examples of postures targeted in this study (collected posture data).

2.1. *Data Collection Experiment*

As a desk task, we set a task to create a report (Word, PowerPoint, etc.) while searching the Internet for a given theme. Each session was 15 minutes long, and we collected data from three sessions per subject, for a total of 45 minutes. RGB images (for posture estimation) were captured at 30 fps from three directions in front of, behind, and to the right of the subject using a Logitech C920n. Motion capture data (for annotation) was collected at 100 fps using NaturalPoint's OptiTrack Flex 13. The motion capture marker positions were set as shown in Fig.2 so that the degree of curvature of the spine and the tilt of the pelvis could be seen.

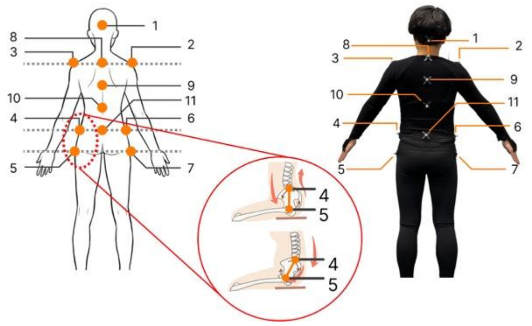


Fig.2 Marker positions for motion capture.

2.2. *Annotation*

Due to slow postural changes during desk work, we down-sampled the collected image data to 15 fps for posture estimation. Thus, the number of images for posture estimation was approximately 1,215,000 (= 30 subjects × 45 minutes × 60 seconds/minute × 15 fps). For supervised learning, it is necessary to assign true posture labels to each image data. However, since no medical indices exist to evaluate posture quality, a quantitative annotation based on rules is challenging. Additionally, manually annotating all data is time-consuming and labor-intensive, making it impractical. Hence, this study manually annotated part of the data and trained an estimation model for annotation, using the trained model to perform automated annotation on the remaining data.

2.2.1. Manual Annotation

Sampled image data every 25 seconds was used for manual annotation. Due to missed frames in some motion capture data, 2,955 images were extracted. Using images captured from the side view, where the spinal and pelvic tilt is most visually apparent, annotators labeled the data in three stages: good, slightly poor, and poor. Five annotators labeled each data piece to account for potential subjective judgment variability, and an integrated posture label was determined from these labels. Specifically, scores of good, slightly poor, and poor posture, denoted as S, were set to 1, 0, and -1, respectively. The average score $\bar{S}$ from five annotators determined the posture labels L as in Equation 1. In addition, the variance, σ^2, was calculated, and data with a variance value of less than 0.5 was treated as valid annotation data.

$$L = \begin{cases} Good\ Posture\ if\ \frac{1}{3} \leq \bar{S} \leq 1, \\ Slightly\ Poor\ Posture\ if\ -\frac{1}{3} < \bar{S} < \frac{1}{3}, \\ Poor\ Posture\ if\ -1 \leq \bar{S} \leq -\frac{1}{3}. \end{cases} \quad (1)$$

2.2.2. Learning Estimation Model and Automatic Annotation

Motion capture data is measured with less than 1 mm error for each marker's 3D coordinates, containing precise information on spinal and pelvic tilt, thus the input for the estimation model comprised 33 dimensions of features, being 3D coordinates (x, y, z) of each marker point (11 points). Support Vector Machine (SVM) [4] was employed for the 3-class classification of good posture, slightly poor posture, and poor posture. RBF kernel was used, and hyperparameters C and γ were optimized via grid search, resulting in $C = 20$ and $\gamma = 0.1$.

Performance evaluation results upon dividing manually annotated data into training and test data are shown in Table1. Table1 indicates that although the estimation accuracy of slightly poor posture is somewhat inferior compared to other postures, it is overall accurately estimated. Using the trained SVM model, automated annotation was performed on the remaining data. Class-wise data distribution is shown in Table2. Through these procedures, all image data were labeled as truth data, allowing them to be utilized for learning and evaluating posture estimation models.

Table 1: Annotation Estimation Performance.

	Precision	Recall	F1-score
Good Posture	0.98	0.94	0.96
Slightly Poor Posture	0.86	0.94	0.89
Poor Posture	0.96	0.91	0.94

Table 2: Breakdown of Annotation Labels.

	Manual Annotation	Automatic Annotation
Good Posture	1,181	487,579
Slightly Poor Posture	1,039	388,479
Poor Posture	704	355,884

3. Posture Estimation Method and Evaluation Experiment

Utilizing input data such as RGB images from the side, silhouette images, and Yolo's posture estimation features, we conducted a 3-class classification of seated postures. Table3 summarizes the experiments.

Each model utilized the Cross Entropy Loss as the loss function and Adam for optimization. The epoch number was set to 1000, and training was terminated through EarlyStopping if the F1 score decreased for 10 consecutive epochs.

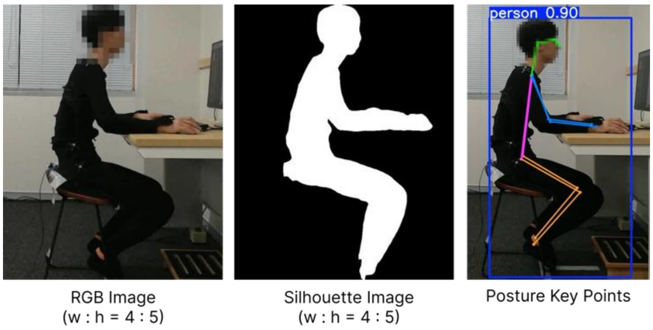


Fig.3 Image Preprocessing.
【Silhouette Image】: Extracted using DeepLabV3 with ratio adjustments to 5:4 height-to-width based on the subject's height.
【RGB Image】: Similarly extracted subject regions using DeepLabV3, adjusting the width to a 5:4 aspect ratio.
【Posture Key Points】: Acquired 17 key points' 2D coordinates with YOLO's pose estimation results.

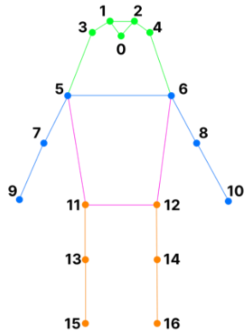


Fig.4 YOLO Pose Estimation Key Points.
0: Nose, 1: Left Eye, 2: Right Eye, 3: Left Ear, 4: Right Ear 5: Left Shoulder, 6: Right Shoulder, 7: Left Elbow, 8: Right Elbow 9: Left Wrist, 10: Right Wrist, 11: Left Hip, 12: Right Hip 13: Left Knee, 14: Right Knee, 15: Left Ankle, 16: Right Ankle

Table 3: Experiments

	Method	Input Data
Experiment 1	ResNet50	RGB Images
Experiment 2	ResNet50	Silhouette Images
Experiment 3	GCN	YOLO Pose Estimation Results

3.1. *RGB Images*

Using RGB images as input, ResNet50 [5] was utilized for 3-class classification of good posture, slightly bad posture, and bad posture. The RGB image of the subject, segmented using DeepLabV3 [3] as shown in Fig.3, was used as input. For model training, the hyperparameters were as follows: learning rate of 0.001, batch size of 1024, and a weight decay parameter of 0.01. Adam's parameters $\beta 1$, $\beta 2$, and eps were set to 0.95, 0.995, and 1e-06, respectively. For the final layer of ResNet50, the existing fully connected layers were replaced to correspond to the 3-class classification. During training, the cross-entropy loss was used, and the softmax function was used to classify the outputs into 3 classes.

3.2. *Silhouette Images*

Silhouette images were used as input with ResNet50 [5] for 3-class classification of good posture, slightly bad posture, and bad posture. Silhouette images were generated using DeepLabV3 [3] as outlined in Fig.3. The training hyperparameters were set to the same values as the RGB images. The input layer for ResNet50 was modified to have a channel count of 1 to accommodate silhouette images, and the final layer was replaced for a 3-class classification. During training, cross-entropy loss and the softmax function were employed to distinguish the outputs into 3 classes.

3.3. *Posture Key Points*

Using the posture feature points from YOLO's pose estimation as input, GCN [8] was employed for 3-class classification of good posture, slightly bad posture, and bad posture. The results of YOLO's pose estimation consisted of 17 key point 2D coordinates, as illustrated in Fig.4. The hyperparameters for training were configured as follows: learning rate of 0.01, batch size of 512, and weight decay parameter of 0.0. Adam's parameters $\beta 1$, $\beta 2$, and eps were set to 0.9, 0.999, and 1e-08, respectively. The input layer of GCN was set to have a channel count of 17 based on YOLO's pose estimation results. A fully connected layer for 3-class classification was added at the end, and cross-entropy loss was employed for training, classifying the output into 3 classes through a softmax function.

3.4. *Evaluation Method*

For performance evaluation, subjects were divided into six groups of five subjects each, performing a 6-fold cross-validation. Each group was set as test data once, with another group set as validation data, and the remaining four as training data, continuing the cycle until all groups served as test data once. While automated annotation data was used for training and validation, only manually annotated data was used in testing to account for potential estimation errors. Random undersampling ensured a balanced class distribution for training data across groups. Models were evaluated using those from epochs exhibiting the highest validation data performance, with accuracy and macro-F1 score as evaluation metrics.

3.5. *Results and Discussion*

The results of the sitting posture estimation are shown in Table4. When using RGB images as input, when using silhouette images as input, and when using the results of YOLO's pose estimation as input, the average accuracy was 0.729, 0.696, and 0.528, respectively, and the average F1 score was 0.591, 0.589, and 0.459, respectively. The highest Accuracy and F1 score were achieved when the input was an RGB image. The reason for the higher Accuracy and F1 score than for the silhouette images is thought to be that they have the same background and clothing, and there is noise due to false positives in the semantic segmentation of the silhouette images. On the other hand, when the results of Yolo's pose estimation were used as the input, the accuracy and F1 score were the lowest. This is thought to be because, as shown in Fig.5, the results of YOLO's pose estimation do not express the posture features necessary for estimating sitting posture, such as the curvature of the back. However, the silhouette images contain noise due to false positives in the semantic segmentation, and the RGB images contain noise due to background, color information, clothing, etc. Therefore, it is thought that it would be ideal to use data with less noise and more detailed posture features, rather than sparse posture feature points like YOLO.

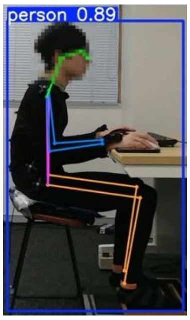


Fig.5 Example of Poor Representation of Posture Features by YOLO Pose Estimation Results.

Table 4: Seated Posture Estimation Results

	Ave-Acc	Ave F1
RGB Image	0.729	0.591
Silhouette Image	0.696	0.589
Posture Key Points	0.528	0.459

4. Conclusion

In this study, we constructed a dataset for estimating natural sitting postures using monocular RGB images. Using the dataset we created, we performed 3-class classification of sitting postures using a deep learning model with RGB images and silhouette images, and low-dimensional posture feature points extracted using Yolo as inputs, and compared the accuracy of each model according to the differences in the characteristics of the input data. As a result, the model using silhouette images showed the highest accuracy, while the model using posture feature points showed the lowest accuracy. However, since RGB images and silhouette images contain information that is noise and not necessary for posture estimation, it is thought that using posture feature points with appropriate features is effective. Therefore, in the future, the challenge will be to propose a method for extracting posture feature points that are optimal for sitting posture. It is thought that posture feature points that are optimal for sitting posture can be extracted using 3D human body models and 3D human skeletal models.

References

1. Bauman, A., Ainsworth, B. E., Sallis, J. F., Hagstr¨omer, M., Craig, C. L., Bull, F. C., Pratt, M., Venugopal, K., Chau, J. and Sj¨ostr¨om, M.: The Descriptive Epidemiology of Sitting: A 20-Country Comparison Using the International Physical Activity Questionnaire (IPAQ), American Journal of Preventive Medicine, Vol. 41, No. 2, pp. 228–235 (2011).
2. Chen, K.: Sitting Posture Recognition Based on OpenPose, IOP Conference Series: Materials Science and Engineering, Vol. 677, No. 3, p. 032057 (2019).
3. Chen, L.-C., Papandreou, G., Schroff, F. and Adam, H.: Rethinking Atrous Convolution for Semantic Image Segmentation (2017).
4. Cortes, C. and Vapnik, V.: Support-vector networks, Machine Learning, Vol. 20, No. 3, pp. 273–297 (1995).
5. He, K., Zhang, X., Ren, S. and Sun, J.: Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016).
6. Journal Editor, P., Iqbal, M. J., Megha, Vyas, N., Kansal, H. and Kulshreshth, D.: A study of Smart Chair for Monitoring of Sitting Behaviour, PRATI- BODH, No. RACON 2023 (2023).
7. Kapoor, R., Jaiswal, A. and Makedon, F.: Light-Weight Seated Posture Guidance System with Machine Learning and Computer Vision, Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments, PETRA '22, New York, NY, USA, Association for Computing Machinery, p. 595–600 (2022).
8. Kipf, T. N. and Welling, M.: Semi-Supervised Classification with Graph Convolutional Networks, International Conference on Learning Representations (2017).
9. Li, L., Yang, G., Li, Y., Zhu, D. and He, L.: Abnor- mal sitting posture recognition based on multi-scale spatiotemporal features of skeleton graph, Engineering Applications of Artificial Intelligence, Vol. 123, p. 106374 (2023).
10. Otoda, Y., Mizumoto, T., Arakawa, Y., Nakajima, C., Kohana, M., Uenishi, M. and Yasumoto, K.: Census: Continuous posture sensing chair for office workers, 2018 IEEE International Conference on Consumer Electron- ics (ICCE), IEEE, pp. 1–2 (2018).
11. Ultralytics inc.: Ultralytics YOLOv8, https://docs.ultralytics.com/ja/models/yolov8. (Accessed on 12/12/2024).

Authors Introduction

Mr. Hitoshi Shimomae

He received his Bachelor's degree in Computer Science and Systems Engineering in 2024 from the Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology in Japan. He is currently a master student at Kyushu Institute of Technology, Japan

Mr. Tsubasa Esumi

He received his Bachelor's degree in Computer Science and Systems Engineering in 2023 from the Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology in Japan. He is currently a master student at Kyushu Institute of Technology, Japan

Dr. Noriko Takemura

She received B.S., M.S., and Ph.D. degrees in engineering from Osaka University, Japan, in 2006, 2007, and 2010, respectively. She is currently an associate professor at Graduate School of Computer Science and Systems Engineering , Kyushu Institute of Tech-nology. Her current research interests include human-centric intelligent systems

Identification of lung nodules based on combining multi-slice CT images and clinical information

Yuto Nishitaki
Kyushu Institute of Technology, 1-1 Sensui-cho, Tobata-ku, Kitakyushu, 804-8550, Japan

Tohru Kamiya
Kyushu Institute of Technology, 1-1 Sensui-cho, Tobata-ku, Kitakyushu, 804-8550, Japan

Shoji Kido
Osaka University, 2-2 Yamadaoka, Suita, Osaka, 565-0871, Japan

Email: nishitaki.yuuto699@mail.kyutech.jp

Abstract

Although chest CT scans are an effective means of diagnosing lung cancer, there are still problems such as the heavy burden on physicians. To solve this problem, computer-aided diagnosis (CAD) systems are being introduced. Conventional CAD systems are based on a method that uses only image information. In this study, we propose a method for identifying nodular shadows that integrates a composite image created from multi-slice CT images and clinical information such as the patient's age, sex, and medical history in the medical record. The proposed method extracts features from the multi-section CT and clinical information, respectively, integrates the features, and then performs binary classification of nodules and vessels using a classifier. The proposed method achieved a very high accuracy of Accuracy=0.983, TPR=0.987, and FPR=0.018.

Keywords: Computer Aided Diagnosis, Deep Learning, Clinical Information, Multimodal, Multi-slice CT

1. Introduction

Malignant neoplasms are the leading cause of death worldwide, and lung cancer is one of the most serious of these diseases; according to a 2022 report by the International Agency for Research on Cancer (IARC), more than 2.4 million new cases of lung cancer will be reported worldwide each year, making it the leading cause of cancer death [1]. In Japan, malignant neoplasms have been the leading cause of death since 1981, and by 2022, lung cancer will be the leading cause of death in men, the second leading cause in women, and the leading cause in both men and women combined [2]. Because lung cancer progresses rapidly, and survival decreases significantly as the disease progresses, early detection and early treatment are important issues. When lung cancer is suspected, a chest X-ray is performed first, and if abnormalities are found, a chest CT scan is widely used as a more precise examination. However, the large number of CT images generated in a single examination places a heavy burden on the physician reading the images. In addition, the possibility of variations in diagnostic accuracy and missed cases due to differences in the skills and experience of the reading physicians has also been pointed out. To address these issues, computer-aided diagnosis (CAD) systems are needed to reduce the burden on the reading physician and improve diagnostic accuracy. Conventional CAD systems are based on a method that uses only image information [3], [4]. However, in actual clinical practice, diagnosis is made by considering background information such as the patient's age, gender, and medical history. For this reason, CAD systems that integrate image and clinical information have been studied in recent years. In this paper, we propose a method for identifying pulmonary nodular shadows by combining multi-slice CT images and clinical information from medical records.

2. Methodology

The overall process is shown in Figure 1. First, image features and clinical information are extracted from the CT images obtained from the examination and the clinical information of patient, respectively. Next, these features are integrated and finally classified into two classes, nodules and vessels, using a classifier.

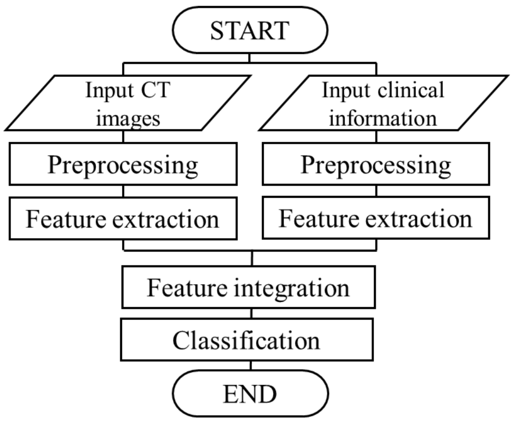


Fig.1. Flow of the proposed method.

2.1. *Image feature*

In this paper, to obtain 3D features of the target with low computational cost, a synthetic image using three images (axial, sagittal, and coronal cross-sections) is fed into an image feature extraction model. For feature extraction, CoAtNet [5], which combines CNN to capture local features and Self-Attention to capture global features, is used as the base model, and an improved CoAtNet that introduces a Convolutional Block Attention Module (CBAM) [6] is used. The CBAM was introduced to estimate attention maps along two different dimensions, channel and spatial, with the goal of highlighting important features and suppressing unnecessary information.

2.2. *Clinical information feature*

In this paper, gender, age, and medical history are used as medical record information. The six types of medical history used are colorectal cancer, stomach cancer, other cancers, tuberculosis, pneumonia, and other lung-related diseases. As a preprocessing step, gender is converted to one-hot coding with 1 for males and 0 for females. Similarly, history is converted to one-hot coding with 1 if the disease is present and 0 if it is absent. Age is standardized to reduce the effect of outliers and to equalize the scales of the explanatory variables. After these preprocessing steps, a multi-layer perceptron consisting of three layers is used to extract clinical information features. Details of the model used are shown in Table 1.

2.3. *Feature Integration*

The acquired 512-dimensional image features and 128-dimensional clinical information features are horizontally combined by concatenating layers to form a 640-dimensional feature vector. Self-Attention with Scaled Dot-Product Attention [7] is then applied to the combined features to highlight important relationships between different modalities. This feature vector is fed into a four-layer multilayer perceptron, which performs a two-class classification of nodules and blood vessels. A summary of the model used is shown in Table 2.

Table 1. Clinical information extraction model

Layer	Input	Output	Activation function
Input	8		-
Dropout.1	-	-	-
Linear.1	8	128	ReLU
Dropout.2	-	-	-
Linear.2	128	256	ReLU
Dropout.3	-	-	-
Linear.3	256	128	softmax
Output		128	-

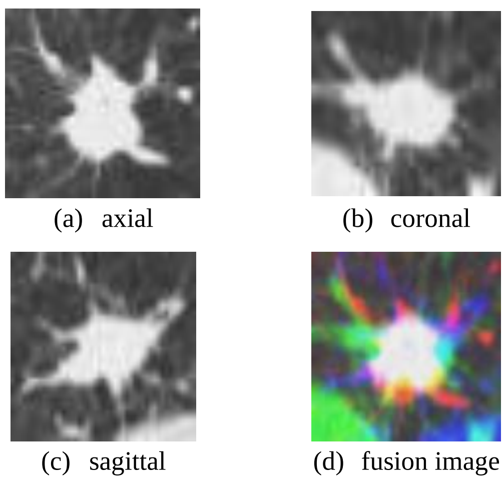
Table 2. Classification model

Layer	Input	Output	Activation function
Input	640		-
Dropout.1	-	-	-
Linear.1	640	1024	ReLU
Dropout.2	-	-	-
Linear.2	1024	512	ReLU
Dropout.3	-	-	-
Linear.3	512	256	ReLU
Dropout.4 Linear.4	256	2	softmax
Output		2	-

3. Experiment

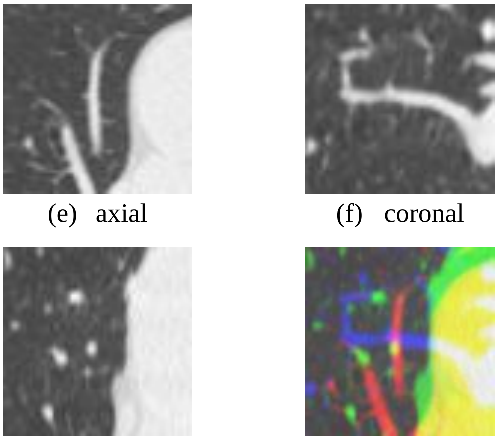
3.1. *Dataset*

In this paper, we use three images from chest CT images: axial, sagittal, and coronal cross-sections to obtain 3D features of the object while minimizing computational complexity. CT images with a nodule of 20 mm or less were used as input images. For the axial section, the slice plane with the largest nodule was selected from a set of CT image data obtained from a chest CT examination. The area of the nodule was manually segmented and modified under the guidance of a physician. In addition, because the region of interest was small, it was cropped from 512 × 512 pixels to 112 × 112 pixels and resized as needed. For the coronal and sagittal sections, the center coordinates of the region of interest were obtained from the axial section, and each image was cropped to 112 × 112 pixels based on these coordinates. The same procedure was used to generate images of blood vessels. The nodule images of 157 nodules from 73 cases were used as positive images, and the vessel images of 325 vessels from 73 cases were used as negative images. The images are shown in Figure 2 and Figure 3.



(a) axial (b) coronal (c) sagittal (d) fusion image

Fig.2. Dataset (lung nodule)


(e) axial (f) coronal

(g) sagittal (h) fusion image

Fig.3. Dataset (blood vessel)

Table 3. Experimental results (image only)

Model (Input image)	Accuracy	TPR	FPR
Improved ResNet [8] (Axial image)	0.938	0.917	0.052
CoAtNet (Axial image)	0.935	0.923	0.061
Proposed model (Axial image)	0.950	0.949	0.049
Proposed model (Fusion image)	0.977	0.980	0.025

Table 4. Experimental results (image + clinical information)

Method	Accuracy	TPR	FPR
Method1	0.968	0.968	0.032
Proposed method	0.983	0.987	0.018

3.2. *Data augmentation*

Deep learning is prone to overlearning when the amount of data is small. Therefore, it is necessary to improve the generalization performance by data expansion. In this paper, the training data was rotated from 15 to 345 degrees in 15-degree increments and flipped horizontally and vertically to increase the data volume 23-fold.

3.3. *Evaluation function*

The metrics used are Accuracy, which is the proportion of correctly classified images in all data, True Positive Rate (TPR), which is the proportion of images correctly identified as positive when the correct answer is positive, and False Positive Rate (FPR), which is the proportion of images incorrectly identified as positive when the correct answer is negative.

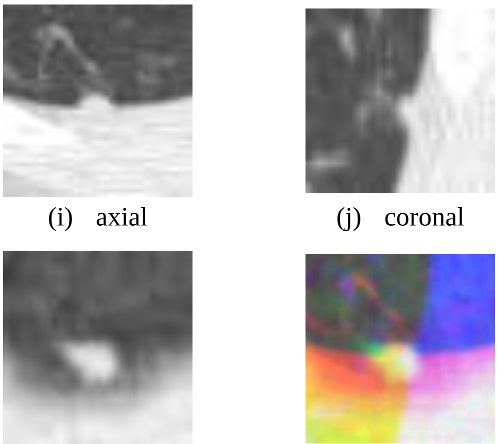
4. Results and Discussion

Table 3 shows the classification results using only images as input. The image feature extraction models used were Improved ResNet, which was proposed in a previous study [8], and CoAtNet, which was the base model before improvement, and the proposed model that introduced CBAM into CoAtNet. Improved ResNet is a ResNet34 [9] with a residual block just before, a convolutional autoencoder is added to ResNet34, and an SE block [10] is introduced within the residual block. The classification results using image and clinical information are shown in Table 4. All the proposed models were used as image feature extraction models. Method 1 is a method in which image features and clinical information are integrated and then directly input into a classification model. In the proposed method, the features transformed by Self-Attention are input into the classification model after feature integration.

Table 3 shows that the proposed method achieves higher accuracy than the other methods in the image-only experiment. This may be due to the introduction of CBAM, which calculates the importance in both the channel and spatial directions for the feature map, thereby highlighting important features. Another contributor to the improved accuracy may be the use of composite images generated from cross-sectional images in three directions, which allows accurate detection of different shape features: spherical or elliptical pulmonary nodules and vessels with elongated tubular or branched structures, while reducing computational complexity.

Table 4 shows that when additional information about the patient was learned in combination with image information, the proposed method using Self-Attention achieved higher accuracy than the method using image information alone. This result can be attributed to the fact that the introduction of Self-Attention was able to dynamically highlight correlations between important features and also effectively integrate complementary information between different modalities.

Examples of nodules misclassified by the proposed method are shown in Figure 4. In this case, the nodule was misclassified because the area occupied by lung tissue other than the nodule was large in the image, and it was difficult to accurately capture the shape features of the target. In addition, the limited number of images of nodules invaded by other lung tissue in the dataset used may have contributed to the misclassification. Future research needs to collect and expand image data showing different types of tissue erosion.



(i) axial (j) coronal (k) sagittal (l) fusion image

Fig.4. Misclassified image

5. Conclusion

In this paper, we proposed a method to identify pulmonary nodular shadows by combining multi-slice CT images and clinical information from medical records. The proposed method achieved a very high accuracy of Accuracy=0.983, TPR=0.987, and FPR=0.018, confirming the effectiveness of the proposed method.

In future studies, image data showing different types of tissue erosion will be collected and tested again. In addition, since image cropping is currently done manually, we hope to automate image generation using temporal subtraction imaging [11] or similar methods in the future.

References

1. International Agency for Research on Cancer, Cancer Today: Data Visualization Tools for Exploring the Global Cancer Burden in 2022, https://gco.iarc.fr/today/en.
2. Foundation for Promotion of Cancer Research, Cancer Statistics 2024, 2024.
3. Tamai, et al., "Detection of Nodular Shadows from Chest CT Images Using Deep Learning," Proceedings of the Biomedical Fuzzy Systems Society Annual Conference, vol. 33, pp. 26–29, 2020.
4. Li, et al., "Development of a Classification Model for Detecting Nodular Shadows in Pulmonary CT Images," Proceedings of the 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 1P4-J-10-01, 2019.
5. Z. Dai et al., "CoAtNet: Marrying Convolution and Attention for All Data Sizes", Advances in Neural Information Processing Systems 34, 2020.
6. S. Woo et al., "CBAM: Convolutional Block Attention Module", European Conference on Computer Vision, pp.3-19, 2018.
7. A. Vaswani et al., "Attention Is All You Need", Advances in Neural Information Processing System, vol. 30, pp5998-6008, 2017.
8. N. Honda et al., "Identification of abnormal tissue from CT images using improved ResNet34", International Conference on Control, Automation and Systems, pp.532-536, 2022.
9. K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.770-778, 2016.
10. J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, "Squeeze-and-excitation networks", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.7132-7141, 2018.
11. A. Kano et al., "Digital image subtraction of temporally sequential chest images for detection of interval change", Medical Physics, Vol.21, No.3, pp.453-461, 1994.

Authors Introduction

Mr. Yuto Nishitaki

He received his Bachelor's degree in Engineering in 2023 from the Faculty of Engineering, Kyushu Institute of technology in Japan. He is currently a master student in Kyushu Institute of Technology, Japan.

Tohru Kamiya, Prof. Ph.D

He received his B.A. degree in Electrical Engineering from Kyushu Institute of Technology in 1994, the Master and Ph.D. degree from Kyushu Institute of Technology in 1996 and 2001, respectively. He is a professor at the Department of Mechanical and Control Engineering at Kyushu Institute of Technology. His research interests are focused on image processing and medical application of image analysis.

Shoji Kido M.D. Ph.D.

He received his M.D. degree from Osaka University in 1988. He received his Ph.D. degrees in Medicine and Information Science from Osaka University in 1992 and 1999, respectively. He is a guest professor of Osaka University Institute for Radiation Science and Osaka University Graduate School of Medicine. His research interests are focused on the use of artificial intelligence in radiology.

Graphical User Interface Design for Mobile Commerce Site for Women Seller in Rural Area

Kho Irene Lee Sa
Faculty of Business and Management, UCSI University Sarawak Campus
Email: irenekho@ucsiuniversity.edu.my

Shayla Islam
Institute of Computer Science and Innovation, UCSI University, Kuala Lumpur, Malaysia

Abdul Samad Shibghatullah
College of Computing & Informatics (CCI), University Tenaga Nasional, Kajang, Selangor, Malaysia

Abstract

Graphical User Interface (GUI) is a medium that allows users to communicate with electronic devices. Good GUI design is important to allow users to learn to use a system in the shortest time and effectively operate the system. Especially in rural areas with limited communication resources, a well-designed GUI for m-commerce sites that integrates women seller's personality types becomes crucial. Mobile-commerce (m-commerce) site should aid women in their efforts to sell their products directly to customers. Unfortunately, the interface design of m-commerce sites caters only to IT literate users with fast internet connections. Existing m-commerce sites' interface designs, are compact to place as much information as possible. This may lead to an interface design that looks busy and messy unless women are given proper guidance by IT experts. A research review using a GUI as a standard to allow women sellers in rural areas to promote and sell their product through an m-commerce site.

Keywords: m-commerce site, women seller, graphical user interface

1. Introduction

According to job statistics for 2021 for Sarawak Malaysia, there are 50.8 thousand unemployed citizens. That's mean about 30% of the citizens of Sarawak are not participate in labor force [1]. It is believed that housewives are the majority group of people who did not join the work force.

Culturally, housewives in rural areas are not allowed to join the work force by their family members. On top of that, with their low wages, women in the rural area will not be able to afford to pay for their children's daycare fees in their absences when they are working [2]. Furthermore, good job opportunities will only be available in cities. Transportation fees and time spent on the road may not be ideal for women in rural areas. As a result, female unemployed individuals should be encouraged to consider becoming m-commerce entrepreneur to work from home.

According to [3], village-based activities were not always sufficient to support household expenses, rural households have become increasingly dependent on various sources of income. These findings also highlight the challenging nature of rural livelihood, which is why diversification is already a survival mechanism for rural areas. Hence, rural households should consider venturing into entrepreneurship. This is crucial, especially for housewives who are unemployed but want to contribute to supporting household expenses.

To encourage women in rural areas to pursue entrepreneurship, a substantial of assistance is required. Initially, women in rural areas were required to possess certain skills in order to produce creative and traditional goods that were difficult to duplicate [2]. For instance, traditional hand craft beaded headpiece, mat-making, basketry and woodcarving. Whereas, traditional food product includes dabai cracker, tebaloi, gula mitai, pepper sweets, sago balls, kuih jala and salai lumek. Then, rural women seller must trade the traditional product in order to earn a living [4]. However, infrastructure is lacking in rural areas. The produced goods will not be able to send out due to the physical accessibility to market. The public may not be able to view the product or even be aware of its existence.

Explore in a digital environment does not only allow women seller in rural environments to have direct access to the customer without intermediaries, they can also sell the product at a higher price to users worldwide [5]. With the internet connection in place, women seller in the rural area are able to use mobile devices such as smart phone or tablet as a tool to connect themselves with the world without going through intermediaries. Considering M-commerce will be viable solution to women seller in the rural area due to devices such as desktop and laptop required complicated setup.

Hence, this research addresses the following research questions as shown below using systematic review.

RQ1: What are the user interface women seller in rural areas are comfortable to have in m-commerce site?
RQ2: What personality types can influence women seller using m-commerce site?

Several criteria must be satisfied, so that the literature will be considered for the review:
- Studies are indexed and published as journal articles, book chapters and proceedings.
- Studies are focused on 'up to date information' within the m-commerce site.
- Studies are easily accessible by the researcher through public or open access.

Table 1 below is a summary table of terms use in this research study. This table will give an inside about the terms use in this research.

Table 1 Summary of Terms

Terms	Literature
Mobile Commerce site	**Definition** M-commerce, also known as mobile commerce, is the trade of products via the Internet while utilizing a mobile device. M-commerce is now accessible from mobile devices that support online services partially due to the development of applications and services in the current world. [6] **Different Types and Categories of M-commerce site** There are two common types operating systems that support M-commerce site. That is IOS and Android. The popular M-commerce site use in Malaysia includes Lazada, Shopee, Mudah, GoShop, and Lelong.com. M-commerce site is crucial for entrepreneurs to explore online business. It cut short various heavy investments in technology for a beginner in online business.
Human Computer Interaction (HCI)	**Definition** The term Human-Computer Interaction/Interfacing (HCI), often referred to as Man-Machine Interaction or Man-Machine Interfacing, was immediately associated with the development of the computer or, more generally, the machine itself. HCI is an investigation of how people and computers communicate with one another. Researchers in HCI examine how people use computers and create novel solutions that enable innovative computer use. A device that permits interaction between a human and a computer is known as an HCI. For the many methods that people and computers communicate to work, an interface between them is necessary. [7] **Graphical User Interface** A graphical user interface (GUI) is a computer program that allows people to connect with computers using symbols, visual metaphors, and pointing devices. The relatively easy GUI interface has replaced the complicated and challenging textual interfaces in earlier computing, making computer use not only simpler to understand but also more enjoyable and natural. The GUI has become a standard computer interface, and its individual components are now recognizable cultural possessions. [8] **User Interface in M-commerce Site** Mobile devices have replaced desktop computers and laptops due to their mobility and ability to be carried from one place to another. Several sizes of mobile devices have developed to cater to different screen needs, such as mobile phones, palm devices, and tablets. M-commerce is an avenue that connects business leaders and their customers for multiple commerce operations via mobile devices. However, there are various challenges in m-commerce that make it harder for customers to complete smoother operations or transactions using their mobile devices, such as restrained mobile memory, user interface simplicity, size of display, and small input method or keypad.

2. Methodology

The process of systematic review includes identifying problems; from there, generating RQ; followed by a literature search strategy, then literature selection by choosing the most relevant literature. Next is results and discussion, conclusion and recommendation will be the last part. (Fig. 1)

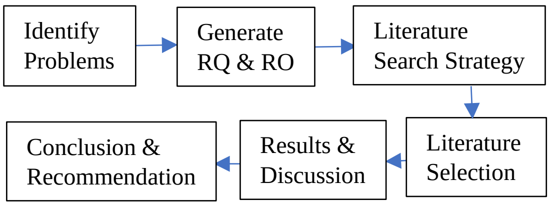


Fig. 1 literature search strategy process

In the literature search strategy process, researchers search for relevant literature related to the topic through electronic journals using related keywords. Common electronic journals used for literature search purposes include Google Scholar, ProQuest, Science Direct, Springer, and Emerald Insight. Most of the time, these common electronic journals are chosen because they are free, can be found in university libraries, have formal ways to access them, are up-to-date, and are easy for researchers to find.

Literature selection is conducted as shown in Fig. 2. With 34,000,000 literature results shown in Google search, researchers only focus on the first 200 search results through preliminary screening. 50 papers are selected based on the title of the research and relevance to the studies. Then 13 papers are selected after accessing the abstracts. Finally, five pieces of literature are selected after reading the full papers.

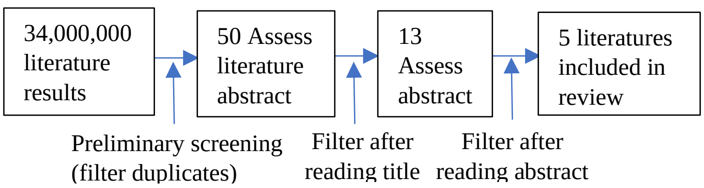


Fig. 2 Literature selection process

After going through all the literature reading, this paper will summaries the finding based on women seller in rural area intention to m-commerce site to sell their product. The outcome will be summaries based on personality types of women sell in rural area and the GUI requirement comfortable for them to start using the m-commerce site as shown in Fig. 3.

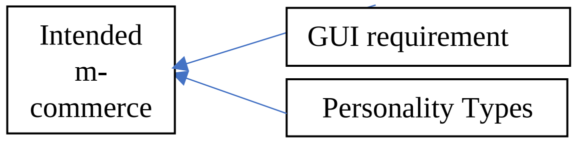


Fig. 3 Intended m-commerce site requirement

3. Result & Discussion

After a systematic search strategy, the outcome and findings will be presented and discussed. The selected five journal articles are the closest to the research topic. All the journal articles do not have a good match with the research topic. Some of the journal's topics are mainly discussed in the education area rather than in m-commerce selling products. Not only that, the journals are found to study women consumers rather than women seller selling products. Nevertheless, the journals discuss the topic of user interface in m-commerce, which fits well for carrying out a systematic review. All five selected articles range from the years 2017 to 2023.

RQ1: What are the user interface women seller in rural areas are comfortable to have in m-commerce site?

Research on m-commerce site adopted by women seller in rural area in related literature not easily identified. An area of studies related to this topic will be adopted in this study as shown in Table 2.

Table 2 Summaries of literature on m-commerce site adopted by women seller in rural area

Author	Research Title	Existing Research	Weakness of M-Commerce System	Existing Problems for Women
[9]	The roles of user interface design and uncertainty avoidance in B2C ecommerce success: Using evidence from three national cultures	This research incorporates user interface design (UID), visual aspects (VA), picture aspects (PA), and third-party seals (TPSs) as predictors of system quality use in e-commerce.	This research conducted on consumer from Kuwait, Poland and Latvia which has different culture from Malaysia. Their perception may not fit to Malaysia contact.	Not mention about women seller Respondent of the study are university students where they are digital natives Using e-commerce instate of m-commerce

			PA and formatting affect system quality of e-commerce, while TPS not affect system quality. This may not applicable to m-commerce.	Mention customer point of view instate of women merchant
[10]	Determinants for M-Commerce Adoption in Malaysia Small to Medium Enterprise (SME): A Conceptual Framework	This research state the importance of SME adopts m-commerce. There is an increase trend of m-commerce in Malaysia	SME owners could not accept modern and sophisticated technology to their company operations. SME not able to adapt globalization pressure SME feels m-commerce is not important. SME do not adopt mobile devices for monetary transaction.	Not mention about women to use m-commerce apps Majority of SME adoption rate is low

Continuation from Table 2

Author	Research Title	Existing Research	Weakness of M-Commerce System	Existing Problems for Women
[11]	Interface designs with Myers-Briggs Type Indicator (MBTI) personality types	make use of human personality types using MBTI and human computer interaction and user interface designs in electronic learning	Discuss e-learning of student in higher education would increase efficiency when user interface design customizes the individual personality of a student E-learning prefer more text driven, more graphics and mind mapping.	Not mention about women to use m-commerce apps women seller's condition is different from e-learning in higher education. It is hard to compare with the two situations.
[12]	Development of measurement instrument for visual qualities of graphical user interface elements (VISQUAL): a test in the context of mobile game icons.	a survey conducts towards 569 participants to measure their aesthetics with graphical user interface element to creative a successful graphical asset. The researcher divides GUI into 5 dimensions of perceived VISQUAL: Excellence / Inferiority, Graciousness / Harshness,	focus on discuss the singular interface elements (VISQUAL) that can be used in multiple ways in several contexts related to visual human-computer interaction, interfaces and their adaption only the survey is based on 68 game app icons	Not mention about women to use m-commerce apps

		Idleness / Liveliness, Normalness / Bizarreness and Complexity / Simplicity. a measurement instrument for perceptions of visual qualities of graphical user interfaces and/or singular interface elements VISQUAL that can be used in multiple ways in several contexts related to visual human-computer interaction, interfaces and their adaption		

Continuation from Table 2

Author	Research Title	Existing Research	Weakness of M-Commerce System	Existing Problems for Women
[6]	M-Commerce in Malaysia - Opportunities and Challenges	Journal review paper that states the current scenario of m-commerce apps in various industrial in Malaysia.	Existing m-commerce apps are complicated. Insecure wireless infrastructure makes it easy for intruder to get in.	Not mention about women to use m-commerce apps
		The journal also states the challenges and opportunities using mobile apps.	Language barrier is another weakness of apps available. Many Malaysians prefer to use their indigenous language. Low internet connectivity (2G, 3G) not able to support m-commerce apps Make user still prefer use mobile apps as a medium to find out more about products, but purchase offline.	
[13]	M-Commerce-Interface Design: A Review of Literature	This journal discusses several interface design principles that m-commerce apps must follow in order to be a good app. m-commerce interface design element include context, content, community, customization, communication, connection and commerce elements.	M-commerce apps has limited mobile memory, simplicity of user interface, screen size and small input method and keypad	This paper discusses from the customer point of view using m-commerce apps to purchase products This paper didn't discuss from the merchant point of view

Women sellers in rural areas are the minority groups that researchers tend to forget to investigate their needs. Research studies of m-commerce site for women sellers to sell their products online need more attention. With existing studies by Hassna [9], shoppers appreciate good picture quality and formatting quality. Picture quality refers to real pictures of products with different sizes and clear pictures. As for formatting quality, font sizes to differentiate title and content, color to differentiate the information and background displayed is an important elements of m-commerce site design. Digital natives' shopper preferences may not represent the women seller's viewpoints. However, it can serve as a good guide as a basic user's requirement.

According to Yahaya [10], SME Malaysia will adopt m-commerce if it is flexible, effective, and suitable. Minimum maintenance, user-friendliness, and ease of monitoring will be considered to be adopted using m-commerce. As for individual women sellers, simplicity is the major concern on top of SME concerns. Similarly, SME owners and women sellers share the common viewpoint that they are not ready for modern and complicated technology yet. This may due to their cultural background.

Learners at higher education institutions are not comparable with women sellers in rural areas. Undeniably, a common viewpoint of learner and women seller are gaining new knowledge. In order for them to understand a new knowledge, simple and easy ways to learn are important to them. According to Subaramaniam [11], different personality types will absorb knowledge differently. Hence, women sellers are believed to adopt m-commerce if the user interface design suits their personality types. This is tally with Jylha [12] where his study focuses on VISQUAL for mobile game users. His study enforces using a visual human-computer interface where women sellers are believed to be able to adopt better compared to text explanations. So, a simple GUI based on women sellers's personality types that comes with a visual GUI will be helpful for them to adopt m-commerce.

Authors Ahmad [13] and Al-Abiary [6] have written the review paper mentioning m-commerce interface design and condition. Al-Abiary indicates that m-commerce apps are complicated due to many details. Malaysians believe m-commerce is insecure; low internet connectivity, added to the language barrier, makes them prefer to purchase offline. With those discouragements, the women sellers may refuse to adopt m-commerce to sell their products. The majority of m-commerce sites used in Malaysia are adopted from foreign founders. That may not fit into the social and cultural context of Malaysian. If adjustments can be made to cater to Malaysian contact, that may improve the usability of the m-commerce site by women sellers in the future. Ahmad [13] has pointed out that a good interface design element may encourage women sellers to sell online for long-term business expansion. Established women sellers may consider expanding their business through m-commerce sites.

RQ2: What personality types can influence women seller using m-commerce site?

All the research studies related to personality types conducted in e-learning environment in state of m-commerce site may adopted by women seller in rural area. The respondents in the journal found are higher education learner using web-based learning system as shown in Table 3.

Table 3 The personality types studies

Author	Research Title	Personality Types	Research Finding
[11]	Interface designs with MBTI personality types	Myers-Briggs Type Indicator (MBTI)	Research focus on interface design preferred by higher education learner using e-learning tools MBTI has eight dominant dimensions, that include Introvert and Extrovert, Sensing and Intuitive, Feeling and Thinking, Judging and Perception. Based on the finding, the highest number of respondents are from the ESTJ, ISFJ and INFJ personality type. User interfaces in the mobile application has developed based on the preferred three personality types.
[14]	A First Look at the Effectiveness of Personality Dimensions in Promoting Users' Satisfaction with the System	Big Five personality	This research discusses personalization of user interface based on user's personality characteristics help to generate student's satisfaction in a mobile learning context. Big five personality make out of Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness dimensions. MUIDE instrument include 11 design principles includes information structure, navigation, layout, font style, size, button, color, list,

			information density, support and alignment the finding is neuroticism group was higher satisfaction when using their preferred interface. If learners are satisfied with the UI design, they will have high interaction with mobile content.
[15]	User Interface Design for Effective E-Learning based on Personality Traits	Eysenck Personality Traits	This research finds out the learner personal traits preferred web-based e-learning environment. Eysenck has three orthogonal higher order factors: extraversion, neuroticism and psychoticism. Based on the study, majority of learner are extraversion prefer blue color and times new roman font style. Incorporate user interface design with the preferred font and color will improve the performance of students.

User interface is an important medium for women sellers to input instruction to m-commerce site. Researchers Subaramaniam [11], Sarsam [14], and Arockiam [15] carried out studies on the preferred user interface design for learners online. Subaramaniam studied the preferred mobile interface design based on MBTI personality types of learners. Whereas Sarsam [14] studied learner's preferred interface design based on Big five personality types. Arockiam [15] studied learner's preferred interface design based on Eysenck personality traits. All three researchers give advices to guide design the user interface for m-commerce. These advices are useful to implement on m-commerce for women seller to sell products online.

4. Conclusion

The journal articles about women sellers in rural areas using m-commerce sites to sell their products are hardly found. Based on the previous research studies, learners at higher education institutions are not comparable with women sellers in rural areas. The journal's collected and discussed may not represent the true perception of women sellers. However, basic user interface design may be applicable to women sellers. Further research studies should carry out to cater to women sellers in rural areas to have a better understanding of their needs.

References

1. Jabatan Tenaga Kerja Sarawak. (2021). *Sarawak Labour Outlook 2021(Jan-Dec).* Kuching: Jabatan Tenaga Kerja Sarawak.
2. Yap, C.S., Keling, W., & Ho, P.L. (2023). *Determinants of entrepreneurial performance of rural indigenous women entrepreneurs in Sarawak, Malaysia.* Gender in Management: An International Journal. doi: 10.1108/GM-06-2021-0170
3. Ellis, F. (2000). The determinants of rural livelihood diversification in developing countries. *Journal of agricultural economics*, *51*(2), 289-302.
4. Rosli, A., Daud, A., & Saad, J. M. (2021). *Iban Women and Side-Income Generation in Bintulu.* International Journal of Academic Research in Business and Social Sciences, 11(17), 254–262
5. United Nations Conference. (2020). *Digital platforms and value creation in developing countries:* Geneva: United Nations.
6. Al-Abiary, S. M.-U. (2018). *Mobile Commerce in Malaysia—Opportunities and Challenges.* Journal of Computational and Theoretical Nanoscience, 24(6), 4126-4128. doi:10.1166/asl.2018.11555
7. Jyoti, G. K. (2023, March - April). *Research Paper on Human Computer Interaction.* International Journal for Multidisciplinary Research (IJFMR), 5(2), 1-9.
8. Levy, S. (2023, March 23). graphical user interface. Encyclopedia Britannica.
9. Hassna, G., Rouibah, K., Lowry, P. B., Paliszkiewicz, J., & Mądra-Sawicka, M. (2023). *The roles of user interface design and uncertainty avoidance in B2C ecommerce success: Using evidence from three national cultures.* Electronic Commerce Research and Applications, 61, 101297.
10. Yahaya, S., Hamid, S. N. A., & Nafi, S. N. M. (2022). *Determinants for m-commerce adoption in Malaysian SMEs: A conceptual Framework.* International Journal of Business and Economy, 4(1), 138-149.
11. Subaramaniam, K., Ern-Rong, J. L., & Palaniappan, S. (2020). *Interface designs with MBTI personality types.* The 7th Mechanical Engineering Research Day, 178-179.
12. Jylhä, H., & Hamari, J. (2019). *Development of measurement instrument for visual qualities of graphical user interface elements (VISQUAL): a test in the context of mobile game icons.* User Modeling and User-Adapted Interaction, 30(5), 949-982.
13. Ahmad, Z., & Ibrahim, R. (2017). Mobile commerce (M-commerce) interface design: A review of literature. J. Comput. Eng, 19, 66-70.
14. Sarsam, S. M., & Al-Samarraie, H. (2018). *A first look at the effectiveness of personality dimensions in promoting users' satisfaction with the system.* Sage Open, 8(2), 2158244018769125.

15. Arockiam, L., & Selvaraj, J. C. (2013). *User interface design for effective e-learning based on personality traits*. International Journal of Computer Applications, 61(14).

Authors Introduction

Irene Kho Lee Sa

She received her Bachelor in Computer Science in 2001 from university of Wollongong, Australia. In 2006, she received her Corporate Masters in Business Administrative UNIMAS, Malaysia. She is currently a Doctoral student in UCSI university, Malaysia.

Associate Professor Ts. Dr. Shayla Islam

She is a Graduate Engineer of the Board of Engineers Malaysia and the Institution of Engineers Bangladesh (MIEB-M/40624). She has also evaluated her Ph.D. thesis at the international level as the foreign examiner. She is currently a Deputy Director and Head of Research at the Institute of Computer Science and Digital Innovation (ICSDI), UCSI University, Malaysia.

Associate Professor Dr. Abdul Samad Bin Shibghatullah

He received his Bachelor's degree in Accounting from Universiti Kebangsaan Malaysia, Bangi, Malaysia, in 1999, his M.Sc. degree in Computer Science from Universiti Teknologi Malaysia, Skudai, Malaysia, in 2002, and his Ph.D. degree in Computer Science from Brunel University, Uxbridge, United Kingdom. He is currently an Associate Professor at the College of Computing & Informatics (CCI), Universiti Tenaga Nasional, Kajang, Malaysia.

Obstacle-aware Autonomous Flipper Control Method Based on Terrain Geometry

Kotaro Kanazawa
Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Aichi, 466-8555 Japan
Email: k.kanazawa.837@nitech.jp

Noritaka Sato
Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Aichi, 466-8555 Japan

Yoshifumi Morita
Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Aichi, 466-8555 Japan

Abstract

Rescue robots play crucial roles in search and rescue missions at disaster sites. Crawler mechanisms, which are valued for their high traversability on unstructured terrain, can be equipped with flipper arms featuring four single-rotation joints to enhance their performance. However, operating these flippers is challenging. Although various control methods have been developed recently, difficulties persist in adapting to three-dimensional uneven terrains and in avoiding interference between flippers and the environment. In this study, we propose an obstacle-aware autonomous flipper control method that actively adjusts the flippers based on the progress of the robot and terrain geometry.

Keywords: Autonomous Flipper Control, Rescue Robotics, Crawler-type Robot

1. Introduction

Recently, there has been a growing demand for remote controlled rescue robots that can gather information at disaster sites. At disaster sites, where there is rubble and uneven ground, many rescue robots use a crawler mechanism called a flipper (sub-crawler), which is a mobile mechanism with four joints, to adapt to uneven ground [1]. Crawler robots equipped with flippers can climb high steps, this allows them to expand the range of their activities. Flippers can also be used to change the posture of the vehicle body to ensure stability on uneven ground [2], and can also be used to cushion the impact on the robot when it descends steps. However, when operating a remotely controlled robot, it is extremely difficult to operate the four flippers independently. Particularly, when remotely controlling a robot in an unstructured environment, such as rubble, there is a risk of the robot falling over. Thus, research has been conducted on the autonomous control of flippers. As a strategy for traversing rough terrain for crawler-type robots, the contact area with rough terrain is maximized and the robot posture is appropriately transitioned using flippers. As a method of semi-autonomous control that utilizes the interaction between the robot and the environment, a method using passive flippers that move in response to gravity and the reaction force of obstacles [3] and a method of changing the joint torque in response to the robot posture based on rules [4] have been proposed. These methods are useful because they do not require high-load computational processing. However, the robot may unintentionally get stuck in the environment and immobilize itself because of the lack of terrain information. Recently, 3D LiDAR sensors have been used in crawler-type robots for environmental information gathering [1], [5], [6], [7], [8]. In these studies, it is possible to plan the movement of the flippers according to the terrain from the sensor information [5], [6]. Additionally, it is possible to maintain a stable posture even on uneven terrain by estimating the contact points between the robot and the terrain [7]. Furthermore, the path following on uneven terrains can be achieved using an environmental map [8]. However, many studies only consider obstacles in the forward direction and often drive the left and right flippers in sync [5], [6], [7]. As a result, they might struggle with 3D uneven terrains, such as entering a slope from the side. Even in methods where each flipper adapts to the terrain, side collisions during turns are not considered [1], [8]. Yuan et al. proposed a control method that predicts the contact between the flipper and the ground [9]. They demonstrated that overcoming a step diagonally or approaching a slope from the side by driving the four flippers independently is possible. However, because the goal is to maximize the contact area with the terrain, the pitch angle may increase if the step is large. In this study, we developed a flipper control method that considers contact with the environment from the flipper's side, utilizing environmental information on 3D uneven terrain. Furthermore, by using a grid map, which is a smooth representation of a 3D map, we ensured the continuity of the robot's posture. We evaluated the proposed method through physical simulations on 3D uneven terrain to determine its traversability.

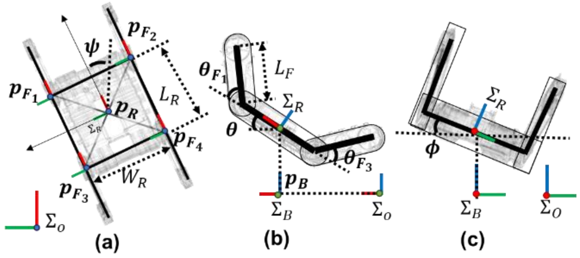


Fig. 1 Coordinate system of the crawler robot: (a) top view, (b) side view from the left, and (c) front view.

2. Methodology

2.1 Terrain Processing

We used the contact conditions of a crawler robot with the flipper skeleton model and the terrain expansion used by Yuan et al. [9]. A simplified model that connected each link with a straight line and expanded the terrain with the radius of the pulley. In this method, the diameter of the flipper pulley must be unique for the crawler and flipper. The flipper angle is constrained within a range of −90° and 90°. When the robot exists in the world coordinate system Σ_0, its position and posture are represented as $P_0^R = [x, y, z]^T$ and $\Phi_0^R = [\psi, \theta, \phi]^T$, respectively (Fig. 1). The base coordinate system Σ_R, which is defined as z = 0 in the vertical direction below the center of the robot, is defined as follows:

$$\mathbf{p}_B^o = [x, y, 0]^\top, \quad \mathbf{R}_B^o = R(\psi, 0, 0), \tag{1}$$

where $R(\psi, \theta, 0)$ is a rotation matrix with yaw angle ψ.

Target posture of the robot was generated by smoothening the terrain. In this study, we assumed that the terrain was accurately represented as a point-cloud map and that two layers were generated using a grid map (Fig. 2). h_F is the terrain expanded by the wheel radius. If the skeleton model is in contact with h_F, it is inferred that the actual robot is in contact with the ground or wall. This layer was used for angle estimation based on the estimation of the flipper in contact with the ground, which is described later. h_R is a terrain that gently connects to uneven ground. This method is based on the method of obtaining the body-referenced posture of a four-legged robot by Fabian et al. [10]. In this study, we adjusted the parameters of each filter to apply the proposed method to crawler robots. The following inequality must hold:

$$(k_2 + k_{3x}) * \alpha_x \geq 2L_F + L_R \tag{2}$$

$$(k_2 + k_{3y}) * \alpha_y \leq 2W_R \tag{3}$$

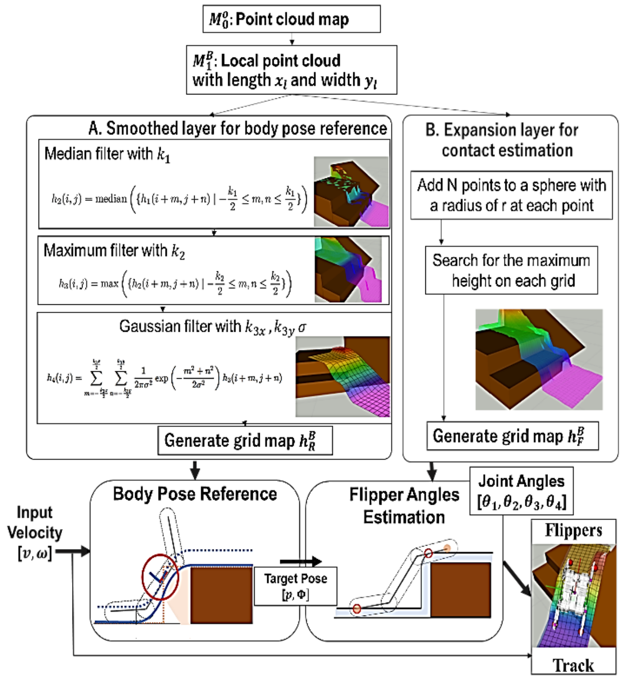


Fig. 2 By processing the point cloud around the robot, two grid map layers were obtained. Each layer is used for pose reference and flipper angle estimation.

here, alpha represents the resolution for each respective variable.

2.2 Autonomous Flipper Control Method

The aim of the control system is to avoid collisions between the flippers and the environment and to overcome obstacles when the robot is given a speed input. The body posture is obtained based on the terrain information, and the four flipper angles were calculated. The target posture of the robot is calculated using gradient information in the grid map $= h_R^B$. When the speed input $v_R = [v_x, 0, 0]^T$ and $\omega = [0, 0, \omega_z]^T$ is given to the crawler, the next position is calculated. First, the velocity input was converted from the robot coordinate system to the base coordinate system V_B and ω_B. Subsequently, the next robot position from the current base coordinate system was calculated. The next position $p_R^{B\prime}$ is updated as follows:

$$\mathbf{p}_R^{B\prime} = (V_x^B/\omega_z) \cdot [\sin(\omega_z \Delta t), 1 - \cos(\omega_z \Delta t), 0]^\top, \tag{4}$$

where Δt denotes a time step. The height z' and posture p_{ref}^B of the robot in the next step are obtained from the smoothened surface $h_R^B(x, y)$, where x and y are the positions of the robot of $p_R^{B\prime}$, and h_R^B returns the height corresponding to that position. Also, the pitch and roll angles were calculated based on the slope of the height information in the grid map. The flipper angle is the maximum angle at which each flipper contacts the ground when the body is at its target position. Based on the flipper angle and length, we determined whether the position of the endpoint of each flipper in the base coordinate system, $p^B_{F_i e} = (x^B, y^B, z^B)$, contacted the grid map h^B_F. In the robot coordinate system, the endpoint position of the flipper was calculated as follows:

$$\begin{aligned} x^R &= x_i + (-1)^{\lfloor \frac{i-1}{2} \rfloor} \cdot L_F \cdot \cos(\theta), \\ y^R &= y_i, \\ z^R &= L_F \cdot \sin(\theta). \end{aligned} \tag{5}$$

where x_i, $y_i$are the joint positions of the flipper, L_F is the is the flipper's length, and L_F to r are the values of the range of contact. θ is the flipper's angle, and$(-1)^{\lfloor \frac{i-1}{2} \rfloor}$ is a symbol indicating whether the flipper is in front or not.

3. Experimental Results

To evaluate the performance of traversing an uneven terrain in three dimensions, we conducted a simulation of the movement of the crawler robot using Gazebo, an ROS standard physics simulator as shown in Fig.3. In the experiment, we verified (a) the ability to follow the target posture in random steps, and (b) the ability of the flipper to avoid obstacles when turning. We also investigated the limits of the ability of the robot to traverse slopes and steps without falling over by changing the angle of incidence in (c) traversing slopes from the side and (d) traversing steps from the side. The robot moves straight ahead at a constant speed in (a), (c), and (d), and turns in (b). In the simulation of the crawler belt in the Gazebo, we used the method of Okada et al.[11]. We use the following parameters for the experiment: $L_B = 0.45$, $W_B = 0.46$, $L_F = 0.29$, $r = 0.082$, $k_1 = 7$, $k_2 = 4$, $k_{3x} = 11$, $k_{3y=}7$, $\sigma = 5.0$, $\alpha = 0.04$.

Fig. 4 illustrates the robot's movements during the task. As shown in (a), even when the robot approached the step at a 45° angle, it was able to climb over the 30 cm high step. Fig. 5 presents the time series data of the target and actual values in Task (a). When the robot approached the step, the right front flipper rotated according to the step height following the Expansion Layer. Afterward, according to the Smoothed Layer, target values for roll and pitch angles were provided, and by moving the four flippers separately, the target posture was followed. However, when the robot proceeded at a 50° angle, the roll angle increased significantly before the left front flipper contacted the step, causing the robot to fall. Thus, it was found that traversability differed depending on the direction of movement. When using the robot in practice, this issue could be addressed by imposing movement direction restrictions based on the Smoothed Layer during motion. In (b), even when the robot turned close to an obstacle, the flippers did not collide with the wall surface, following the Expansion Layer. In (c), the robot was able to climb a 30° slope from the side. When

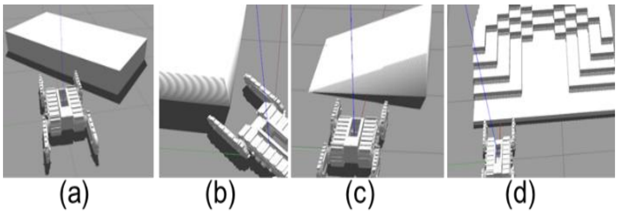


Fig.3 Test environment in Gazebo.

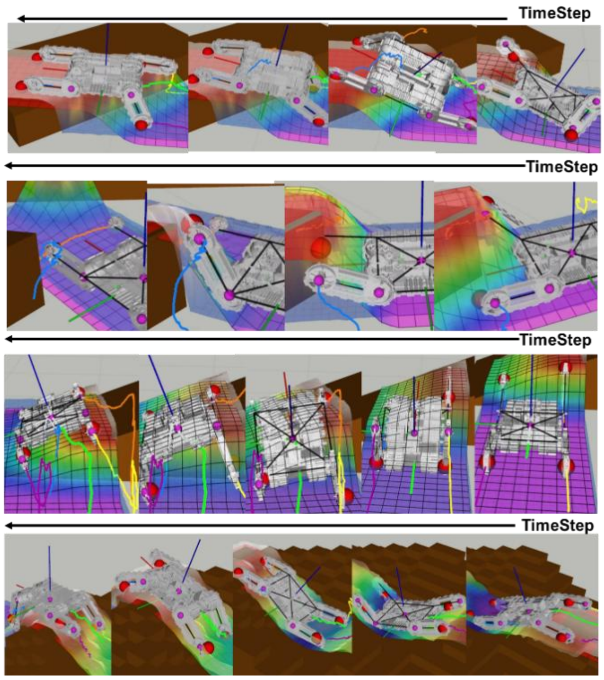


Fig. 4 Robot movement during the task. From top to bottom, (a), (b), (c), (d). All time steps are from the right image to the left image.

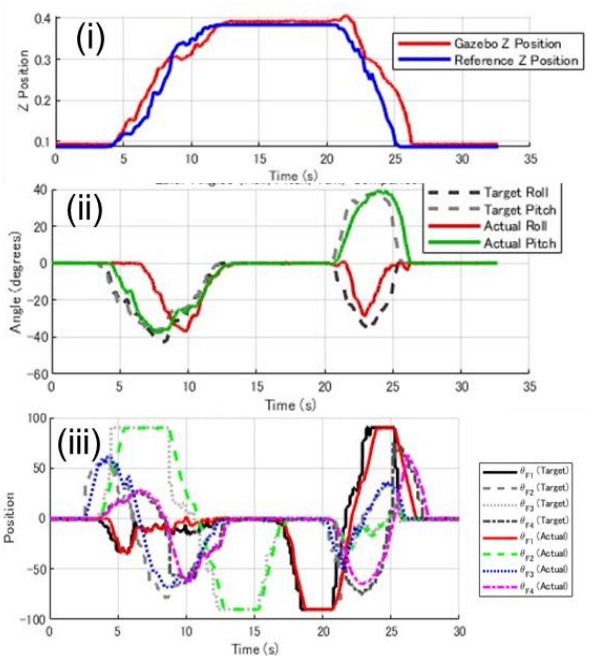


Fig. 5 Time series data of the target and actual values for (i) the robot's center of gravity (z-coordinate), (ii) the robot's posture, and (iii) each joint angle in Task (a).

approaching the obstacle, the angles of the left and right flippers changed according to the height difference. After climbing, roll angle commands were given according to the Smoothed Layer, and the robot proceeded by aligning its body to the slope of the ground. In (d), even on irregular terrain where the pitch angle fluctuated, the robot was able to adjust its posture according to the terrain shape and proceed. Additionally, the flippers did not get stuck during the execution of these steps.

4. Conclusion

In the proposed method, a flipper control strategy that considers contact with the environment during turning was developed, demonstrating the possibility of overcoming obstacles on uneven terrain in three dimensions. The target posture of the robot was continuously generated using smooth curved surfaces. This can be applied for navigation purposes, using a path-following algorithm. However, in the existing methodology, the target posture and joint angles are calculated at each time step. It should be noted that the continuity of the flipper angle values cannot be guaranteed. Furthermore, since the experiments in this study assume that the true values of the robot's posture in its environment and world coordinate system can be obtained, sensor noise and position errors must be addressed when applying the method to an actual robot. Additionally, as the objective was to ensure that the robot's posture followed smooth terrain, the stability of the robot was not considered. Future work includes generating target angles that match the flipper drive speed and developing a control method that accounts for stability.

Acknowledgements

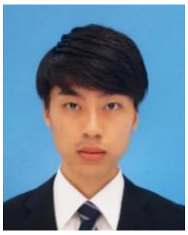
This work was supported by JST SPRING, Grant Number JPMJSP2112.

References

1. Y. Okada, K. Nagatani, K. Yoshida, S. Tadokoro, T. Yoshida, and E. Koyanagi, “Shared autonomy system for tracked vehicles on rough terrain based on continuous three-dimensional terrain scanning,” Journal of Field Robotics, vol. 28, no. 6, pp. 875–893, 2011.
2. F. Rocha, A. Cid, M. Delunardo, R. P. Junior, N. S. Thiago, L. Barros, J. D. Domingues, G. Pessin, G. Freitas, and R. Costa, “Body posture controller for actively articulated tracked vehicles moving over rough and unknown terrains,” in 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2023, pp. 2330–2337
3. A. Watanabe, T. Mitsuhashi, M. Okugawa, K. Ogane, T. Kimura, T. Kinugasa, Y. Ohtsubo, "Ground Adaptability of Crawler Mobile Robots with Sub-Crawler Rotary Joint Compliance," *Journal of Robotics and Mechatronics*, 2024, Vol. 36, No. 3, pp. 732-745.
4. R. Michikawa, W. Xixun, Y. Okuda, Y. Morimoto, T. Shibuya, F. Matsuno, “Semi-autonomous flippers control to aid crawler robots in maneuvering over rough terrain without terrain information”, Proceedings of the 67th Annual Conference of the Institute of Systems, Control and Information Engineers (ISCIE), vol. 67, pp. 355–361, 2023.
5. B. Chen, K. Huang, H. Pan, H. Ren, X. Chen, J. Xiao, W. Wu, and H. Lu, “Geometry-based flipper motion planning for articulated tracked robots traversing rough terrain in real-time,” Journal of Field Robotics, vol. 40, no. 8, pp. 2010–2029, 2023.
6. Z. Xu, Y. Chen, Z. Jian, J. Tan, X. Wang, and B. L. Liang, “Hybrid trajectory optimization for autonomous terrain traversal of articulated tracked robots,” IEEE Robotics and Automation Letters, vol. 9, no. 1, pp. 755–762, 2023.
7. M. Norouzi, J. V. Miro, and G. Dissanayake, “Planning stable and efficient paths for reconfigurable robots on uneven terrain,” Journal of Intelligent & Robotic Systems, vol. 87, pp. 291–312, 2017.
8. M. Gianni, F. Ferri, M. Menna, and F. Pirri, “Adaptive robust three-dimensional trajectory tracking for actively articulated tracked vehicles,” Journal of Field Robotics, vol. 33, no. 7, pp. 901– 930, 2016.
9. Y. Yuan, Q. Xu, and S. Schwertfeger, “Configuration-space flipper planning on 3d terrain,” in 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR). IEEE, 2020, pp. 318–325.
10. F. Jenelten, R. Grandia, F. Farshidian, and M. Hutter, “Tamols: Terrain-aware motion optimization for legged systems,” IEEE Transactions on Robotics, vol. 38, no. 6, pp. 3395–3413, 2022.
11. Y. Okada, S. Kojima, K. Ohno and S. Tadokoro, "Real-time Simulation of Non-Deformable Continuous Tracks with Explicit Consideration of Friction and Grouser Geometry," *2020 IEEE International Conference on Robotics and Automation (ICRA)*, Paris, France, 948-954, 2020.

Authors Introduction

Mr. Kotaro Kanazawa



He received his Bachelor's and Master’s degree in Engineering,2022 and 2024, respectively, from the Department of Engineering, Nagoya Institute of technology in Japan. He is currently a Ph.D. student in Nagoya Institute of Technology, Japan.

Dr. Noritaka Sato

He is an Associate Professor in the Electrical and Mechanical Engineering Program, Department of Engineering, at Nagoya Institute of Technology. He received his PhD in Engineering from the University of Electro-Communications in 2009. His current research interests include rescue robots and human-robot interaction.

Prof. Yoshifumi Morita

He is a professor in the Electrical and Mechanical Engineering Program at the Department of Engineering, Nagoya Institute of Technology. He received his Ph.D. from the Nagoya Institute of Technology in 1998. His current research interests include robots and devices for rehabilitation assistance, and robot teaching.

An Adaptive Control Method for a Knee-Joint Prosthetic Leg Toward Dynamic Stability and Gait Optimization

Yiqian Ge
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Choisuren Purevdorj
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Shintaro Kasai
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hiroaki Wagatsuma
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Email: ge.yiqian503@mail.kyutech.jp, purevdorj.choisuren544@mail.kyutech.jp, kasai.shintaro660@mail.kyutech.jp, waga@brain.kyutech.ac.jp

Abstract

This paper presents a hybrid control strategy that combines Model Predictive Control (MPC) and Linear Quadratic Regulator (LQR) to achieve robust and stable tracking of human knee joint motion. The state-space model of the system is discretized to facilitate real-time implementation. MPC is employed to track the knee joint trajectory during dynamic motion, while the LQR controller is activated at critical points, particularly when the joint angle approaches zero, to stabilize the system and ensure safety.

Keywords: Model Predictive Control (MPC), Linear Quadratic Regulator (LQR), Trajectory Tracking, Lower-Limb Prosthetics, Knee Joint Motion

1. Introduction

In recent decades, millions of individuals have faced challenges in using their lower limbs due to conflicts, illnesses, traffic accidents, and natural disasters. Consequently, many have lost their capacity to work and are unable to engage in normal social activities [1]. Studies have explored the potential of motorized prostheses to help patients regain their walking ability [2]. Active prostheses have inspired numerous control strategies. However, despite advancements in control strategies, challenges remain in the mechanical design of active prostheses [3].

Sup et al. (2008) [4] developed an electrically driven prosthetic knee and ankle using a ball-screw mechanism. The control strategy employed a finite state machine-based impedance control to ensure stable torque output across gait phases.

Martinez-Villalpando et al. (2009) [5] developed an agonist-antagonist active knee prosthesis featuring two parallel-series elastic actuators that mimic the biomechanical characteristics of the human knee.

Salman and Kadhim (2022) [6] proposed a dual-degree-of-freedom knee prosthesis control method based on backstepping control. Utilizing Lyapunov stability theory, they designed a recursive controller to ensure dynamic stability of the closed-loop system. The control parameters were further optimized using the bat algorithm, significantly improving trajectory tracking accuracy and robustness.

While Taherian et al. [7] explored the application of MPC and LQR for collision avoidance in autonomous driving, their approach also provides valuable insights for developing control strategies applicable to prosthetic systems. Similarly, Capron et al. [8] highlighted the stability and performance benefits of robust LQR-MPC control in industrial processes, offering a solid theoretical foundation for designing effective control strategies in dynamic environments such as prosthetic systems.

This paper presents an MPC+LQR-based approach for tracking the human knee joint angle trajectory during walking. To reduce tracking errors at critical walking points, a switching mechanism to LQR control is designed when the prosthetic knee reaches the support point (0 degrees). The leg model is developed using the Lagrange equation and transformed into a discrete state-space form for the implementation of the MPC+LQR controller. Simulation results demonstrate that compared to using MPC alone, the integrated control strategy significantly reduces trajectory errors at critical points and improves the accuracy of tracking human knee motion. The inclusion of LQR control further enhances system stability and robustness.

2. Control Objective

The objective of this study is to establish a mathematical model of the leg and design a controller to regulate knee joint angle variations, ensuring they align with the knee motion trajectory observed during human movement (Fig.1).

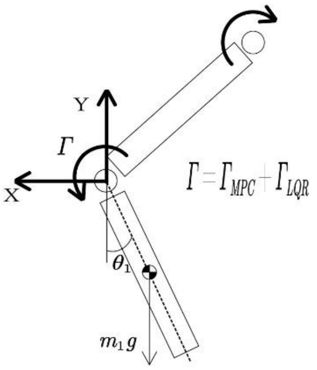


Figure.1. Mathematical Modeling

Based on the measurements conducted by Thomas Seel et al. (2014) [9], the knee joint angle variations can be approximated as a periodic motion with a duration of 1.1 seconds. In this paper, we use a sine function to fit the bending process of the knee joint and apply smoothing techniques to simulate the smooth fluctuations of the knee angle when the foot makes contact with the ground, ensuring the function remains differentiable within the period.

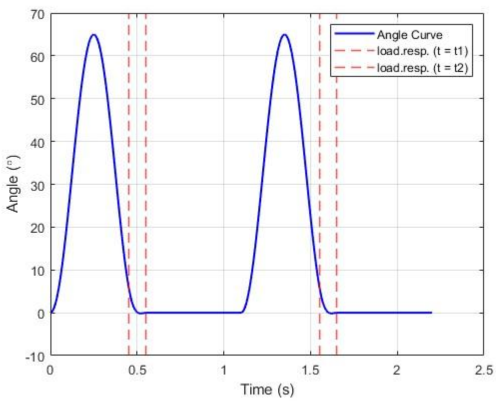


Figure.2. Angle Curve

In the figure (Fig.2), t_1 to t_2 are the start and end times of foot contact with the ground, making the knee Angle smoothly transition to zero.

3. Methodology

3.1. Model Predictive Control modeling

In the simulation of the legs, we employ two standard homogeneous rods connected by revolute joints to replicate the human leg motion mechanism. The lengths and masses of the rods are set as estimated values based on standard adult male anthropometric assumptions [10]. The schematic diagram is presented in Figure (Fig.3):

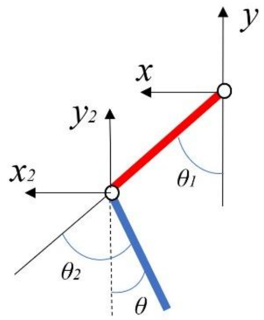


Figure.3. Coordinate diagram of leg mechanism

The coordinate system is defined with the hip joint rotation point as the origin. The X-axis is horizontal to the left. The Y axis goes straight up, and the gravity direction is the -y axis. The whole system is negative on the Y-axis.

The Angle of rotation of the joint is defined as θ_1, θ_2. θ_1 is the Angle between the thigh connecting rod and the y axis, positive in the clockwise direction. θ_2 is the Angle between the thigh extension line and the calf, positive in the counterclockwise direction. Due to the anatomical structure of the human body, θ_2 values greater than 0.

Where, θ_2 is the relative Angle, the absolute Angle θ of the calf relative to the coordinate system is:

$$\theta = \theta_1 - \theta_2 \tag{1}$$

The following Table.1 describes other physical data：

Table. 1. Physical data

Physical quantity	Value	Unit
m_1	7	kg
l_1	0.5	m
m_2	3	kg
l_2	0.45	m
c	0.1	

3.2. Lagrange equations

Equations of Lagrange dynamics [11]:

$$L = T - U \tag{2}$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}_i}\right) - \frac{\partial L}{\partial \theta_i} = Q_i \tag{3}$$

Analyzing the dynamic model of each component.

Centroid of thigh:

$$x_1 = \frac{l_1}{2}\sin\theta_1 \qquad y_1 = -\frac{l_1}{2}\cos\theta_1 \tag{4}$$

Thigh kinetic energy:

$$T_1 = \frac{1}{2} I_1 \dot{\theta}_1^2 + \frac{1}{2} m_1 \left(\frac{l_1}{2}\right)^2 \dot{\theta}_1^2 \tag{5}$$

Thigh potential energy:

$$U_1 = -m_1 g \frac{l_1}{2}\cos\theta_1 \tag{6}$$

The calf movement is completed by the coupling of the calf rotation and the thigh rotation, wherein the calf rotation Angle θ_2 relative to the thigh is the relative Angle, and the absolute Angle of the calf is defined in Eq.1. The absolute position of the center of mass of the calf is:

$$\begin{aligned} x_2 &= l_1 \sin\theta_1 + \frac{l_2}{2}\sin(\theta_1 - \theta_2) \\ y_2 &= -l_1\cos\theta_1 - \frac{l_2}{2}\cos(\theta_1 - \theta_2) \end{aligned} \tag{7}$$

Calf kinetic energy:

$$T_2 = \frac{1}{2} m_2 \left(\dot{x}_2^2 + \dot{y}_2^2\right) + \frac{1}{2} I_2 \left(\dot{\theta}_1 - \dot{\theta}_2\right)^2 \tag{8}$$

Calf kinetic energy:

$$U_2 = -m_2 g \left(l_1 \cos\theta_1 + \frac{l_2}{2}\cos(\theta_1 - \theta_2)\right) \tag{9}$$

Find the Lagrange equation for the whole system:

$$\frac{d}{dt}\begin{bmatrix} \frac{\partial L}{\partial \dot{\theta}_1} \\ \frac{\partial L}{\partial \dot{\theta}_2} \end{bmatrix} - \begin{bmatrix} \frac{\partial L}{\partial \theta_1} \\ \frac{\partial L}{\partial \theta_2} \end{bmatrix} = \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} \tag{10}$$

From an energy perspective, the Lagrange method offers an efficient and systematic approach to modeling the entire system. This method not only effectively handles the dynamic modeling of complex systems, particularly those with rigid multi-links, but also preserves the physical consistency and energy conservation properties of the system.

3.3. State-space equation

For calf movements, define state variables:

$$x = \begin{bmatrix} \theta_2 \\ \dot{\theta}_2 \end{bmatrix}, \ u = \tau_2 \tag{11}$$

The state space model is:

$$\dot{x} = Ax + Bu \tag{12}$$

Obtained from the Lagrange equations:

$$\ddot{\theta}_2 = \frac{1}{M_{22}}(\tau_2 - C_2) \tag{13}$$

Where M_{22} is the equivalent inertial term, derive from $\frac{\partial^2 T}{\partial \dot{\theta}_2^2}$. C_2 represents the equivalent Coriolis, centrifugal, and gravitational terms.

The state space equation is derived as follows:

$$\begin{bmatrix} \dot{\theta}_2 \\ \ddot{\theta}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{\partial C_2}{\partial \theta_2}/M_{22} & -\left(\frac{\partial C_2}{\partial \dot{\theta}_2} + c\right)/M_{22} \end{bmatrix} \begin{bmatrix} \theta_2 \\ \dot{\theta}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{M_{22}} \end{bmatrix} u \tag{14}$$

Calculated by MATLAB:

$$\begin{bmatrix} \dot{\theta}_2 \\ \ddot{\theta}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{g\, l_2\, m_2 \sin(\theta_1 - \theta_2)}{2\left(\frac{m_2\, l_2{}^2}{4} + I_2\right)} & -\frac{c}{\frac{m_2 l_2^2}{4} + l_2} \end{bmatrix} \begin{bmatrix} \theta_2 \\ \dot{\theta}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{\frac{m_2\, l_2{}^2}{4} + I_2} \end{bmatrix} u \tag{15}$$

Where c is the damping coefficient, representing the resistance to be overcome by rotating the knee joint. The damping coefficient is related to the angular velocity, and the equation is

$$f = c \cdot \dot{\theta}_2 \tag{16}$$

4. MPC+LQR trajectory tracking

4.1. MPC modeling

By introducing Model Predictive Control (MPC) into the mathematical model, the controller can predict the system's behavior over a finite time horizon and continuously optimize the control input to achieve the optimal solution at the current time step. MPC achieves this by solving an online optimization problem at each control interval, which considers system dynamics, constraints, and performance objectives [12]. This predictive capability allows MPC to handle multi-variable systems, incorporate input and state constraints, and adapt to disturbances, making it particularly suitable for complex and dynamic systems.

From the previous Eq.12, it is concluded that the state-space equation of the system can be rewritten into discrete form:

$$x_{k+1} = A_d x_k + B_d u_k \tag{17}$$

Where x_{k+1} is the state of the next moment, determined jointly by the state of this moment x_k and the control input of this moment u_k, with A_d representing the state transition matrix and B_d representing the input matrix in the discrete-time system.

The discrete state space equation is incorporated into MATLAB for simulation and numerical computation. To reduce computational load, the equation is linearized around the thigh lift action. Taking $\theta_1 = \pi/2$ and $\theta_2 = \pi/3$.

Concurrently, to simulate the knee joint constraints during real walking motion, the range of θ_2 is designed between 0-70 degrees, while the target trajectory angle range is set between 0-60 degrees. In addition, to model the errors and external disturbances encountered in real-world scenarios, the disturbance function is introduced as:

$$d(t) = A_d dsin(\omega_d t) \tag{18}$$

MPC possesses the capability to track nonlinear motion trajectories, which can be achieved by inputting the trajectory data into the MPC controller. By incorporating the target trajectory into the MPC controller, the tracking plot can be obtained as figure.

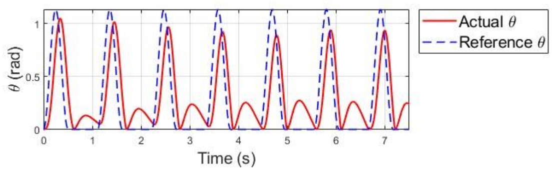


Figure. 4.A MPC Angle Tracking Curve

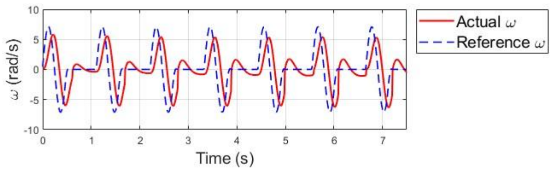


Figure. 4.B MPC Velocity Tracking Curve

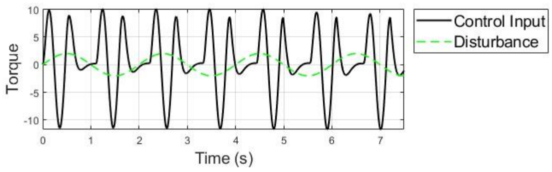


Figure. 4.C Control Input and Disturbance

Figure. 4. MPC tracking curve

The green dashed line represents the disturbance applied to the system. As shown in the figure (Fig.4), with output weights set to [25, 0.5], the MPC achieves most of the trajectory tracking under the combined influence of the disturbance function and system damping, with a calculated Root Mean Square (RMS) Error of 0.28851 rad. However, significant errors occur at the point where the system angle equals zero.

4.2. MPC+LQR Control scheme

To reduce tracking errors to the point where the angle equals 0, we introduced LQR control and designed a control logic to switch to LQR when the tracking trajectory approaches 0 degrees.

Linear Quadratic Regulator (LQR) is a linear control scheme where the optimal gain matrix is computed off-line and applied in real-time for system regulation [13]. By introducing the performance evaluation equation as:

$$J = \sum_{k=0}^{\infty} (x_k^T Q x_k + u_k^T R u_k) \tag{19}$$

By setting the values of the state weight matrix Q and the disturbance weight matrix R, the feedback gain matrix K is determined to satisfy:

$$u_k = -Kx_k \tag{20}$$

After introducing the LQR control with the state weight matrix $Q = \begin{bmatrix} 300 & \\ & 1 \end{bmatrix}$ and input weight $R = 0.05$ at the point where the angle is zero, the simulation results are shown in the figure (Fig.5) below:

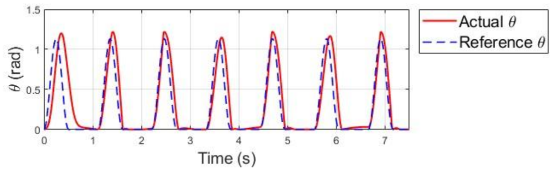


Figure. 5.A MPC+LQR Angle Tracking Curve

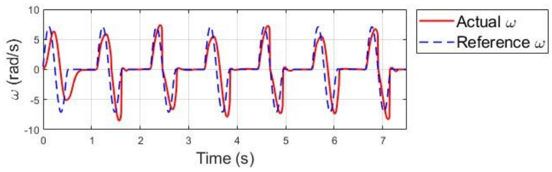


Figure. 5.B MPC+LQR Velocity Tracking Curve

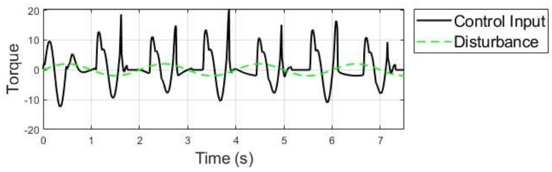


Figure. 5.C Control Input and Disturbance

Figure. 5. MPC + LQR tracking curve

As shown in the figure (Fig.5), with the same tuning parameters, the combined MPC and LQR controller reduced the RMS Error to 0.20437 rad, and the simulation scheme demonstrates better tracking performance when the knee joint angle is at 0 degrees.

5. Discussion

5.1. Interpretation of result

Both MPC and LQR controllers are derived from the state-space equations, ensuring that the transitions of knee joint angle and angular velocity align with human motion patterns during movement. This is also evident from the simulation results, where both the MPC controller (Fig.4) and the MPC+LQR controller (Fig.5) successfully track the trajectory for most of the time.

However, as shown in the simulation results of the standalone MPC controller (Fig.4), a significant error occurs in the predicted model when the angle approaches zero. The Root Mean Square (RMS) error for the standalone MPC controller is 0.28851 rad, reflecting this limitation.

This is primarily due to the combined effects of MPC prediction errors and the physical limitations of the knee joint. When the knee angle reaches zero, corresponding to full knee extension in physical terms, the angular velocity reduces to zero due to physical impacts. This reduction introduces significant errors in the MPC's predicted future control path, leading to inaccuracies in the actual trajectory at the zero-degree position.

These errors are inherent to the predictive control principle of MPC. While adjusting control parameters can reduce the amplitude of fluctuations, residual oscillations will still exist. Importantly, when the angle equals zero, the prosthetic knee acts as the stance leg during human walking, a critical phase where even minor oscillations in the knee joint can destabilize the wearer's center of mass.

To address this issue, LQR control is introduced when the knee angle approaches zero to stabilize the system and ensure safe and reliable gait performance for the wearer. From the simulation parameters (Fig.5), it is evident that the combined MPC+LQR control strategy not only conforms to the human motion trajectory but also reduces the RMS error to 0.20437 rad, ensuring that the knee joint is controlled to a safe position at critical points and maintaining motion stability.

5.2. Reliability and Scalability

In LQR control, the optimal gain K is typically precomputed offline by solving the Algebraic Riccati Equation, thereby avoiding real-time computational burden during implementation [14]. Compared to MPC, which excels in handling constraints and trajectory tracking tasks, LQR provides a computationally efficient solution and performs well in systems requiring fast and robust disturbance rejection [15]. Studies have demonstrated that while MPC offers smoother control signals, LQR's simple structure and reduced computational demand make it suitable for applications with limited real-time resources [16].

Although the LQR controller is limited to tracking linear tasks, the combined MPC and LQR control scheme enables the system to not only track nonlinear trajectories but also switch to linear tasks at critical points, ensuring system stability. This approach enhances the overall robustness and scalability of the system. Regardless of the type of motion trajectory being tracked, the LQR controller maintains stability when the prosthesis transitions to the supporting leg, providing greater safety for the wearer. Additionally, the system allows for the customization of MPC tracking paths by measuring the wearer's unique motion trajectories, offering more walking modes for powered prostheses.

6. Conclusion

In this paper, we have successfully developed a hybrid control strategy that integrates Model Predictive Control (MPC) and Linear Quadratic Regulator (LQR) to address the challenges of controlling the knee joint angle in a prosthetic system. By achieving the challenge of tracking the target trajectory, we also established a reliable control logic that ensures stability at critical points during movement. Through Lagrangian modeling of the human leg mechanism, we derived the state-space equations for the lower limb, enabling precise mathematical representation of its dynamics.

Simulation results demonstrate that this control scheme exhibits better performance in both trajectory tracking and maintaining stability at critical points. The combined LQR+MPC control strategy not only improved the overall trajectory tracking performance but also enhanced system robustness and safety by minimizing angular oscillations at critical phases. Furthermore, the transition between MPC and LQR control was shown to align well with natural human motion, providing smooth and physically consistent angle and velocity changes. Future work will focus on applying this control strategy to real lower-limb prosthetics. By measuring the user's gait, the MPC tracking trajectory can be customized, enabling the prosthetic to better adapt to the user's individual movement patterns.

Acknowledgements

This work was supported in part by JSPS KAKENHI (JP17H06383, JP24K07387), Grant-in-Aid for JSPS Fellows (23KJ1754).

References

1. X. Zhang, J. Li, Z. Hu, W. Qi, and L. Zhang, "Novel design and lateral stability tracking control of a four-wheeled rollator," Applied Sciences, 9(11), 2327, 2019.
2. R. Borjian, "Design, modeling, and control of an active prosthetic knee." University of Waterloo, Ph.D. thesis, 2008.
3. M. R. Tucker, J. Olivier, A. Pagel, H. Bleuler, M. Bouri, O. Lambercy, J. R. Millan, R. Riener, H. Vallery, and R. Gassert, “Control strategies for active lower extremity prosthetics and orthotics: a review,” Journal of NeuroEngineering and Rehabilitation, Vol.12, No.1, 2015.
4. F. Sup, H. A. Varol, J. Mitchell, T. Withrow, and M. Goldfarb, "Design and control of an active electrical knee and ankle prosthesis," 2008 2nd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics, Scottsdale, AZ, USA, pp. 523-528, 2008.
5. E. C. Martinez-Villalpando, and H. Herr. "Agonist-antagonist active knee prosthesis: a preliminary study in level-ground walking," Journal of Rehabilitation Research & Development, vol.46, No.3, pp.361-374, 2009.
6. M. A. Salman, and S. K. Kadhim. "Optimal backstepping controller design for prosthetic knee joint," Journal Européen des Systèmes Automatisés, Vol.55, No.1, pp.49-59, 2022.

7. S. Taherian, K. Halder, S. Dixit, and S. Fallah, "Autonomous collision avoidance using MPC with LQR-based weight transformation." Sensors, Vol.21, No.13, 4296, 2021.
8. B. D. O. Capron, and D. Odloak, "A robust LQR-MPC control strategy with input constraints and control zones," Journal of Process Control, Vol.64, pp.89-99, 2018.
9. T. Seel, J. Raisch, and T. Schauer. "IMU-based joint angle measurement for gait analysis." Sensors, Vol.14, No.4, pp.6891-6909, 2014.
10. J. L. Durkin, " Routledge Handbook of Biomechanics and Human Movement Science: Measurement and estimation of human body segment parameters," 1st edn, Routledge, 2008.
11. M. Zefran, and F. Bullo. " Robotics and Automation Handbook: Lagrangian dynamics," CRC Press, 2005.
12. K. S. Holkar, and L. M. Waghmare. "An overview of model predictive control." International Journal of Control and Automation, Vol.3, No.4, pp.47-63, 2010.
13. B. D. O. Anderson, and J. B. Moore. "Optimal control: linear quadratic methods," 1st edn, Prentice-Hall Inc., 2007.
14. G. Goel, and B. Hassibi. "The power of linear controllers in LQR control," 2022 IEEE 61st Conference on Decision and Control (CDC), Cancun, Mexico, pp.6652-6657, 2022.
15. A. Jezierski, J. Mozaryn, and D. Suski, "A comparison of LQR and MPC control algorithms of an inverted pendulum," Polish Control Conference. Cham: Springer International Publishing, 2017.
16. B. E. Durmaz, B. Kaçmaz, I. Mutlu, and M. T. Söylemez, "Implementation and comparison of LQR-MPC on active suspension system," 2017 10th International Conference on Electrical and Electronics Engineering (ELECO), Bursa, Turkey, pp. 828-835, 2017.

Authors Introduction

Mr. GE Yiqian

He received his Bachelor's degree in Engineering in 2020 from the school of Mechanical Engineering, Hubei University of Technology in China. He is currently a master student in Kyushu Institute of Technology, Japan.

Mr. Purevdorj Choisuren

He received his master's degree in mechanical engineering from Mongolian University of Science and Technology (MUST), Mongolia in 2008. He is currently a doctoral course student in Graduate School of Life Science and Systems Engineering in Kyushu Institute of Technology, Japan.

Mr. Shintaro Kasai

He received his Master's degree in Engineering in 2023 from the Graduate School of Life Science and System Engineering, Kyushu Institute of technology (Kyutech) in Japan,. He is currently a doctoral course student in Kyutech, Japan and JSPS Research Fellow (DC1).

Dr. Hiroaki Wagatsuma

He received his M.S., and Ph.D. degrees from Tokyo Denki University, Japan, in 1997 and 2005, respectively. In 2009, he joined Kyushu Institute of Technology, where he is currently an Associate Professor of the Department of Human Intelligence Systems. His research interests include non-linear dynamics and robotics. He is a member of IEEE.

A Gait Analysis with Multibody Dynamics Toward Energy-Efficient Active Knee Prostheses

Choisuren Purevdorj
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Yiqian Ge
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Shintaro Kasai
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Dondogjamts Batbaatar
Mongolian University of Science and Technology, 8th khoroo, Baga toiruu 34, Sukhbaatar district Ulaanbaatar, Mongolia 14191

Naranbaatar Erdenesuren
Mongolian University of Science and Technology, 8th khoroo, Baga toiruu 34, Sukhbaatar district Ulaanbaatar, Mongolia 14191

Hiroaki Wagatsuma
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Email: purevdorj.choisuren544@mail.kyutech.jp, ge.yiqian503@mail.kyutech.jp, kasai.shintaro660@mail.kyutech.jp, dondogjamts.b@must.edu.mn, denaranbaatar@must.edu.mn, waga@brain.kyutech.ac.jp

Abstract

In general, prosthetic knee users have a large stress in the locomotion due to less smoothness and unnecessary energy consumption. In the passive prosthesis, it is difficult to regulate the stiffness depending on the ground contact force. In consideration of designs for such an adaptivity to improve passive systems, we propose an artificial knee kinematics design to absorb the redundant contact force for the smooth and stable walking and explore necessary constraints for the proposed mechanics to be able to have multi-functions not only for walking but also knee flexion accumulating the power for jumping. In the analysis, we used Multibody Dynamics (MBD) to investigate. This result will contribute to design an integrated dynamic model by incorporating a flexible body and ground contact forces in various purposes.

Keywords: Multibody dynamics (MBD), Contact force model, Prosthetic knee, Ordinary differential equation

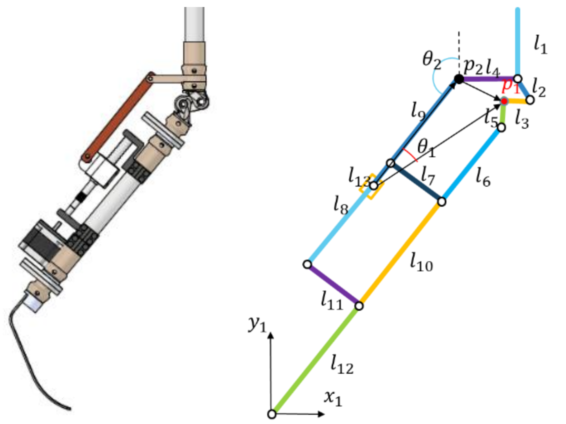
1. Introduction

Knee prosthesis users have high energy expenditure and motion load due to their different gait characteristics [1]. Therefore, developing knee prosthesis designs is an important research area, especially for lower limb amputation patients, to improve their mobility and quality of life. The knee joint is crucial in improving biomechanical efficiency and is a key component in supporting and transmitting load during walking, running, and stair climbing [2], [3]. However, prosthetic users face difficulties in replicating the motion of a real human knee, which leads to increased physical stress and energy consumption, especially during high-impact activities. Active prostheses use actuators and control systems to support joint movement and improve gait kinematics, but they suffer from weight gain, metabolic cost, and energy expenditure [4]. Passive prostheses cannot actively generate knee joint rotation, so above-knee amputees have more difficulty climbing stairs and ramps, and transitioning between sitting and standing, than non-amputees [5]. Therefore, there is a need to improve prosthetic technology to meet the unmet needs of people with above-knee amputees. To consider an adaptive model to improve the passive system, we propose a kinematic design of an artificial knee that incorporates additional hardware to support a smooth and stable gait and explore the constraints required to enable the proposed mechanism to perform multiple functions. Not only walking, but also knee flexion accumulates the force of jumping [6], [7]. In the analysis, we conducted research using Multibody Dynamics (MBD) [8], [9]. This result contributes to the development of a unified dynamic model by characterizing the rigid body interaction forces as a function of the motion constraint conditions [10]. In addition, modeling contact forces allows for detailed gait analysis by simulating the interactions between body parts and motion. By integrating these elements into a multibody dynamics (MBD) framework, we develop an analytical system for evaluating human biomechanics. The proposed approach contributes to the development of energy-efficient prosthetic devices designed to support natural, dynamic movement, increase user comfort and function, and address the limitations of current technologies.

2. Methodology

2.1. Knee prosthesis model

This paper proposes a kinematic design method of passive system prosthetic knee that adds a limited to the new knee joint to lock the knee joint for stable walking. The dynamic model kinematic mechanism proposal can be analyzed. However, it can be modeled by transferring the motion by making one more connecting mechanism as shown in Fig. 1. Considering the knee joint analysis model, the Prosthetic knee motion and trajectory can be analyzed using a simplified connecting mechanism as shown in Fig. 1 (b).



(a) (b)

Fig. 1. Prosthetic knee modified CAD model (a) and Prosthetic knee simplified joint mechanism (b).

2.2. Multibody system formulation for Knee link model

The knee joint angle is a critical factor in human walking, influencing motion and stability. To examine the relationship between the knee joint angle Fig. 1 (b) and the resulting joint motion and angular changes, multibody dynamics (MBD) [9], [10], [11], [12] was employed. Our approach is based on the previous work of Batbaatar & Wagatsuma (2021) [6], [10], In which authors used MBD to study the ground reaction forces during walking and running in a horse leg model. Inspired from their methodology, we analyzed the human walking process and implemented appropriate constraints in our system to identify the key factors affecting to the movement. For our prosthetic knee design, these constraints were adjusted to maintain mechanical stability while operating smoothly in a variety of activities. The mathematical model of the artificial knee joint mechanism is described using multibody dynamics (MBD), and the position and orientation of each body in the leg mechanism can be written as Eq.1 which is a 39-element vector called a generalized coordinate q.

$$q = [q_1^T, q_2^T, q_3^T, \dots, q_{13}^T]^T \tag{1}$$

The vector of q contains $q = R_N \times 3$ elements which are 13 rigid links and their center of position x_i, y_i and orientation ϕ_i are obtained in the generalized coordinates in the present analysis. Where the first 38 elements of the column matrix of kinematic constraint equation $\Phi^{K}(q)$ are derived from the absolute constraints between body and fixed ground node. The last elements in $\Phi_{(q,t)}$ defines the driver constraint of the proposed leg mechanism. Considering the initial configuration (Linear actuator), the driver constraint equation can be written as Eq.2.

$$\Phi^{D}_{(q,t)} = \begin{bmatrix} x_{13} \\ y_{13} \end{bmatrix} - \left(\begin{bmatrix} x_8 \\ y_8 \end{bmatrix} + A_8 \begin{bmatrix} -\frac{l_8}{4} cos(\omega t) \\ 0 \end{bmatrix} \right) \tag{2}$$

The differential-algebraic equations of motion for the knee stiffener model were effectively solved using the generalized acceleration matrix, allowing for the accurate computation of joint angles and angular velocities. These results were obtained through numerical integration employing the Runge-Kutta Gill's method [13].

The partial derivative of kinematic constraint equation with respect to the generalized absolute Cartesian coordinates q is Jacobian matrix Φ_q is obtained as Eq.3.

$$\Phi_q = \left[\frac{\partial \Phi(q,t)}{\partial q} \right]_{39\times 39} \tag{3}$$

Where it allows us to investigate placement, velocity and acceleration analyses kinematically. The forward dynamics analysis introduces the mass matrix $M = (39 \times 39)$ as Eq.4, and the generalized external force vector $h^{(a)} = (39 \times 1)$, as Eq.5.

$$M = diag(M_1, M_2, \dots, M_{13})$$
$$\{M_i = [m_i, m_i, J_i]^T | i = 1,2,\dots,13\} \tag{4}$$

$$h^{(a)} = \left[h_1^{(a)^T}, h_2^{(a)^T}, \dots, h_{13}^{(a)^T} \right]^T$$
$$\{h_i^{(a)} = [0, -m_i g, 0]^T | i = 1,2,\dots,13\} \tag{5}$$

Where m_i is the mass of rigid link to point, $J_i = 2l_i/3$ is the polar moment of inertia of rigid link to point i, and g is the gravitational acceleration. The equation motion of the system for the computer system analysis can be is expressed in general matrix form as Eq.6.

$$\begin{bmatrix} M & \Phi_q^T \\ \Phi_q & 0 \end{bmatrix} \begin{bmatrix} \ddot{q} \\ \lambda \end{bmatrix} = \begin{bmatrix} Q^A \\ \gamma \end{bmatrix} \tag{6}$$

This system of the equation can be solved for acceleration $\ddot{q}$, and Lagrange multipliers λ. In order to obtain coordinates q and velocities $\dot{q}$, acceleration is integrated at instant of time $t = t + \Delta t$. For the forward dynamic analysis, to ensure the numerical accuracy in the general solution of motion equation, constraint stabilization Baumgarte method is used with the following parameters α and β. In the forward dynamic analysis, new coordinates and velocities require two arrays for $\dot{q}$ and $\ddot{q}$ for the time step $t + \Delta t$ as Eq.7.

$$u = \begin{bmatrix} q \\ \dot{q} \end{bmatrix}, \quad \dot{u} = \begin{bmatrix} \dot{q} \\ \ddot{q} \end{bmatrix}, \quad \dot{u} \xrightarrow{yields} u(t + \Delta t), \qquad (7)$$

At the starting point of the numerical simulation, initial configurations of target mechanisms are given according to the primary operation in forward dynamics analysis.

3. Results & Discussion

Numerical analysis was performed using kinematic and dynamic simulations based on the above MBD formulation. This constraint allowed us to calculate the foot motion and visualize the time evolution of the ground reaction force within the knee mechanism.
All simulations and computational experiments were performed using MATLAB, which is perfectly suited for implementing the MBD framework and analyzing the dynamic behavior of the system.

Table 1. Parameters used in numerical simulations.

Kinematic/Dynamic analysis		
Gravitational acceleration [m/s^2]	g	9.81
The velocity of the driving crank [rad/s]	ω	$\pi/3$
Total simulation time [s]	t	$0 \leq t \leq 12$
Baumgarte parameter	α	0.8
Baumgarte parameter	β	35
Time step [s]	dt	1.0×10^{-3}

Summarizes the numerical simulation parameters used to analyze the contact forces in the foot mechanism (Table 1). The MBD differential-algebraic equations of motion for the knee rigid-joint model were successfully solved using the generalized acceleration matrix, and the angles and angular velocities were calculated by numerical integration using the Runge-Kutta Gill's method [13].

Computer experiments were performed using the MATLAB program.

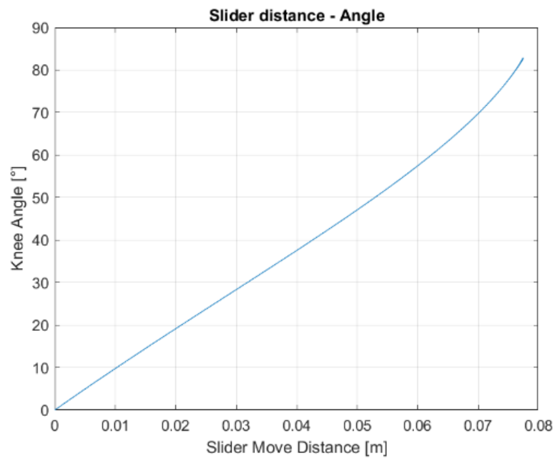


Fig. 2. Slider distance Knee Angle

The plot in Fig. 2 illustrates the relationship between the slider's displacement and the knee angle expressed in degrees. This relationship is nonlinear, indicating that even small changes in the sliding distance can lead to changes in the knee angle. This diagram highlights the sensitivity of the knee joint to sliding motion. This relationship is essential for applications such as robotics or biomechanical simulations, where precise control of the joint angle is important.

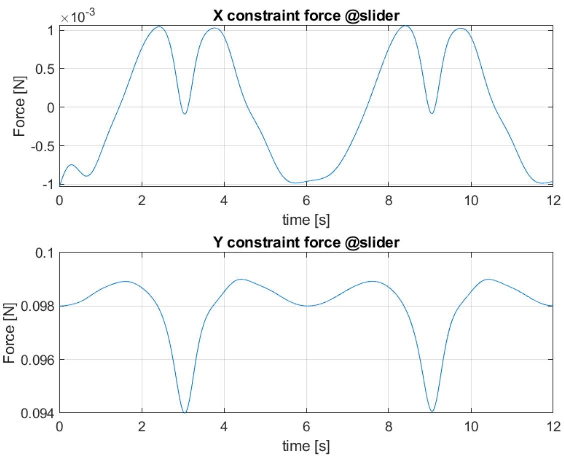


Fig. 3. X and Y Constraint Forces at the Slider

Fig.3 shows the forces (X and Y) acting on the slider over time due to mechanical restraints. This motion is clearly oscillating with the slider's cycle of motion.

This oscillation represents the dynamic interaction between the slider and the other components of the system.

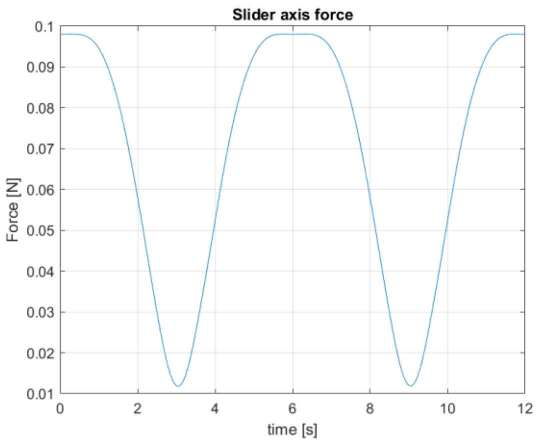


Fig. 4. Slider Axis Force Over Time

Fig.4 shows the net force along the axis of motion of the slider, which is the sum of the forces in the X and Y directions in the local coordinate system of the slider. The force varies over time in accordance with the cycle of the slider's motion. Analyzing the net force along the axis of the slider is essential to ensure that the system operates within safe force limits. It also provides insight into the energy transfer and efficiency of the mechanism. The relatively small forces acting on the slider and associated components, as shown in Fig. 1 indicate that the system is designed for precise use beyond the mass-weight relationship. Also, the strong dependence of the knee angle and the slider displacement, shown in Fig. 2 indicates the need for precise control of the slider motion. The dynamic forces acting on the various parts of the system demonstrate the complexity of the interactions between the components, requiring careful calculations for robust constraints and load analysis.

4. Conclusion

The critical kinematic and kinetic factors were successfully identified using MBD (Multibody Dynamics) analysis for the knee joint model. In addition, the motion constraints of the system were clarified by identifying the motion constraints, especially the critical positions that limit the motion. The high sensitivity of the knee joint behavior affects the motion of the slider. This combined analysis of MBD and MATLAB simulations provides valuable insights for the design and optimization of mechanical systems in robotics, biomechanics, and other fields where precise motion control and load management are required. In the future this result will contribute to design an integrated dynamic model by incorporating a flexible body and ground contact forces in various purposes.

Acknowledgments

This work was supported in part by JSPS KAKENHI (JP17H06383, JP24K07387), Grant-in-Aid for JSPS Fellows (23KJ1754) and the joint research project (J23A16) supported by Mongolia-Japan higher Engineering Education Development (MJEED-JICA).

References

1. L. Zhang, G. Liu, B. Han, Z. Wang, Y. Yan, J. Ma and P. Wei, “Knee Joint Biomechanics in Physiological Conditions and How Pathologies Can Affect It: A Systematic Review,” Applied Bionics and Biomechanics, 2020, Article Number 7451683.
2. S.D. Masouros, A.M.J. Bull, A.A. Amis Biomechanics of the knee joint. https://doi.org/10.1016/j.mporth.2010.03.005
3. Wei Liang, Zhihui Qian, Wei Chen, Hounan Song, Yu Cao, Guowu Wei, Lei Ren, Kunyang Wang and Luquan Ren, Mechanical and component design of prosthetic knees: A review from a biomechanical function perpective. https://doi.org/10.3389/fbioe.2022.950110
4. Minh Tran, Lukas Gabert, Sarah Hood and Tommaso Lenzi, A lightweight robotic leg prosthesis replicating the biomechanics of the knee, ankle, and toe joint doi: 10.1126/scirobotics.abo39
5. Thomas Schemalz, Siegmar Blumentritt and Rolf Jarasch, Energy expenditure and biomechanical characteristics of lower limb amputee gait:: The influence of prosthetic alignment and different prosthetic components https://doi.org/10.1016/S0966-6362(02)00008-5
6. D. Batbaatar and H. Wagatsuma, A Viscoelastic Contact Analysis of the Ground Reaction Force Differentiation in Walking and Running Gaits Realized in the Simplified Horse Leg Model Focusing on the Hoof-Ground Interaction, Journal of Robotics, Networking and Artificial Life, Vol. 8(2); September (2021), pp. 78-84.
7. K. Komoda and H. Wagatsuma, Energy-efficacy comparisons and multibody dynamics analyses of legged robots with different closed-loop mechanisms, Multibody System Dynamics 40, 2017, pp. 123–153.
8. P. E. Nikravesh, Planar Multibody Dynamics: Formulation, Programming with MATLAB, and Applications, 2nd edn., CRC Press, Boca Raton, 2018.
9. D. Batbaatar and H. Wagatsuma, A Proposal of the Kinematic Model of the Horse Leg Musculoskeletal System by Using Closed Linkages, Proceedings of the 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO), Dali, China, 2019, pp. 869–874.
10. A. A. Shabana, "Computer Implemantation of the Absolute Nodal Coordinate Formulation for Flexible Multibody Dynamics," Nonlinear Dynamics, vol. 16, pp. 293-306, 1998.
11. L. G. Maqueda, A. A. Mohamed and A. A. Shabana, "Use of General Nonlinear Material Models in Beam Problems: Application to Belts and Rubber Chains," Journal of Computational and Nonlinear Dynamics, vol. 5, pp. 21003-1-21003-10, 2010.
12. Flores, P., Machado, M., Silva, M.T. et al. On the continuous contact force models for soft materials in multibody dynamics. Multibody Syst Dyn 25, 357–375 (2011). https://doi.org/10.1007/s11044-010-9237-4
13. Wolfram Research, Inc., Runge-Kutta Gill’s method, https://mathworld.wolfram.com/GillsMethod.html

Authors Introduction

Mr. Choisuren Purevdorj

He received his master's degree in mechanical engineering from Mongolian University of Science and Technology (MUST), Mongolia in 2008. He is currently a doctoral course student in Graduate School of Life Science and Systems Engineering in Kyushu Institute of Technology, Japan.

Mr. Yiqian GE

He received his Bachelor's degree in Engineering in 2020 from the school of Mechanical Engineering, Hubei University of Technology in China. He is currently a master student in Kyushu Institute of Technology, Japan.

Mr. Shintaro Kasai

He received his Master's degree in Engineering in 2023 from the Graduate School of Life Science and System Engineering, Kyushu Institute of technology (Kyutech) in Japan. He is currently a doctoral course student in Kyutech, Japan and JSPS Research Fellow (DC1).

Dr. Dondogjamts Batbaatar

He received his M.S. in the field of mechatronics from Mongolian University of Science and Technology (MUST), Mongolia and Ph.D. degree from Kyushu Institute of Technology, Japan in 2015 and 2021. He is currently a vice dean of research affair in School of Mechanical Engineering and Transportation at MUST. His research interests include computational non-linear dynamics and bio-inspired robotics.

Dr. Naranbaatar Erdenesuren

He received his M.S. degree in IT/Mechatronics and Ph.D. degree in Mechanical and Automotive Engineering from University of Ulsan, South Korea in 2008 and 2013, respectively. He is currently the associate professor at the School of Mechanical Engineering and Transportation, MUST. His research interests include Mechatronics, Robotics and AI

Dr. Hiroaki Wagatsuma

He received his M.S., and Ph.D. degrees from Tokyo Denki University, Japan, in 1997 and 2005, respectively. In 2009, he joined Kyushu Institute of Technology, where he is currently a Professor of the Department of Human Intelligence Systems. His research interests include non-linear dynamics and robotics. He is a member of IEEE.

Suppressing of Multi-Axial Vibration Caused in Carried Objects by Robot Using a Heuristic Algorithm Based on Evaluation Actual Machine Information

Yusuke Ueno
Faculty of Production Systems Engineering and Sciences, Komatsu University,
Nu 1-3 Shicho-machi, Komatsu-shi, Ishikawa,923-0971, Japan
Email: yusuke.ueno@komatsu-u.ac.jp

Fumitoshi Shimono
Faculty of Production Systems Engineering and Sciences, Komatsu University,
Nu 1-3 Shicho-machi, Komatsu-shi, Ishikawa,923-0971, Japan

Ryu Shinmura, Hiroki Noguchi
Division of Production System Science, Komatsu University,
Nu 1-3 Shicho-machi, Komatsu-shi, Ishikawa,923-0971, Japan

Hiroshi Tachiya
Advanced Mobility Research Institute, Kanazawa University, Kakuma-machi, Kanazawa-shi, Ishikawa, 920-1192, Japan

Abstract

Residual vibrations induced in objects carried by robots cause the positioning accuracy to deteriorate, which makes the installation of carried objects difficult. This study proposes a method for determining a trajectory that can suppress residual vibration using a heuristic algorithm, based on the behavior of a carried object, which is measured by actually operating a robot. Trajectories were generated for a commercially available industrial robot to suppress residual vibrations in two axial directions for a pendulum-shaped object to be transported. By optimizing the path shape and acceleration/deceleration, a trajectory that reduces the amplitude of residual vibration in each axis direction by 70% was successfully generated in a few actual machine operations.

Keywords: Robot arm, Residual vibration, Trajectory planning, Heuristic algorithm

1. Introduction

When an industrial robot transports an object at high speed, residual vibrations are likely to occur in the transported object after the robot has stopped. Residual vibrations in the carried object cause a deterioration in positioning accuracy and an increase in working time. Various methods for suppressing vibration by adjusting the trajectory, which is the change in the position and posture of the robots has been proposed [1], [2], [3]. However, most of them consider only the vibration characteristics of the robot itself and not the object to be transported. They also require an accurate kinetic analysis of the vibration control target. It is not easy to construct precise kinematic models for kinetic analysis for commercially available robots and transport objects, and it is difficult to carry out kinetic analysis [4], [5]. Therefore, this study proposes a method for generating trajectories that suppresses residual vibrations by a carried object without using kinetic analysis by optimizing the trajectory using a heuristic algorithm based on the evaluation value measured by actually operating the robot. In the present study, the residual vibration in two-axial directions caused by the object to be transported is targeted, and trajectories that reduce the vibration amplitude of the object are actually generated using a commercially available industrial robot to demonstrate the usefulness of the proposed method.

2. Outline of method for suppression residual vibration

2.1. *Variables to adjust*

Fig. 1 shows the outline of method for suppressing the residual vibration in two-axial direction. The output point of the robot with the carried object is moved from position S to the end position E. Normally, a trajectory prepared by the manufacturer is used for the movement of the load to be transported, with a straight line interpolated from position S to position E. In this study, this trajectory is defined as the standard trajectory. When two-axial vibration occurs in the carried object at position S, the vibration remains after the object is moved along the standard trajectory. In the proposed method, in order to suppress the residual vibration, new target positions P_1 and

P_2 are set in the horizontal plane immediately before the end position shown in the figure, and the output point in moved in the order of positions S, P_1, P_2 and E. By adjusting these target positions and the time required for movement between them, appropriate acceleration and deceleration is applied to the carried object in both directions perpendicular and parallel to the path of standard trajectory, thereby suppressing residual vibration in the two-axial direction. In this paper, position P_1 is placed on the path of standard trajectory and only position P_2 is placed outside the path. The operation displacement between each target position in the direction parallel to the path of standard trajectory and the operation displacement perpendicular to the path are defined as l_1, l_2, l_3 and e, respectively. The operation time between each target position is defined as t_1, t_2 and t_3. Here, if there are no constraints on the operation times t_1, t_2 and t_3, residual vibrations are suppressed but may increase the transport time. Therefore, the values of the aforementioned variables are adjusted such that the output point has the same transport time as the standard trajectory. Thus, these variables satisfy the following Eq. (1), where the operation displacement and target velocity measured in the standard trajectory are defined as L and V, respectively:

$$t_1 + t_2 + t_3 = \frac{L}{V} \tag{1}$$

L and V were obtained from the standard trajectory measured by actually operating the robot. Thus, by determining t_1 and t_2, t_3 can be subordinately obtained from Eq. (1). In addition, since the end position E is fixed, l_3 is also obtained subordinately obtained by determining l_1, l_2 and e. In trajectory generation, the residual vibration measured by the actual operation of the robot was evaluated, and the optimal values of five variables (l_1, t_1, l_2, t_2, e) were determined using the heuristic algorithm described in next section.

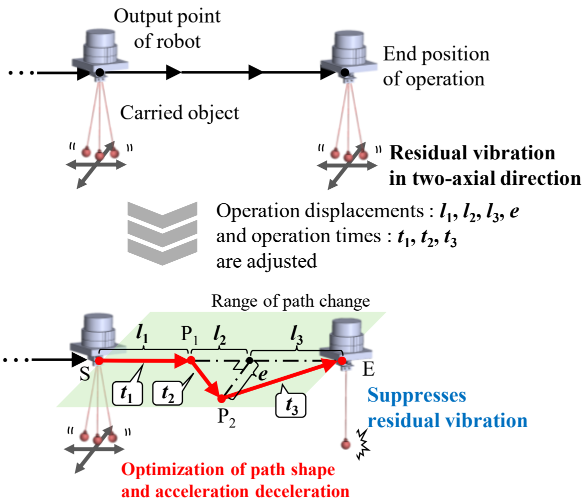


Fig. 1. Outline of method for suppressing the residual vibration.

2.2. *Heuristic Algorithm "SHA"*

Trajectory generation by the heuristic algorithm iteratively updates the trajectory based on evaluation of measurements obtained by operating the robot. Lin's heuristic algorithm (SHA) [6], [7], proposed by Lin et al. and confirmed by the authors in the trajectory generation of a robot manipulator [8], is used to update the trajectory. SHA is a type of heuristic algorithm which searches for the best combination of values of several variables in a short time so that set evaluation value is the best. The SHA expresses the design parameters in the variable matrix, as shown in Fig. 2. Each column from the first column to column n in the variable matrix represents a type of variable to be searched, and each element from the first row to row k in each column represents the discrete values taken by the variables. As indicated by the black circle in Fig. 2, The combination of discrete values determined by selecting elements from each column of the variable matrix one by one. In trajectory generation, one discrete value combination represents the robot's trajectory, as (l_1, t_1, l_2, t_2, e) is placed in the variables of the variable matrix. In SHA, multiple trajectories are generated by randomly selecting elements called look-ahead-base-points (LBPs) in a certain column, as shown by the red circles in Fig. 2. By repeatedly performing comparison of the evaluation values measured by operating the robot for each generated trajectory, the trajectory with the best evaluation value, i.e., the optimal trajectory, is determined.

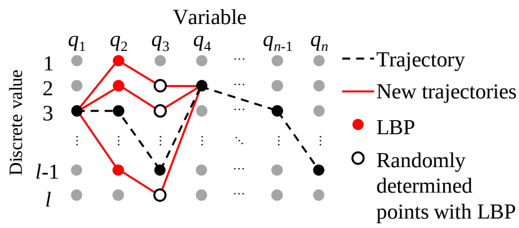


Fig. 2. Variable matrix.

3. Target Robot Arm and Carried Object

The method described above was applied to a commercial 6-DOF vertically articulated robot (Kawasaki Heavy Industries, Ltd., RS020N). The robot is shown in Fig. 3. As shown in the figure, the joints of the robot were designated as JT1 to JT6 from the base side of the robot. The motion and trajectory of the output point P of the robot are represented by the absolute coordinate system O-*XYZ*, in which the origin is the center of JT1 and the vertical direction is the *Z*-axis. Operational commands are given using the company's AS language, which is based on the Standard Language for Industrial Manipulators (SLIM) language. In this paper, trajectory generation is performed for residual vibration in elastic bodies. Therefore, a pendulum with a natural frequency of 1.5 Hz was used as shown in Fig. 4. The carried object was attached to the output point of the robot via the six-axis force sensor

(Leptrino CFS034CA301A). The residual vibrations induced by the motion of the robot are measured by measuring the reaction moment generated at the fixed end using the force sensor.

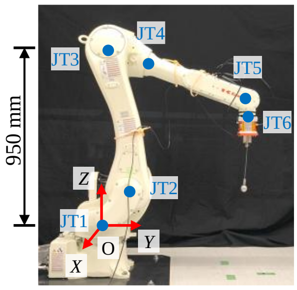


Fig. 3. Industrial manipulator RS020N.

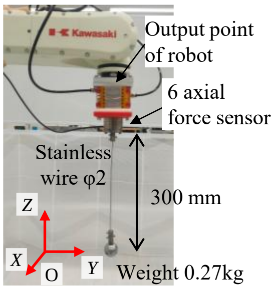


Fig. 4. Carried object.

4. Condition for Searching the Trajectory by SHA

Fig. 5 shows the path targeted by trajectory generation. As shown in the figure, the path S-E of the standard trajectory was a straight line parallel to the *Y*-axis direction with a length of 600 mm (L = 600 mm). In trajectory generation, the optimal trajectory was searched by repeatedly changing the trajectories within the operation range indicated by the green rectangle in the figure and evaluating the residual vibrations induced in the carried object at position E. To prevent breakdowns of the carried object, from preliminary experiments, the acceleration/deceleration of the output point at trajectory S-E was set to 20% and the operation range to ±150 mm in the *X*-axis direction. Before starting the motion at position S, the output point is moved in the order of positions A, B and S to apply initial vibration. To increase the amplitude of the initial vibration, in paths A-B and B-S by trial and error, the velocities were set to 10% and 5%, and the acceleration / deceleration were set to 30% and 15%, respectively. The output point was allowed to wait for 0.7 and 1.32 s at positions B and S. The output point with the carried object was moved while maintaining a constant attitude. During the search, the output point was returned to position S for the next operation after the evaluation of the residual vibration was completed. The vibrations induced during this process were suppressed by the plate shown in the bottom right-hand corner of the figure.

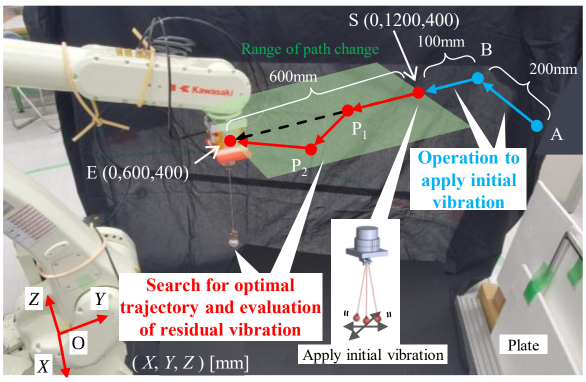


Fig. 5. Path of the robot.

The target velocity V of the standard trajectory was measured by moving the output point in a straight line from positions S to E at the maximum target velocity that the robot could exert. This resulted in V = 953 mm/s and L/V in Eq. (1) was 0.62 s. The evaluation value in SHA was the sum of the standard deviations of the moments about the *X*-axis and *Y*-axis at the fixed end of the carried object from 1 to 6 seconds after the end of the operation at position E. The sampling period of the moments was 32 ms. The number of elements in each row of the matrix and the number of LBPs were set to 100 and 3, respectively.

5. Result of the Trajectory Generation

Fig. 6 shows the changes in the evaluation value obtained during the search for the trajectory by the SHA. The red squares indicate the best value among the evaluation values obtained up to that operation. The dotted lines indicate the evaluation values obtained for the standard trajectory. The best evaluation value decreased as the search progressed. It generated 20 trajectories with evaluation values reduced by more than 70% compared to values obtained for standard trajectory. The trajectory with the smallest evaluation value, namely the optimal trajectory was obtained at 37th. The reduction in the evaluation value of the optimal trajectory was 78.2%. The search test required approximately 50 min in all 76 robot operations.

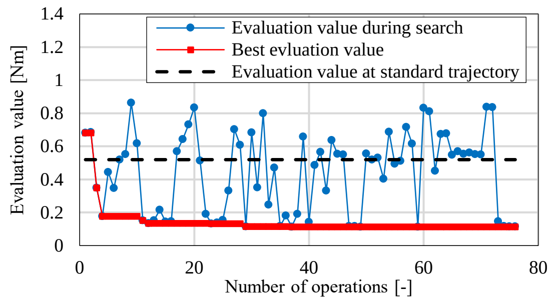


Fig. 6. Change in the evaluation value with the number of operations during a search by the SHA.

Fig. 7 shows the change in the velocity of the output point and *X*-axis moment, *Y*-axis moment with time for the optimal trajectory. As shown figure, the output point stopped at a time of about 6.5 s for both the optimal and standard trajectories. The waveform after that time represents the residual vibration caused in the carried object. In the optimal trajectory, the moments in both the *X*-axis and *Y*-axis decreased remarkably at the end of the operation in contrast to the standard trajectory. The amplitudes of the moment about both axes were reduced by more than 70% in comparison to that of the standard trajectory.

Fig. 8 shows the paths of the output point and the weight of the carried object in optimal and standard trajectories. The blue circle in the figure indicates the position of the weight. Here, the positions of the weight were determined by calculating the proportionality constant between the moment and the displacement of the weight from a static load test on the carried objects. As shown in the figure, the initial vibration with large amplitudes occurs in both axial directions at position S. In the standard trajectory, the output point is only moved in a straight line, resulting in a residual vibration at position E with an amplitude equivalent to the initial vibration. On the other hand, in the optimal trajectory, the output point is meandered to position P_2 so that the displacement direction of the weight matches that of path P_2-E. This changes the two-axial vibration occurring in the carried object to a one-axial vibration along the motion in path P_2-E. The residual vibration is suppressed by accelerating the output point in the same direction as the weight is followed.

As described above, in the optimal trajectory obtained by SHA, the two-axial vibration is changed to one-axial vibration by adjusting the path, and the subsequent acceleration motion applies vibration in the opposite phase to the one-axal vibration. Thus, the vibrations cancel each other out and the vibration amplitude.

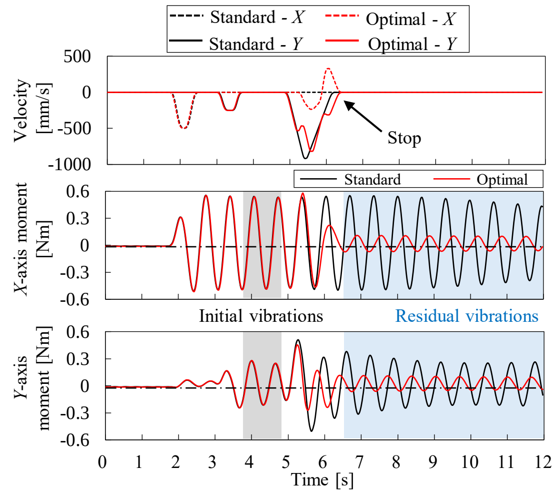


Fig. 7. Changes in the output point velocity and moment with time.

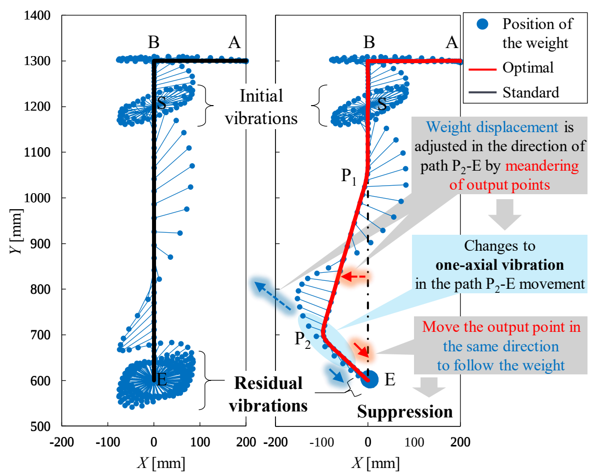


Fig. 8. Paths of the output point and the weight of the carried object for optimal and standard trajectories.

6. Conclusion

In this study, we propose a method for suppressing the residual vibration in the two-axial direction caused by a robot in a carried object using the heuristic algorithm SHA. In the proposed method, because the residual vibration is measured by actually operating the robot is evaluated, a vibration suppression trajectory can be generated without requiring kinetic analysis of the robot or the carried object. Trajectory generation was performed by adjusting the path and acceleration / deceleration using a commercial industrial robot with a pendulum as the carried object. The result was a trajectory in which the amplitude of the residual vibration was reduced by more than 70% compared to the amplitude in the manufacturer's standard trajectory in both axial directions. For carried objects with same specifications, the proposed method can generate trajectories with a sufficiently small number of robot operations, approximately 40 times. However, when the specifications often change, the time required for trajectory search becomes unacceptable. Therefore, in the future, we will plan to construct a method for generating trajectories in a real-time by training a neural networks using obtained optimal trajectories for various carried objects by SHA.

References

1. K. T. Dan, S. K. Rune, B. Ole and Z. Xuping, Vibration control in industrial robot arms by multi-mode varying input shaping, Mechanism and Machine Theory, Vol.155, 2021, 104072.
2. H. C. Pyung and S. P. Hyung, Time-varying input shaping technique applied to vibration reduction of an industrial robot, Control Engineering Practice, Vol. 13, No. 1, 2005, pp. 121-130.
3. N. Michael, L. Kaiyue and K. Matt, Suppression of robot vibrations using input shaping and learning-based structural models, Journal of Intelligent Material Systems and Structures, Vol. 32, No. 9, 2021, pp. 1001-1012.
4. A. Klimchik, B. Furet, S. Caro and A. Pashkevich, Identification of the manipulator stiffness model parameters in industrial environment, Mechanism and Machine Theory, Vol. 90, 2015, pp. 1-22.
5. A. Montazeri, C. West, S. D. Monk and C. J. Taylor, Dynamic modelling and parameter estimation of a hydraulic robot manipulator using a multi-objective genetic algorithm, International Journal of Control, Vol. 90, No. 4, 2017, pp. 661-683.
6. S. Lin and B. W. Kernighan, An effective heuristic algorithm for the Traveling-Salesman Problem, Operations Research, Vol. 21, No. 2, 1973, pp. 498-516.
7. T. W. Lee, F. Freudenstein, Heuristic combinatorial optimization in the kinematic design of mechanisms part 1: theory, Journal of Manufacturing Science and Engineering, Vol. 98, No. 4, 1976, pp. 1277-1280.
8. Y. Ueno and H. Tachiya, Suppressing residual vibration caused in objects carried by robots using a heuristic algorithm, Precision Engineering, Vol. 80, 2023, pp. 1-9.

Authors Introduction

Dr. Yusuke Ueno

He received M. S. of Engineering in 2017 from Kanazawa University in Japan. He received D. Eng. in 2020 from Kanazawa University in Japan. Currently, he is assistant professor in Faculty of Production System Engineering and Sciences, Komatsu University in Japan. His research interests are focused on robotics, machine tools.

Mr. Fumitoshi Shimono

He is currently pursuing Bachelor of Engineering as final year student in the Faculty of Production System Engineering and Sciences, Komatsu University in Japan.

Mr. Ryu Shinmura

He received his Bachelor's degree in Engineering in 2023 from Faculty of Production System Engineering and Sciences, Komatsu Univesity in Japan. He is currently a master student in Graduate School of Sustainable Systems Science, Komatsu Univesity in Japan.

Mr. Hiroki Noguchi

He received M. S. from Engineering in 2024 from Graduate School of Sustainable Systems Science, Komatsu University in Japan.

Dr. Hiroshi Tachiya

He received M. S. in 1987 from Tokyo Institute of Technology in Japan. In 1989, he joined Kanazawa University in Japan. He received D. Eng. in 1997 from Tokyo Institute of Technology in Japan. Currently, he is a professor at the Advanced Mobility Research Institute in Kanazawa University in Japan. His research interests are focused on intelligent tires, robotics and machine tools.

Developing Low-Cost BCI-Based Brain-Limb Interaction Device with Prosthetic Hand

Nethika Jayith Rajapakse
International Bilingual School at Tainan Science Park, No. 1, Ln. 888, Xilaya Blvd., Xinshi Dist., Tainan City 744094, Taiwan (R.O.C.)
Email: ibst311207@nnkieh.tn.edu.tw

Abstract

Inventions that are designed to heal the body and/or mind should always be sought. Currently, too many people are lacking access to cheap prosthetic devices, especially those that allow neural connections to be gained with such devices. This is the reason why this study intends to propose a low-cost invention that is able to enable people to gain back concentration with limb use using Python Programming language, an EEG Neurosky headset, an Arduino, and a 3D-printed prosthetic hand. With our implementation, the proposed method can explore the boundaries of improving attention and continue to develop higher-level BCIs for the mind-limb connection, improving on current-day prosthetics and concentration/limb connection rehabilitation

Keywords: Low-cost prosthetic, BCI, Brain-limp, Interaction, 3D-printed prosthetic

1. Introduction

Technology to help is the greatest form of technology. And the biggest help that can be done is for the medical industry. Prosthetics will always be in demand. It is always good to develop solutions for human augmentation because there will always be people who need that technology. The technology surrounding the Brain-Computer Interface is still quite new. Thus, it is still very much an emerging field that should be actively pursued. Prosthetics are quite the engineering marvel. Prosthetics often require having multiple systems to work, especially precise prosthetics where the brain acts as the controller for the prosthetic. These kinds of robots must require handling data transmission, and building robot mechanisms, and must include the use of integrated circuits. Through the development of the prototype, it also was quite clear that the prosthetic could be used as some sort of attention rehabilitation. Because the prosthetic movement is based on the user's attention, it was apparent that this robot could be used as an interactive game to improve the attention of the user.

In this present study, a low-cost BCI-based interactive device is proposed to be used for attention rehabilitation for the user. This proposal is meant to be a prototype for further BCI prosthetic development.

2. Background

2.1. *Prosthetics and their Cost*

The cost of prosthetic limbs is quite steep. Osseointegrated prosthesis are prosthetics that are attached to the bone instead of the traditional socket-suspended prosthesis, which uses a socket on the amputated limb. Furthermore, socket prosthetics have introduced more microprocessor implementation, which leads to much easier use and better quality of life for the user [1].

But in the case of both types of prosthetics, the process of getting a prosthetic is quite expensive. In the case of prosthetics, the average healthcare cost in a 12-month post-amputation is 99,409 USD. An earlier recipient would have approximately 25% lower total direct healthcare cost compared to people to get their prosthetics later. This shows that not only are prosthetics causing a lot of cost in total healthcare costs (rehabilitation, etc.), but also that earlier access can reduce cost and help the patient [2].

This shows that a major goal in the rehabilitation field should be to improve the cost for better accessibility and work on solutions that make prosthetics faster to make and easier to use. Making prosthetics faster to make also reduces wait time, allowing for the patient to quickly recover and decrease the cost through faster access.

2.2. *Attention at a Decline*

As technology advances, it affects our psychological limits. The idea of shrinking attention spans is gaining a lot of traction in the media and in research. But one thing is certain: digital distractions can cause our attention spans to shorten. Companies have also started to investigate improving the attention spans of their employees [3].

Children are also extremely affected by the digital age, and so there is documented an increase in attention problems from 1983 to 2017. A possible cause for this effect is that children usually spend more time on the screen multitasking which can really affect their attention regulation skills [4].

Either way, the idea of increasing attention spans is a highly researched topic. This is definitely something that has a demand in our society, as more and more people give importance to their attention spans.

2.3. *Neurosky Mindwave Mobile*

Fig.1 The Neurosky Mindwave Mobile 2, the Headset used in this paper [5].

The Neurosky Mindwave Mobile 2, as seen in Fig. 1, is an optimal EEG headset because it is cheap considering the EEG headset market. More importantly, if it is possible to make a product out of the Neurosky Mindwave Mobile 2, then it means that there is a chance to develop a cheap BCI prosthetic, which would disrupt the prosthetic market. Most professional EEG headsets are used for medical purposes, thus aren't easily accessible to the public. However the Neurosky headsets are mostly used for commercial purposes. These headsets are used for small games, or for small research, thus it is one of the best candidates for this interactive device if this was to be commercial.

The Neurosky headset has a specific set of parameters which make it easy to interpret brainwave data. These parameters are *Attention, Meditation, RawValue, Delta, Theta, LowAlpha, HighAlpha, LowBeta, HighBeta, LowGamma, MidGamma, PoorSignal*. The Attention and Meditation parameters are both defined on the NeuroSky website as an *eSense* Meter, in which their own parser, ThinkGear Connector, does the calculations from the headset's brainwave data, allowing the user to know their attention and meditation levels without having to program a separate algorithm to calculate the attention or meditation of the user. Although the brainwave data is provided, for them to be used effectively, a deep learning algorithm would be needed to analyse and find patterns in the data. The *RawValue* is a data value of the raw wave sample, consisting of 2 bytes, one byte for the lowest value and the other byte for the highest value. This is the raw wave value before it is converted into individual brainwave values. The *PoorSignal* is a value that represents the headset signal strength. The *RawValue* and *PoorSignal* are mainly used for debugging purposes.

There are 2 ways of extracting data from the headset using Python, Serial communication, and a Telnet connection. The Serial communication bypasses Neruosky's own connector/parser, ThinkGear Connector. So if the code were to establish a serial connection, then it would also have to parse the data as well. The serial connection connects to the headset by sending it the connection byte, and so the headset connects to the program without any help from the ThinkGear Connector. This is a less restrictive way of gaining the data and allows for the user to force connect to the headset without any issue encountered in previous testing. Not only this but Serial connection allows for better control of the data, but that comes with the cost of inaccuracy because the parser has to be custom written.

The Telnet connection connects to the ThinkGear Connector. So, for the telnet connection to work, the headset must already be connected to the ThinkGear Connector. Then the ThinkGear Connector will parse the data from the headset, and the telnet connection allows for parse data to go through in a JSON dump. This allows for more accurate data but is harder to run as previous testing showed that the ThinkGear Connector was unstable.

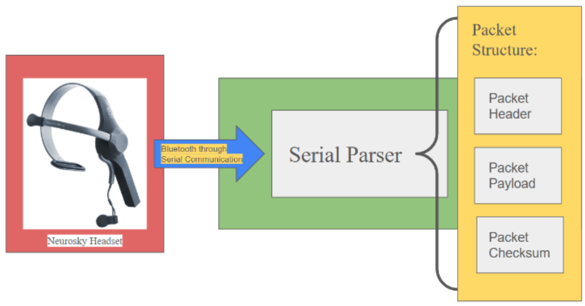


Fig.2 Serial Parser on the EEG Headset data.

3. Methodology

3.1. *NeuroSky Data over Serial Communication*

Many other people had used a Serial connection. So it was also decided to implement this method as it is easy to use and makes for the most direct and straightforward path to connecting to the headset. Other ways, like using a lab streaming layer, or using C or C# code, along with NeuroSky's own SDK was either not stable or not reliable. The NeuroSky Mindwave Mobile 2 has a very unique packet protocol, which consists of a header, payload and a checksum, as seen in Fig. 2. The Packet Header defines the packet's status code. It is very important to pay attention to this header, as this header is used to show what the payload is in reference to which parameter. For example, the code (0x04), which is used to label packets responsible for Attention. The packet payload contains the data for the specific parameter, which is saved into a local variable. The packet checksum to verify the packet before the process starts. As seen in Fig. 2, the headset would be force connected and start sending packets to the

Serial Parser. The first 2 packets are read and have been sorted and verified by its status code. The packet's payload is saved into the Serial Parser for later manipulation. The checksum is used to make sure that the packet was verified by that specific headset. The packet header defines where the payload information would go in the Serial Parser.

3.2. *Arduino data over Serial Communication*

Once sufficient data is collected, a new serial connection is made to communicate with the Arduino Uno [6]. For the prosthetic to correspond to the user's Attention data, a new serial connection has to be made, connecting the Arduino to the program. As seen in Fig. 3, a *DataManager* class is responsible for the Arduino receiving the data possible to have the prosthetic to move.

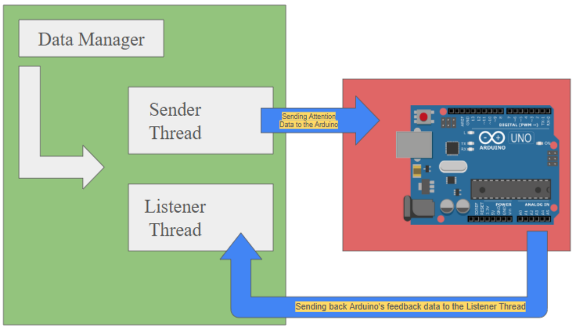


Fig. 3 Overall Serial Communication to the Arduino.

The *DataManager* would have 2 threads open, as shown in Fig. 3. The sender thread is responsible for sending the attention data to the Arduino. The listener thread is responsible for listening to the Arduino's feedback data.

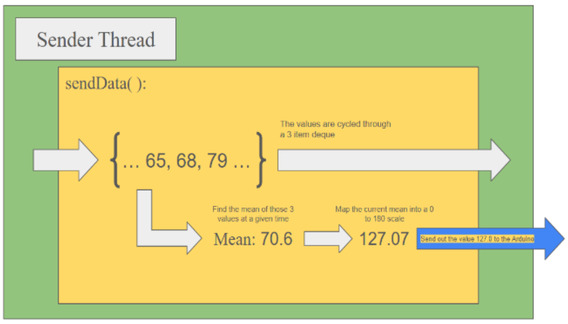


Fig. 4 Data Manipulation done in the Sender Thread.

The sender thread runs another target function called *sendData()*, which is shown in Fig. 4. The *sendData()* function keeps on sending a mean of attention values to the Arduino and rescales them to fit for the servos.

This *sendData()*, as seen in Fig. 4, fills a 3-item deque for data manipulation. The 3 data points are then converted to a mean which is rescaled to 0 ~ 180. For every new set of data, a new mean is calculated and rescaled. This rescaled mean is then sent to the Arduino for the servos to move. The rescaled mean acts as a target position for the servos to reach. This creates a correspondence between the user's attention and the prosthetic. The listener thread is used to verify the Arduino has gotten the data. It is also used to verify whether the data on both ends of Serial communication is the same. The listener thread is only used to check the feedback and debug the program as seen in Fig. 3.

3.3. *Prosthetic Controlled by Arduino*

Prosthetic is controlled by the Arduino end of the code. The Arduino opens up its Serial connection and connects to the main program. The Arduino will listen for the target position of the servos, based on an average rescaled value of the user's attention. The servo's position is based from a 0 to 180 angle.

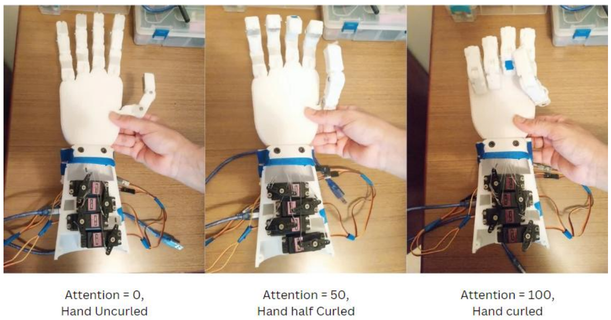


Fig. 5: Different Positions of the Hand based on the varying Attention level.

Thus, the prosthetic movement and the attention of the user is one to one, as shown in Fig. 5. This allows for more precise movement, instead of a simple Boolean condition, allowing for the prosthetic to be in a range of orientations from palm to fist.

3.4. *Mechanics of Prosthetic*

The prosthetic works off of 4 Tower Pro MG996R [7] Digital servo motors. Each servo has a wire that wraps around a whole finger as seen in Fig. 6. The wire runs through the palm and then goes through each segment of the finger before going back into the palm and then being tied to the same servo. This means that when the servo pulls clockwise, the finger curls into a fist. When the servo pulls counter-clockwise, the fingers extend into a palm. The reason why there are only 4 servos is because one of the servos has both the pinky and ring finger attached to it, shown in Fig. 6 (right image).

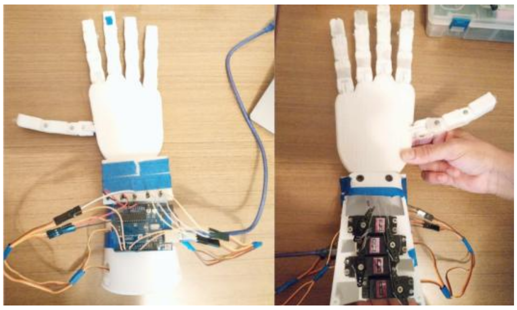


Fig. 6: The first image is of the back of the hand, the second image is of the front of the hand.

4. Results

The user takes the headset and after booting up the program, the data is already being streamed to the computer and then to the Arduino. A demonstration is shown in Fig. 8. Thus the user can increase or decrease their focus and cause the hand to curl up in the same interval that the data shows.

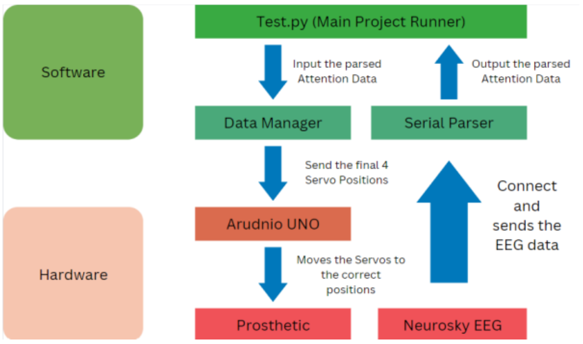


Fig. 7: Flowchart of the final prosthetic

As the user defocuses and focuses, the hard curls in a way that matches that same rate of attention given by the user due to the data manipulation done in the Data Manager in Software processing, as seen in Fig. 7 where the whole system is outlined.

Fig. 8: A Person wearing the headset and controlling the prosthetic

5. Reflections

5.1. *Future Technical Improvements*

An actual brain computer interface would use some sort of machine learning/deep learning to get accurate predictions of when the person is thinking of grasping or extending their fingers. EKG data can be involved in the analysis to allow for a completely synchronized prosthetic with the user's muscles and brain. All the extra analysis will allow for an actual prosthetic in which the user can think about the action they want to perform instead of focusing and focusing. For the future, a neural network would have to be implemented and with that better data would be needed as well. The neural network would need raw brain wave values to give it better pattern recognition. The headset could also be improved as well. Because the Neurosky Mindwave Mobile 2 only has 2 nodes, the data that it can collect isn't the best quality of data. The actual raw brain wave data and its parsing would also need to be improved. As for a neural network would need to run off of a lot more data and from more than 2 channels.

5.2. *Future Mechanical Improvements and Potential*

Because the prosthetic works by the wire pulling from one side to pulling the finger down or up, it isn't accurate at all. The mechanism must change if the user wants a high degree of accuracy in movement and high dexterity in the prosthetic. As it stands, the same positional data doesn't mean the same position on the prosthetic every time. This is because there are too many factors at play when moving the finger, primarily, the friction of the wire and the loosening of the wire over time.

Generation 2 of the prosthetic must have some sort of tension spring in the finger's joints. This would mean that without tension, the finger would spring back to the same place. If the same amount of tension was applied to the spring, then theoretically the finger would move to the same position proportionally to the tension. Thus, this makes for a much more dexterous prosthetic. This final product on the end of this development path would be the generation 3 of the prosthetic. Generation 3 would be a prosthetic with deep learning from both EEG and EMG data for a better prediction of what the user is thinking. From what action the user is thinking, the prosthetic will do that action from the prediction it makes from the EEG and EMG data. But as it stands this is a good proof of concept for a cheap Brain Computer Interface Prosthetic.

6. Conclusion

This prototype includes the communication ability between the headset and the computer. With this proof of concept, an experimental BCI prosthetic was developed to tie with the user's attention for interactive rehabilitation purposes. It should be noted that this result is definitely a prototype at best. As listed before in the problems, there are many sectors that we can improve on that can make for a much better system in the case of developing a full BCI prosthetic.

References

1. A. Amsan, A. Nasution, M. Ramlee,.A Short Review on the Cost, Design, Materials and Challenges of the Prosthetics Leg Development and Usage, International Conference of CELSciTech 2019 - Science and Technology track (ICCELST-ST), 2019
2. T. Miller, R. Paul, M. Forthofer, S. Wurdeman, "Impact of Time to Receipt of Prosthesis on Total Healthcare Costs 12 Months Postamputation." American journal of physical medicine & rehabilitation vol. 99, 11, 2020.
3. K. Subramanian, Myth and Mystery of Shrinking Attention Span.International Journal of Trend in Research and Development, Volume 5(3), 2018.
4. B. Vekety, H.N Logemann, J. Protzko, Z. Takacs, KIDS THESE DAYS!' A META-ANALYSIS OF CHANGES OF ATTENTION PROBLEMS IN REPRESENTATIVE SAMPLES OF CHILDREN, International Conference on Education and New Developments, 2022
5. https://store.neurosky.com/pages/mindwave
6. https://docs.arduino.cc/hardware/uno-rev3/
7. https://towerpro.com.tw/product/mg996r/

Authors Introduction

Mr. Nethika Jayith Rajapakse

He is a high school student. He is currently a driver coach at this School's FRC Robotics team. Their team has gotten into the World Finals in 2022 and 2023. He has also done nation wide science fairs, entering in his robots. He has a current interest in Neural Networks, Machine Learning, Robotics and Prosthetics, and Procedural Generation.

A Study on Local Airports Contributions to Tourism Industry in Japan

Tsutomu Ito*, Seigo Matsuno
Department of Business and Administration, Ube National College of Technology,
2-14-1 Tokiwadai, Ube, Yamaguchi, 755-0096, Japan
E-mail: t_ito@ube-k.ac.jp
** Corresponding author*

Makoto Sakamoto, Satoshi Ikeda
Faculty of Engineering, University of Miyazaki,
1-1 Gakuen Kibanadai-Nishi, Miyazaki, 889-2192, Japan

Takao Ito
Graduate School of advanced Science and Engineering, Hiroshima University,
1-4-1 Kagamiyama, Higashi-Hiroshima, 739-8527, Japan

Abstract

This study examines the impacts of local airports on the regional tourism industry in Japan. A series of indexes of the tourists and a new measure of the time index have been developed based on our four-cell model to analyze the correlation between different indexes such as the number of airport tourists and the time used by public transportation. During the verification process, a correlation has been discovered between the number of tourists in the city where the airport is located and the number of airport users for specific regionally managed airports. Furthermore, regarding access to tourist attractions from airports, a strong negative correlation is confirmed between the evaluation index and the state of development of public transport, suggesting the importance of developing public transport, including airports, in attracting tourists.

Keywords: Tourism industry, Local Tourism, local airport, public Transport

1. Introduction

In recent years, tourism industry in Japan is becoming a hot issue as one of the attractive policies to revitalize regional economy. According to publications from the Japan Tourism Agency of the MLIT [1], the tourism industry is expected to have an economic scale of 7.0 trillion yen, and it will employ 6.13 million people as a large industry in both economic and employment viewpoints. From a global perspective, as can be seen from documents from the World Tourism Organization [2], it is an industry to be expected to grow. In order to develop this industry efficiently, it is required to statistically analyze the behavior of tourists and promote the destination appropriately. On the other hand, tourists need a means of transportation to visit the destinations they are interested in, and the convenience and maintenance of such means of transportation is also expected to have a strong impact on tourists' plans. In Japan, the fastest and most convenient way to travel from overseas or urban areas to local tourists' destinations is by plane. Therefore, this study aims to make contribution to new insights into efforts to revitalize the tourism industry in regional areas by clarifying the impact that regional airports have on regional tourism, and the impact that the development of airport-based transportation networks has on the tourism industry.

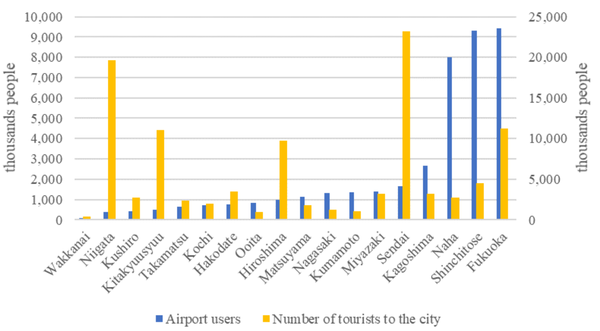
2. Literature review

There is a plethora of studies on the tourism industry from various perspectives. Tiefenbacher et al. [3] focused on the presence of repeat tourists in tourism and analyzed the impact of repeat tourists in the analysis of tourism success. Anna Ju. Aleksandrova et al, [4] focused on the periodicity of tourism activity and mathematically examined the periodicity of international tourism development at the global, regional, and local levels. Mahdi Samadzad et al. [5] aimed to characterize future urban air mobility travel demand for weekly business, airport access, and regional tourism trips in Iran, and showed that urban air mobility for weekly business trips is the most feasible market segment. Clement Kong et al [6]. empirically examined the benefits and impacts of airport subsidies to promote domestic tourism development in China, and empirically showed that the government's airport subsidy system for small and medium-sized airports is effective in supporting the development of aviation and tourism in ethnic minority areas, poor areas, and remote areas with poor land transportation. Tsutomu Ito et al. [7] analyzed the factors that attract visitors to regional tourism, and in their paper, a 4-cell model is proposed. The 4-cell model performs positioning in each quadrant based on the time series changes in the number of tourists and the number of tourist spots, and the first quadrant has a weight of 4, the second quadrant has a weight of 3, the third quadrant has a weight of 2, and gives a weight of 1 to the fourth quadrant. Ito et al. analyzed the characteristics of prefectures that are successful in attracting tourists and those that are not, depending on the public transportation development status of each prefecture in Japan. In addition, there are studies that have investigated the relationship between airport development and tourism, such as [8] and [9], but previous studies have mainly focused on tourists and tourist elements, and very few analysis has been conducted on the relationship between the development of infrastructure that supports tourism and regional tourism.

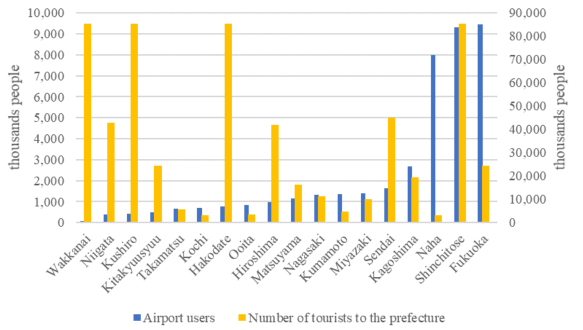
3. Classification of airports in Japan

In Japan, airports are classified according to the Airport Act[10]. In this study, we focus on regional domestic routes and regional, second-class airports.

Most of the second-class airports are airports that are established and managed by the national government. Nineteen relatively large airports, including Tokyo International Airport, Shin Chitose Airport, Itami Airport, Kitakyushu Airport, and Fukuoka Airport, are classified as nationally managed airports. Local governments have little involvement in the management and operation of nationally managed airports, and all funding is provided by the national government. These airports include Tokyo International Airport and 18 other airports designated by government ordinance. For Tokyo International Airport, the national government covers the full cost of construction of basic facilities and ancillary facilities. For airports designated by government ordinance, the national government covers the full cost of ancillary facilities, while the national government covers two-thirds of the basic facilities, and the local government covers one-third. For airports in Hokkaido, Okinawa, and remote islands, the national government's share of the cost of basic facilities is set at between 80% and 95% by the Airport Act and Special Regional Acts.

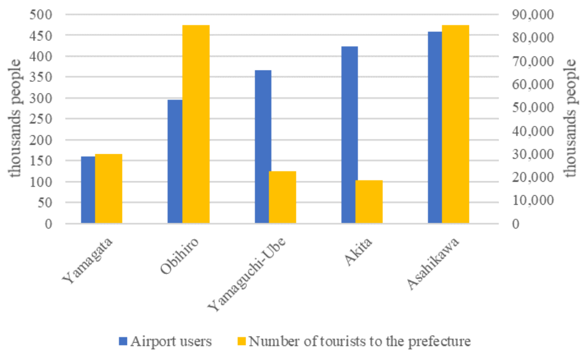


(a) Number of airport users and of tourists to the city



(b) Number of airport users and of tourists to the prefecture

(c) Number of designated airport users and of tourists to the city



(d) Number of designated airport users and of tourists to the prefecture

Figure 1 Comparative Result of Airport Users and Tourist Number.

On the other hand, there are five airports that were established by the national government and managed by local governments. They are Yamaguchi Ube Airport, Yamagata Airport, Akita Airport, Obihiro Airport, and Asahikawa Airport. These are called designated local airports. As stipulated in Item 6, Paragraph 1, Article 4 of the Airport Act, these are "airports designated by government ordinance as hub airports for international or domestic air transport networks," and the national government is to cover 55% of the construction costs for basic facilities, while local governments are to cover 45%. Ancillary facilities are to be borne by local governments, but the national government can subsidize up to 55% of the construction costs. For airports in Hokkaido, Okinawa, and remote islands, the proportion of construction costs borne and subsidized by the national government is higher due to the Airport Act and Special Regional Acts.

4. The impact of airport location on the tourism industry

We analyze whether the number of tourists that an airport can attract extends to the city where the airport is located or to the entire prefecture.

To work through this, we analyze the correlation between the number of airport users and the number of tourists in the city and prefecture where the airport is located and clarify the extent of the effect of airport location on the tourism industry. The number of airport users and the number of tourists is the dataset collected from 2021[11]. Figure 1(a) shows the number of tourists to regional airports and cities. Similarly, Figure 1(b) shows the number of tourists to regional airports and prefectures. Figure 1(c) and Figure 1(d) show the number of users of designated regionally managed airports and the number of tourists to cities and prefectures, respectively. No correlation was found between regional airports and the number of tourists in both cities and prefectures. This indicates that air routes to regional areas have business demand and tourism is not mainstream. On the other hand, in the case of designated regionally managed airports, no correlation was found between airport users and the number of tourists to prefectures. We confirmed that there is a correlation coefficient of 0.77 between the number of tourists to cities and the number of airport users. The P value is 0.13, which is not statistically significant, but this is thought to be influenced by the fact that there are only five samples of designated regionally managed airports. From the results mentioned above, it is seen that designated regional airports in which local governments are involved in the operation have a certain degree of tourist attraction effect on the cities in which they are located.

5. Access to tourist attractions from the airport and tourism attraction

5.1. *4-cell model*

In order to clarify the factors that attract visitors to each tourism element, an index has been proposed that classifies the number of tourists by purpose and evaluates them [7]. Seven types of tourism elements have been published in the Japan Tourism Agency's reference. They are: 1) nature, 2) hot springs, 3) history and culture, 4) sports, 5) urban tourism, 6) events, and 7) others. Thus, we use the Japan Tourism Agency's classification, as well as dataset of the number of visitors and tourism facilities, to analyze. We calculated the increase or decrease in the number of tourists for all elements and calculated the slope of the change in the number of tourists over a seven-year period. In addition, the increase or decrease in the number of tourist destinations was also applied as time-series data to calculate the slope of the change. A method called a four-cell model has been proposed that uses this slope to simultaneously express the fluctuations in the number of tourists and the number of tourism facilities. This four-cell model creates an evaluation index by taking the slope that represents the change in the number of tourist destinations on the horizontal axis and the slope that represents the increase or decrease in the number of tourists on the vertical axis.

5.2. *Evaluation of travel time*

We measured the evaluation index obtained from previous research and the travel time from airports in each prefecture to famous tourist destinations when traveling by car and/or using public transportation. Because the area of each prefecture varies greatly, we calculate the travel time divided by the area to smooth out the difference in travel time due to the size of the prefecture and make it easier to compare.

The estimated travel time is calculated using the following formula.

$$\text{Travel time index} = \frac{\mathit{Travel\ time\ from\ the\ airport\ to\ tourist\ spot}}{\mathit{Area\ of\ Prefecture}}$$

5.3. *Results and discussion*

Comparing the evaluation travel time and the evaluation index of the four-cell model, it is confirmed that the travel time per unit area tends to be short in Hokkaido and Hiroshima Prefecture when using public transportation, while it takes longer in Shimane Prefecture and Ehime Prefecture when using public transportation. When using a car, the results showed that it took less time overall, and it is confirmed that it took short time to travel by car in Shimane Prefecture and Hokkaido in particular. We assumed that when tourists from urban areas or abroad use local airports, it is more convenient to use public transportation rather than traveling by rental car. We compared the evaluated travel time with the evaluation indexes based on the four-cell model. The results are shown in Figure 2. Orange is the evaluation travel time

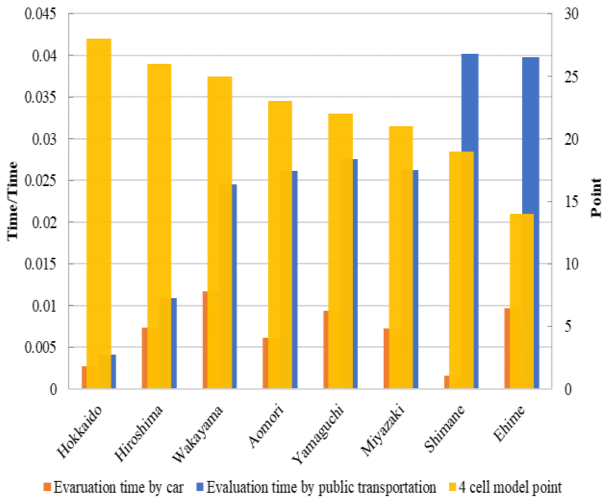


Figure 2 Comparison of evaluation travel time and evaluation indicators by prefecture.

when using a car, and blue is the evaluation travel time when using public transportation. Yellow is the evaluation index based on the four-cell model. From the figure, it is found that the higher the evaluation index, the shorter the time it takes to travel by public transportation. It was found that there was an inversely proportional relationship between the evaluation index and the unit travel time using public transportation from the airport to the tourist destination, with a strong negative correlation coefficient of -0.901. In addition, this result was confirmed to be statistically significant with a p-value of 0.002. Based on this result, it can be considered that the development of public transportation from the airport to the tourists' destinations is important for attracting tourism to rural areas.

6. Conclusion

In order to consider how airports can be utilized for regional tourism, this study aimed to understand the range of activities of airport users, and to clarify whether access from airports to tourists' destinations has an impact on attracting tourists. We confirmed whether there is a correlation between the number of airport users and the number of tourists, and conducted verification using a method called a four-cell model.

During the verification process, a correlation has been found between the number of tourists in the city where the airport is located and the number of airport users at specific regionally managed airports. Furthermore, regarding access to tourist elements from airports, a strong negative correlation has been confirmed between the evaluation index and the state of public transport development, suggesting the importance of developing public transport, including airports, in attracting tourists.

References

1. Japan Tourism Agency, Ministry of Land, Infrastructure, Transport and Tourism, Research study on economic impacts of tourism in Japan, March 2024. retrieved November 27, 2024
2. UN Tourism, UN Tourism Barometer, https://www.unwto.org/un-tourism-world-tourism-barometer-data, retrieved November 27, 2024
3. John P. Tiefenbacher, Frederik A. Day, Joseph A. Walton, Attributes of Repeat Visitors to Small Tourist-Oriented Communities, The social science journal Vol37/No.2. April 2000
4. Aleksandrova A.Ju., Dombrovskaya V.E. (2023). Cyclicality as a Pattern of Development of World Tourism.Outlines of Global Transformations: Politics, Economics, Law.vol.16, no.1, pp.86–106 (in Russian). /2023/02/05
5. Mahdi Samadzad, Fatemeh Ansari, Mohammad Amin Afshari Moez, Who will board urban air taxis? An analysis of advanced air mobility demand and value of travel time for business, airport access, and regional tourism trips in Iran, Journal of Air Transport Management, Volume 119, August 2024, 102636
6. Clement Kong Wing Chow, Wai Hong Kan Tsui, Hanjun Wu, Airport subsidies and domestic inbound tourism in China, Annals of Tourism Research 90 (2021) 103275
7. Tsutomu Ito, Seigo Matsuno, Makoto Sakamoto, Takao Ito, Discovering the relationship between tourists and tourist spots in Japan, The 2022 International Conference on Artificial Life and Robotics (ICAROB2022), January 20 to 23, 2022.
8. Ane Elixabete Ripoll-Zarraga, Josep Maria Raya, Tourism indicators and airports' technical efficiency, Annals of Tourism ResearchVolume 80, January 2020, 102819.
9. Walanchalee Wattanacharoensil, Markus Schuckert, Anne Graham, An Airport Experience Framework from a Tourism Perspective, Transport Reviews Volume 36, Issue 3, 2016, Pages 318-340
10. Ministry of Land, Infrastructure, Transport and Tourism website, Airport list, retrieved November 27, 2024
11. Ministry of Land, Infrastructure, Transport and Tourism Statistics and Data Aviation 2021 Airport Rankings, retrieved November 27, 2024

Authors Introduction

Dr. Tsutomu Ito

Dr. Tsutomu Ito is Assistant Professor of the Department of Business Administration at National Institute of Technology, Ube College, Japan. He has published many papers in refereed journals and proceedings, particularly in the area of industrial management, and computer science. His current research interests include internet of things (IoT), mechanical engineering, artificial intelligence (AI), automata theory, quantitative analysis of Japanese Keiretsu. He was one of the winners of the Best Paper Award in the International Conference on Artificial Life and Robotics (ICAROB) in 2015and 2016. Dr. Ito earned his doctor degree of Engineering from Hiroshima University, Japan in 2018.

Dr. Seigo Matsuno

Seigo Matsuno is a professor of the Department of Business Administration at National Institute of Technology, Ube College, Japan. He received his Ph.D. degree in Economics from Kyushu University, Japan in 2004. His current research interests are in the areas of IT management and strategy, digital transformation, and inter-firm relationship management. Dr. Matsuno has published several research papers and books in that area. His articles have been appeared in journals such as Procedia Computer Science, Journal of Robotics, Networking and Artificial Life, and so on. He is a member of the Board of the Directors of Japan Society for Information and Management.

Dr. Makoto Sakamoto

Makoto Sakamoto received a Ph.D. degree in computer science and systems engineering from Yamaguchi University. He is presently Professor in the Faculty of Engineering, University of Miyazaki. He is a theoretical computer scientist, and his current main research interests are automata theory, languages, and computation. He is also interested in digital geometry, digital image processing, computer vision, computer graphics, virtual reality, augmented reality, entertainment computing, complex systems and so on

Dr. Satoshi Ikeda

PhD, Hiroshima University. He is an associate professor in the Faculty of Engineering, University of Miyazaki. His research interest includes graph theory, probabilistic algorithm, fractal geometry and measure theory.

Dr. Takao Ito

Dr. Takao Ito is Professor of Management of Technology (MOT) in Graduate School of Engineering at Hiroshima University. He is serving concurrently as Professor of Harbin Institute of Technology (Weihai) China. He has published numerous papers in referred journals and proceedings, particularly in the area of management science, and computer science. He has published more than eight academic books including a book on Network Organizations and Information (Japanese Edition). His current research interests include automata theory, artificial intelligence, systems control, quantitative analysis of inter-firm relationships using graph theory, and engineering approach of organizational structures using complex systems theory.

Analysis of Careless Mistakes Using Gaze Information

Ryota Yabe
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, 820-8502, Japan
Email: yabe.ryota884@mail.kyutech.jp

Noriko Takemura
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, 820-8502, Japan
Email: takemura@ai.kyutech.ac.jp

Abstract

Careless mistakes are caused by a lack of concentration, time pressure, and information overload due to multitasking. If we predict careless errors, they could be helpful in various situations, such as learning support, business efficiency improvement, and medical diagnosis support. However, it is challenging to reproduce careless error situations, and few studies have been on predicting careless errors. Therefore, we focus on Shogi (Japanese chess), where situational awareness is complex, and players must maintain long periods of concentration, making careless mistakes. In this study, we collected eye-tracking data of players during a game and annotated careless errors based on the game's contents and surveys from the players. By analyzing this data, we have examined the conditions under which careless mistakes occur).

Keywords: Careless mistakes, Shogi, Gaze analysis

1. Introduction

Careless mistakes refer to errors due to a lack of attention or insufficient confirmation despite having sufficient knowledge or ability. Careless mistakes can have a significant impact, especially in academics and work.

Several factors trigger careless mistakes. First, a lack of concentration is a significant factor. When concentration is lacking, attention becomes distracted, and one may overlook details. In particular, performing the same task for an extended period or engaging in monotonous work can lower attention levels. Second, lack of time is another important factor. When there are deadlines or time constraints, work tends to become sloppy, and verification steps may be skipped. In such situations, tasks that would normally be performed carefully can lead to a higher likelihood of careless mistakes. Furthermore, information overload due to multitasking also affects performance. By engaging in multiple tasks simultaneously, attention is divided among these tasks, making it difficult to process information accurately. As a result, when attention is required for specific tasks, it is often insufficient, leading to an increase in careless mistakes.

As mentioned above, careless mistakes are caused by various factors, such as a lack of concentration, insufficient time, and information overload due to multitasking. If we could predict careless mistakes, there would be expectations for their application in various scenarios, including learning support, work efficiency improvement, and medical diagnosis support. However, while numerous studies exist on estimating human internal states [1] [2], research on predicting careless mistakes has not yet been conducted. This lack of research is due to the difficulty in collecting data, as careless mistakes cannot be intentionally produced.

Therefore, in this study, we focus on Shogi, a game in which careless mistakes can occur, and create a dataset. Shogi requires complex situational judgment, such as defending while attacking, ignoring the opponent's attack and attacking each other, and defending from the opponent's attack. Additionally, it necessitates prolonged concentration, making it prone to fatigue and a decline in focus. In addition, because of the time limit, it is easy to feel the mental pressure of time. Given these factors, Shogi is highly susceptible to careless mistakes and is suitable for data collection.

Additionally, to predict careless errors, it is first necessary to analyze the circumstances under which careless errors occur. In this research, gaze information is utilized for that analysis. Gaze information can provide insights into what a person is focusing on, what they are struggling with, and signs of fatigue or drowsiness. This information is believed to be crucial for predicting careless mistakes. Although gaze movements are assumed to vary depending on the task being performed, it is expected that the analysis of gaze movement information during a game of Shogi in this study will provide clues for clarifying the mechanism of careless errors.

The subjects annotate the collected gaze data regarding the presence or absence of careless mistakes for each move, thereby the gaze dataset is constructed. For the analysis, various features are extracted from the gaze dataset, and F-tests and t-tests are conducted to analyze whether there is a significant difference in gaze characteristics related to careless mistakes.

2. Construction of a Gaze Dataset

2.1. *Dataset overview*

Five male subjects in their 20s compete against a shogi AI, and gaze information during the match is collected using an eye tracker. Furthermore, after the match, the subjects annotate the presence or absence of careless mistakes for each move, thereby constructing the dataset. The composition of this dataset is shown in Table 1.

Table 1. Composition of The Gaze Dataset.
("#" indicates the total count.)

ID	#Matches	#Moves	#Careless Mistake Moves
1	7	389	1
2	1	36	1
3	5	170	5
4	6	241	5
5	7	195	4
Total	26	1031	16

2.2. *Shogi GUI*

Shogi is a type of traditional Japanese two-player board game. Players move their pieces on a 9×9 square board to capture the opponent's king. The pieces consist of eight types: king, gold general, silver general, knight, lance, pawn, rook, and bishop.

In this study, five male subjects in their 20s who are familiar with shogi rules compete against the Shogi AI 'Suisho5' [3] using ShogiGUI [4] under the rule that each move must be made within 20 seconds. ShogiGUI (Fig. 1) is a graphical user interface software for Shogi that runs on Windows. In ShogiGUI, players make their moves by dragging and dropping the pieces. To adjust the strength of the opponent, the number of positions 'Suisho5' searches is adjusted by setting the node limit between 50 and 100, allowing for a match where humans can win. Before the experiments, the subjects adjust the settings by playing against the AI to ensure it is neither too strong nor weak as an opponent.

2.3. *Gaze features*

To obtain gaze information, we use the Tobii Pro X3-120 [5]. This eye tracker is a screen-based model installed at the bottom of the display with a sampling rate of 120Hz. This eye tracker can collect gaze coordinates and pupil diameter on the display; in this research, we use the gaze coordinates. The coordinates originate at the top left, with the horizontal direction as the x-axis and the vertical direction as the y-axis. The gaze heat map created from the collected data during the matches is shown in Fig. 2.

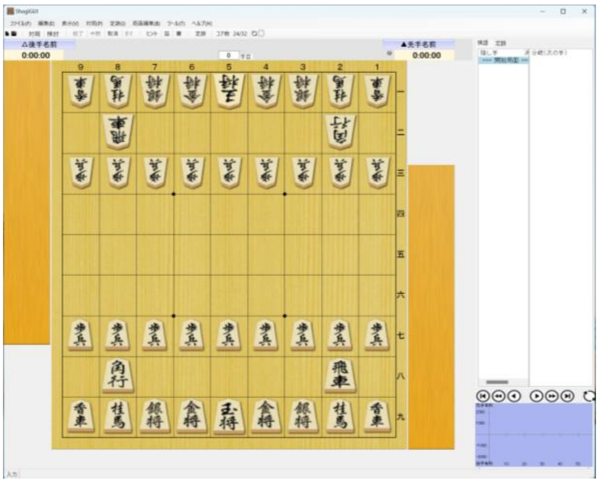


Fig.1 ShogiGUI.

Fig.2 Gaze Heat Map During the Match.

Fig. 2 shows a heat map generated from the gaze information for a single move, with a Gaussian distribution overlay on each gaze coordinate to create the heat map. From the heat map, it is possible to understand what moves the player is considering and what moves they are deliberating between.

2.4. *Annotation*

Annotations of careless mistakes are conducted. After the game, subjects annotate the game while viewing the game in ShogiGUI.

A dataset is created based on the gaze data and annotations during the match. First, the gaze data is divided into segments between the moment the opponent places a piece and when the subject places their piece. This study does not analyze the gaze data during the AI's thinking period. The moves the subject identified as having a careless mistake are annotated as such, while all other moves are annotated as without careless mistakes. Any missing values in the gaze data are entirely removed.

3. Analysis of Gaze Features

3.1. *Analysis method*

We extract features from the created dataset and conduct tests to analyze whether there is a significant difference in gaze information related to the occurrence of careless mistakes.

3.1.1. *Testing method*

In this study, F-tests and t-tests are conducted. The F-test is a method used to determine whether there is a significant difference in the variances of two data groups. In comparison, the t-test checks for a substantial difference in the means of the two data groups. First, we conduct an F-test to check whether the two data groups have equal variances. If they are equal, a Student's t-test is performed; if not, a Welch's t-test is conducted. Additionally, the significance level is set at 0.05 for both tests.

3.1.2. Gaze features

The features used for the tests are shown in Table 2. The mean and variance are calculated based on the gaze coordinates on the display. Additionally, the movement amount is determined by calculating the Euclidean distance between the gaze coordinates at the current timestamp and those from one timestamp prior, as well as by computing their sum or average. Regarding the skewness and kurtosis of the movement amount, skewness indicates the symmetry of the distribution, while kurtosis reflects the sharpness of the distribution. Based on these features, we conducted tests for each left eye, right eye, and both eyes. The data for both eyes is treated as the average of the left and right eyes.

3.2. *Analysis results*

3.2.1. *F-Test*

The results of the F-test are shown in Table 3. According to Table 3, there is a difference in the variance of both eyes' movement amount, skewness, and the kurtosis of the left eye's movement amount between the times when careless mistakes were made and when they were not.

Table 2. Features Used for the Tests.

Gaze Coordinates	Mean	x-axis
		y-axis
	Variance	x-axis
		y-axis
Movement of Gaze Coordinates	Sum	
	Mean	
	Variance	
	Skewness	
	Kurtosis	

3.2.2. *t-Test*

The results of the two-tailed tests are shown in Table 4. For the variance and skewness of the movement amount of both eyes and the kurtosis of the left eye's movement amount, which were found to have a significant difference in the F-test, Welch's t-test is performed. For the other data, a Student's t-test is conducted.

According to Table 4, there is a significant difference in the skewness and kurtosis of the right eye's movement amount between careless mistakes and not.

According to Table 3 and Table 4, several features show significant differences; however, many do not. One possible reason for this is that the sample size needs to be bigger, which may have affected the effectiveness of the tests. Therefore, increasing the number of data points in the future is necessary.

In this experiment, tests were conducted regarding gaze coordinates and gaze movement, but the treatment of missing values and the discussion of fixations and saccades are insufficient. Therefore, these factors must be considered in future analyses.

Table 3. F-test.

			left eye	right eye	both eyes
Gaze Coordinates	Mean	x-axis	0.413	0.469	0.293
		y-axis	0.346	0.297	0.353
	Variance	x-axis	0.367	0.240	0.266
		y-axis	0.236	0.300	0.238
Movement of Gaze Coordinates	Sum		0.394	0.247	0.156
	Mean		0.339	0.399	0.338
	Variance		0.062	0.098	**0.022**
	Skewness		0.083	0.116	**0.049**
	Kurtosis		**0.006**	0.079	0.131

Table 4. t-test (Two-tailed Test).

			left eye	right eye	both eyes
Gaze Coordinates	Mean	x-axis	0.137	0.161	0.435
		y-axis	0.071	0.080	0.075
	Variance	x-axis	0.745	0.892	0.309
		y-axis	0.488	0.675	0.511
Movement of Gaze Coordinates	Sum		0.279	0.224	0.156
	Mean		0.904	0.756	0.791
	Variance		0.608	0.970	0.741
	Skewness		0.798	**0.026**	0.562
	Kurtosis		0.913	**0.042**	0.709

4. Conclusion

Using gaze information during shogi matches, we employed F-tests and t-tests to test whether there is a significant difference between the gaze associated with careless mistakes and the gaze when no careless mistakes were made.

Through the tests, it was clarified that there are significant differences in features concerning variance and mean of gaze coordinates. In the future, this gaze data and the extracted features are expected to contribute to developing a model for predicting careless mistakes. Additionally, while this study utilized gaze data during shogi, where careless mistakes are likely to occur, it is essential to analyze under what circumstances careless mistakes happen beyond just shogi, broadening the discussion to other events. This remains a key challenge for the future.

References

1. S L Happy, Anirban Dasgupta, Priyadarshi Patnaik, and Aurobinda Routray. Automated alertness and emotion detection for empathic feedback during e-learning. In 2013 IEEE Fifth International Conference on Technology for Education (T4E), pp. 47–50, 2013.
2. Salma Alhagry , Aly Aly Fahmy, and Reda A. El-Khoribi. Emotion recognition based on EEG using LSTM recurrent neural network. Emotion vol. 8, no.10, pp. 355–358, 2017.
3. yaneurao, "YaneuraOu," https://github.com/yaneurao/YaneuraOu, (accessed on December 11, 2024).
4. siganus, "Shogi GUI", http://shogigui.siganus.com/, (accessed on December 10, 2024).
5. tobii, "Tobii Connect", https://connect.tobii.com/s/x3-downloads?language=ja, (accessed on December 10, 2024).

Authors Introduction

Mr. Ryota Yabe

He received his Bachelor's degree in Computer Science and Systems Engineering in 2024 from the Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology in Japan. He is currently a master student at Kyushu Institute of Technology, Japan

Dr. Noriko Takemura

She received B.S., M.S., and Ph.D. degrees in engineering from Osaka University, Japan, in 2006, 2007, and 2010, respectively. She is currently an associate professor at Graduate School of Computer Science and Systems Engineering , Kyushu Institute of Tech-nology. Her current research interests include human-centric intelligent systems

A Mathematical Framework for Logit Model in Transportation Mode Choice Analysis

Ahmad Altaweel
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Kazuhito Mine
School of Computer Science and Systems Engineering, Kyushu Institute of Technology
680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan

Bo-Young Lee
Logistics Revolution Korea Co., Ltd.
5th FI. 47, Gangnam-daero 101-gil, Seocho-gu, Seoul 06034, Republic of Korea

Jang-Sok Yoon
Logistics Revolution Korea Co., Ltd.
5th FI. 47, Gangnam-daero 101-gil, Seocho-gu, Seoul 06034, Republic of Korea

Hiroaki Wagatsuma
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Email: altaweel.ahmad770@mail.kyutech.jp, mine.kazuhito139@mail.kyutech.jp, {bylee, kjsyoon}@logisroad.com, waga@brain.kyutech.ac.jp

Abstract

Traditional transportation demand forecasting has relied on massive zone-specific aggregations, which assume a linear demand increase. Such models may lack the flexibility needed to perform dynamic and context-sensitive analyses. Recently, disaggregated behavioral models have gained prominence for requiring less data and enabling sensitivity analyses in policy decisions. This study explores the feasibility of a model focusing on the mathematical formulation and validation of the transportation mode choice model. The study uses the logit model with a non-linear probability distribution function represented by a logistic curve and incorporates a linear combination of independent predictor variables. The mathematical model is examined for its ability to estimate choice probabilities. The methodology is formulated to be adaptable to diverse contexts that provide an analytical framework for transportation systems independently of geographic or demographic considerations.

Keywords: Transportation Mode Choice, Logit model, maximum likelihood estimation, Gumbel distribution.

1. Introduction

In microeconomics, disaggregated behavioral models such as the Logit model have been used to explain qualitative choice in specific phenomena. The logit model uses a logistic function to fit a probability distribution as a nonlinear S-shaped sigmoid curve. This allows for predicting discrete transitions between different modes. The sigmoid function is frequently utilized in soft computing for binary transitions or classifications. Consistently, the Logit model simulates individual decision-making by assuming rational behavior aimed at maximizing personal utility. Utility is a quantifiable measure of an individual's satisfaction from a specific behavioral change.

This paper is intended to enhance the understanding of model structure and estimation of the Logit model with a focus on the data structure and estimation procedure. Through the incorporation of attributes such as distance, time, and cost, the model evaluates their impact on decision-making processes and the resulting probabilities. Recognizing the challenges associated with collecting data to validate such models, this study adopts a structured data generation approach, enabling control over the attributes and the diversity of decision-making scenarios. This method allows for evaluation of the logit model precision under predetermined conditions which support critical alternatives selection. Section 2 explores the assumptions and formulation of the logit model. In addition, the design of the structured dataset. We then discuss the dataset and its estimation in the results and discussion section and finally the conclusion.

2. Methodology

2.1. Logit Model Assumptions and General Framework

The logit model is a choice model that uses probabilistic approaches to evaluate individual preferences for different modes of transport. It formulates mathematically the relation between attributes of the alternative and characteristics of the decision maker into a utility function. The concept of utility allows one to rank a series of alternatives. The utility maximization rule states that an individual will select the alternative that maximizes his utility [1]. Despite this theoretical framework, there are three primary sources of error in using deterministic utility functions. First: the individual has incomplete or incorrect information or misperceptions about the attributes of alternatives. Second, the observer has different or incomplete information about the attribute. Third: when analysts do not fully understand the decision process.

The general framework for using the logit model in travel mode can be summaries in Fig. 1 in the following steps: (1) Define problem context, (2) Gather mode attributes and observe choices (3) Define utility function and Identify predictor vars (4) Estimate model coefficients (5) Evaluate the model by comparing predicted choices and observed choices (6) Validate the model by performing sensitivity analysis of coefficient values (7) Analyze results to provide policy insights.

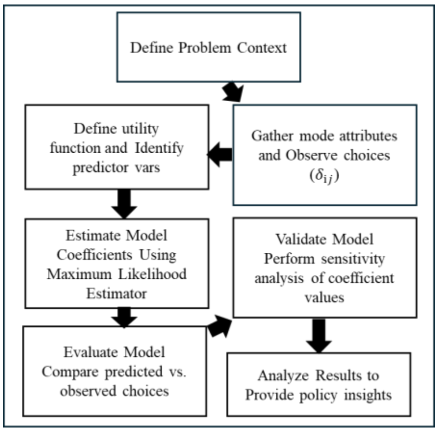


Fig. 1 The general framework for using the logit model

2.2. Logit Model Formulation

The logit model in this study is estimated in three steps:

1. Determine the shape of the utility function.
2. Use the least squares method to approximate the coefficient values.
3. Find a solution that maximizes the log-likelihood function.

Assume that the utility U_{ij} of individual j when choosing option i is expressed as:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \qquad (1)$$

In the model, V_{ij} in Eq. (1) represents the deterministic component of the utility, while ε_{ij} denotes the stochastic component which accounts for random error in the true utility. Assuming i represents a binary choice of either 0 or 1, then the deterministic component V_{ij} is expressed as a linear function of time t_{ij} , cost c_{ij} and distance d_{ij} as:

$$\begin{aligned} V_{0j} &= \alpha_0 d_{0j} + \alpha_1 c_{0j} + \alpha_2 t_{0j} \\ V_{1j} &= \alpha_0 d_{1j} + \alpha_1 c_{1j} + \alpha_2 t_{1j} \end{aligned} \qquad (2)$$

In the Eq. (2), α_k represents an unknown coefficient. Once α_k is known, the deterministic term V_{ij} can be calculated by using time, cost, and distance. If the stochastic component ε_{ij} follows normal distribution, it is possible to use the least squares method to find the coefficient α_{ij}. On the other hand, if ε_{ij} follows Gumbel distribution, the maximum likelihood estimation method is more flexible and can handle non-normal error terms. Since an analytical solution is generally unattainable, iterative methods like Newton's are employed to find a numerical solution.

The probability P_i of choosing option i is expressed as:

$$P_i = \frac{e^{V_i}}{e^{V_0} + e^{V_1}} \qquad (3)$$

Then the binary probability is:

$$P_0 = \frac{1}{1 + e^{V_1 - V_0}}, P_1 = \frac{1}{1 + e^{V_0 - V_1}} \qquad (4)$$

By taking the logarithm of the ratio of probabilities P_0 and P_1, we derive the log-odds ratio or logit:

$$\log \frac{P_1}{P_0} = \log \frac{e^{V_1}}{e^{V_0}} = V_1 - V_0 \qquad (5)$$

$$\begin{aligned} \log \frac{P_1}{P_0} &= \alpha_0 (d_1 - d_0) + \alpha_1 (t_1 - t_0) \\ &\quad + \alpha_2 (c_1 - c_0) \end{aligned} \qquad (6)$$

Thus, the data for J individuals represent matrix:

$$p = \left[\log \frac{P_{10}}{P_{00}}, \cdots, \log \frac{P_{1J}}{P_{0J}} \right]^T$$

$$\alpha = [\alpha_0, \alpha_1, \alpha_2]^T \qquad (7)$$

$$X = \begin{bmatrix} d_{11} - d_{01} & t_{11} - t_{01} & c_{11} - c_{01} \\ \vdots & \vdots & \vdots \\ d_{1J} - d_{0J} & t_{1J} - t_{0J} & c_{1J} - c_{0J} \end{bmatrix}$$

As for the normal equations:

$$X\alpha \approx p$$

$$\alpha \approx [X^T X]^{-1} X^T p \qquad (8)$$

An estimated value of α can be obtained via the least squares method, denoted as $\widehat{\alpha}'$.

Taking the logarithm of the probability in Eq. (4) gives:

$$\log P_{0j} = -\log\left(1 + e^{V_{1j} - V_{0j}}\right)$$

$$\log P_{1j} = -\log\left(1 + e^{V_{0j} - V_{1j}}\right) \qquad (9)$$

In this case, the log-likelihood function $L(\beta)$ over J observations can be written as:

$$L(\alpha) = \sum_{j=1}^{J} \left(\delta_{0j} \log P_{0j} + \delta_{1j} \log P_{1j}\right) \qquad (10)$$

The term δ_{ij} denotes the Kronecker Delta function, which return 1 when individual j chooses option i and 0 otherwise.

2.3. Structured Data Set Design

The dataset design shows intercity structure with nodes representing stops accessible by bus or car, and edges representing roads or pathways. The structure is a square grid with "n" nodes per side. Path selection between nodes has three criteria: travel distance, time, and cost. The bus route follows a fixed, clockwise loop along the grid's outer contour excluding interior nodes (Fig. 2). This predefined loop imitates real bus route and to enhance realism, individuals can walk to the nearest node on the bus route if their starting point is not directly on it. Individuals total travel then is walking to the nearest bus stop, taking the bus along its predefined route and walking to their destination. In contrast, Fig. 3. shows in this structure cars have bi-directional connectivity between all adjacent nodes (two ways paths). Individuals using cars directly access any node bypassing the restrictions in the bus case.

To compute the shortest paths between all pairs of nodes, the Floyd-Warshall algorithm [2] was used. For the bus route, it is assumed to operate continuously moving along the route. As a result, distances between non-adjacent nodes on the bus route are often longer due to directional restrictions. Conversely, car distances are symmetric and consistent regardless of the direction of travel. The distance between adjacent nodes is assumed to be uniform at one distance unit of 1 km.

Assuming the individuals can walk to the nearest node on the bus route and then use the bus to travel to other nodes, a mixed distance, time, and cost were made for bus mode. The design involves the process: (1) identifying nodes that have no adjacent nodes in the adjacency matrix (2) determining the nearest node prioritizing nodes located on the bus route (3) updating the distance matrix by adding the distance between the node and the nearest node.

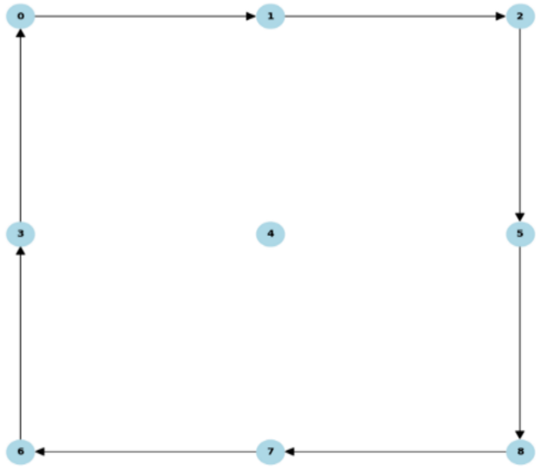


Fig. 2. Grid 3 × 3 represents the bus route

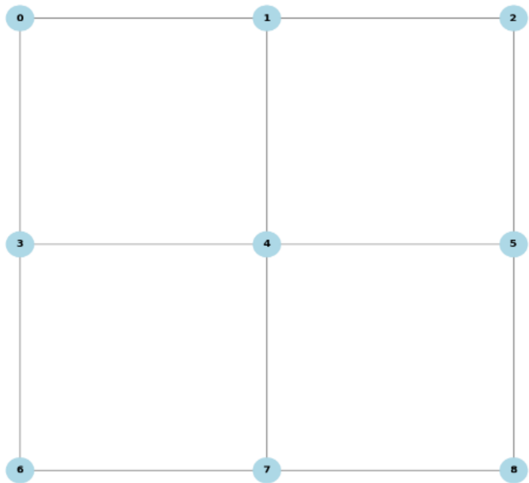


Fig. 3. Grid 3 × 3 for the car route

To calculate travel time, we assumed a bus speed of 30 km/h and a walking speed of 5 km per hour as fixed values across all distances. This assumption allows us to derive the travel time for bus and walking segments by dividing the respective distances by their speeds. For a car trip we assumed a minimum speed of 5 km/h and a maximum speed of 50 km/h with a random distribution on the trips in the dataset. Walking is assumed to have no cost, while the bus cost is 150 units, and the car is 170 units. For the car mode, the cost is then influenced by multiple factors, including driver behavior factor, fuel rate factor, maintenance rate, and vehicle type factor. Similarly, the cost matrix for the mixed-mode bus trips assumes a walking cost of zero for movement between any pair of nodes. Under this framework, for nodes without adjacent connections, the cost of movement is assigned as the bus travel cost from the nearest node that has access to the bus route.

3. Results and Discussion

3.1. Dataset Formation

It is important to consider that decision-makers assign different levels of importance to travel distance, time, and cost. Developing choice models at the decision-maker level allows us to incorporate variables that capture these differences. We assume travelers are business-oriented and prioritize time over cost. For that, time is more critical than cost, which is reflected in the sensitivity relationship: $\beta_{\text{time,work}} = r_{\text{work}} \cdot \beta_{\text{cost,work}}$. By defining base sensitivities as $\beta_{\text{cost,work}} = -0.9$, $\beta_{\text{time,work}} = -3.6$, and $\beta_{\text{distance,work}} = -0.5$ that emphasize the importance of time to the cost for these individuals. Fig. 4 and Fig. 5 illustrate the distribution of the generated utility of buses and cars for time and cost.

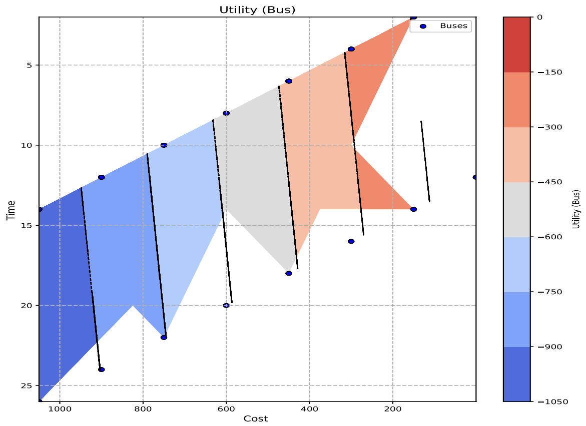


Fig. 4. The generated utility associated with the bus for all combinations of cost and time.

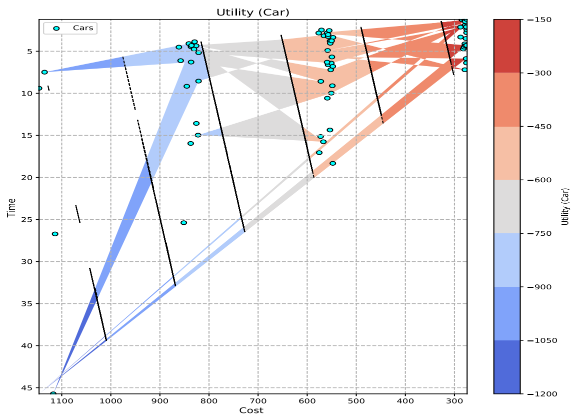


Fig. 5. The generated utility associated with the car for all combinations of cost and time.

Using the logit model, we determined the observed probabilities by maximizing the utility captured by the logit model. Fig. 6 shows the probabilities observed for the bus and Fig. 7 shows the observed probabilities for the car. According to the observed probabilities, the share of buses is 0.5937, while the share of cars is 0.4063.

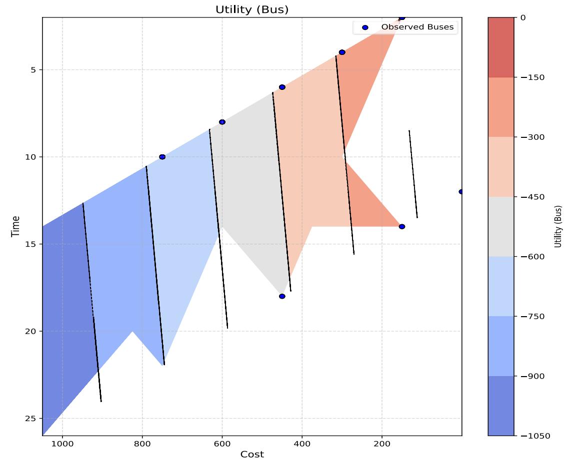


Fig. 3. The probabilities observed for the bus

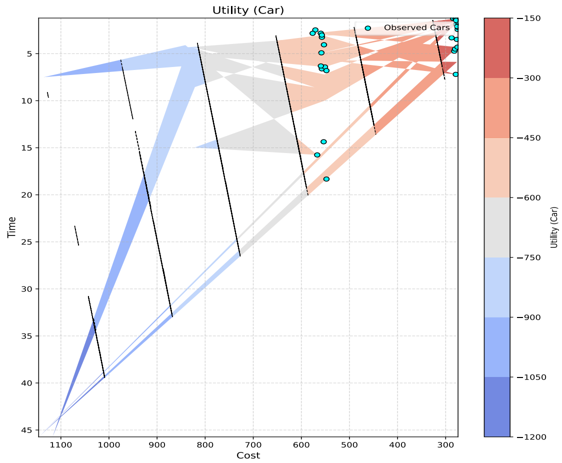


Fig. 7. The observed probabilities for the car

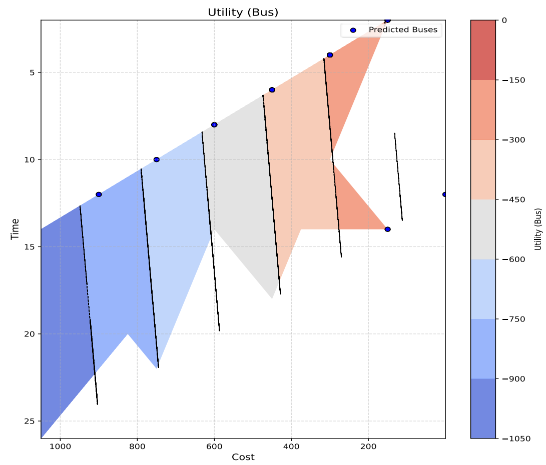


Fig. 8. The estimated probabilities for the bus.

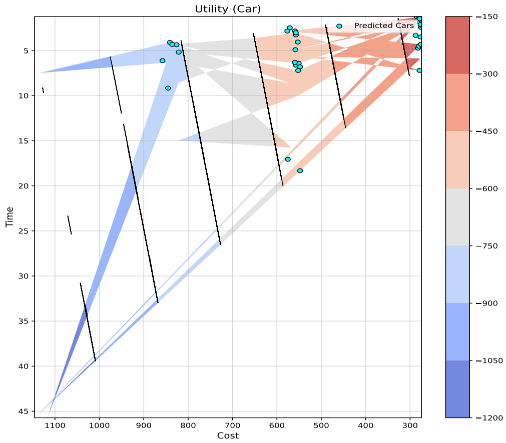


Fig. 9. The estimated probabilities for the car.

To estimate the parameters of the logit model (β) based on distance, time, and cost, we initiated with an estimation obtained using the least squares error method. These β values were then refined through maximum likelihood estimation and Newton's method. The mean squared error (MSE) between the observed probabilities of the bus and the predicted probabilities was calculated to be 0.139. Fig. 8 and Fig. 9 show the estimated probabilities for the bus and the car. After estimation, the estimated share of the bus is 0.5139, and the estimated share of cars is 0.4861.

4. Conclusion

This study evaluates a structured dataset to assess the logit model's accuracy in predicting predesigned observed probabilities. To conclude, relying solely on distance, time, and cost restricts the model's versatility as these attributes make utility and probability differences overly predictable. The simplicity limits the evaluation of the model's ability in this straightforward scenario. Stressing the need for additional attributes to enhance competition between modes and less diverted probabilities estimated by the mode.

Acknowledgments

This work was supported in part by the cooperative research project on digital logistics between the Kyushu Institute of Technology and Logistics Revolution Korea Co., Ltd.

References

1. F. S. Koppelman and C. Bhat, A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models, prepared for U.S. Department of Transportation, Federal Transit Administration, Jan. 31, 2006.
2. R. W. Floyd, "Algorithm 97: Shortest Path," Communications of the ACM, vol. 5, no. 6, p. 345, June 1962.

Authors Introduction

Mr. Ahmad Altaweel

He received his Bachelor's degree in Electronics and Communication Engineering, ALBAATH University, Syria. He is currently a Ph.D. student at Kyushu Institute of Technology, Japan.

Mr. Kazuhito Mine

He is a bachelor's student in the School of Computer Science and Systems Engineering, Kyushu Institute of Technology, Japan

Ms. BoYoung Lee

She received her bachelor's degree in human relations from Keio University and M.S. in logistics from Inha University in Korea and is currently a Ph.D. student. She is a logistics consultant in the Logistics Revolution Korea Co., Ltd.

Dr. JangSok Yoon

He received his M.S. and Ph.D. degrees from Kyunghee University, in South Korea, in 2004 and 2008. He is the founder, CEO, and Consultant over 23 years of Logistics Revolution Korea Co., Ltd .

Dr. Hiroaki Wagatsuma

He received his M.S., and Ph.D. degrees from Tokyo Denki University, Japan, in 1997 and 2005, respectively. He is currently an Associate Professor at Kyushu Institute of Technology.

A Computational Approach for Global Trade Analysis in Korea Contributing to the Forecasting of Future Efficacy in Global and Domestic Korean Transportations

Bo-Young Lee
Logistics Revolution Korea Co.
Ltd.,5th FI. 47, Gangnam-daero 101-gil, Seocho-gu, Seoul 06034, Republic of Korea

Ahmad Altaweel
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Kazuhito Mine
School of Computer Science and Systems Engineering, Kyushu Institute of Technology
680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan

Jang-Sok Yoon
Logistics Revolution Korea Co., Ltd.
5th FI. 47, Gangnam-daero 101-gil, Seocho-gu, Seoul 06034, Republic of Korea

Hiroaki Wagatsuma
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Email: bylee@logisroad.com, {altaweel.ahmad770, mine.kazuhito139}@mail.kyutech.jp,
kjsyoon@logisroad.com, waga@brain.kyutech.ac.jp

Abstract

Economic forecasting studies are integral for shaping strategic policy decisions by providing data-driven insights that guide resource allocation, logistics and transportation, and long-term planning. This study investigate the trade dynamics of the Republic of Korea through the Global Trade Analysis Project Recursive dynamic GTAP-RD model with the GTAP v11 database to forecast economic scenarios and Shared Socioeconomic Pathways (SSPs) serve as growth trajectories. The analysis centers on the Republic of Korea's key trading partners, as identified by the GTAP database, and top trading sectors from the Korea Transport Database (KTDB) to compare the key influencers on Korea's trade thereby providing deeper strategic economic planning. This study investigates further the involvement of Korea's logistics and transportation by focusing on changes in import/export tonnage to inform infrastructure planning and strategic transport development. The evaluation of Korea's trade in the global context is of the essence to ensure adaptive logistics, backing economic resilience, and aligning with changeable global trade conditions.

Keywords: Computable general equilibrium (CGE), General Equilibrium Modelling PACKage (GEMPACK), Global Trade Analysis Project Recursive Dynamic (GTAP-RD)

.

1. Introduction

Computable General Equilibrium (CGE) Models have emerged as competitive tools for policy evaluation. Their widespread applications include research on trade policy, regional infrastructure development, and environmental protection. Developments and investments in the Republic of Korea’s infrastructure require a meticulous economic analysis to build a data driven decisions based on a clear vision. In this study, we focus on forecasting economic quantities that have an impact on the development of transportation infrastructure in the Republic of Korea (ROK). The method used in the research guarantees a wide range of economic trajectories to extract insights about Korean economic development in the context of global changes. This supports a reliable evaluation of the uncertainty in future narratives.

The main objectives of the study are to forecast the values of bilateral trade of the Republic of Korea using the Global Trade Analysis Project Recursive dynamic model GTAP-RD [4] and the latest version of the GTAP database ver. 11 [1] with 2017 as a reference year. In addition, by implementing the forecasting changes from the GTAP-RD model which represents the global changes based on the Shared Socioeconomic Pathway narratives [6], we calculated the bilateral Volume of the Export and Import starting from the volume of commodities extracted from the Korean transportation database KTDB [8]. This study serves as a standing point for decision makers in their efforts for data informed strategies. The methodology section outlines the modeling procedure and utilized database, while section 2.2 describes the experimental setup of the baseline and policy Scenarios. Section 3 presents the results and Section 4 concludes the study and findings.

2. Methodology
2.1. Model and Database Preparation and Tools

First developed in 1992, the Global Trade Analysis Project GTAP model has become the de facto standard and starting point for many economy-wide analyses of global trade issues. The first full documentation of the GTAP model became available in Hertel's book [2]. also, the dynamic extension of the GTAP model has been included in a dynamic version, called GDyn, which is described in [3], and Global Trade Analysis Project Recursive Dynamic GTAP-RD the so-called WTO Global Trade Model [4] which we use in this study to forecast the trade dynamic of the Republic of Korea. Within the model, each region has its own economy, and the regions are linked through inter-regional trade flows. In each region, firms perform production activity using a multi-level, nesting production structure. Top-level nesting, lower-level nesting structure, and bottom-level nesting [5].

In this study, we use the GTAP Database which integrates diverse data sources on a global scale, providing a rich dataset on value transactions, quantities, and various tax instruments across time, making it a cornerstone for studies on global economic issues. Version 11 of GTAP includes time series data for five reference years across 65 economic sectors within a total of 160 countries and broader regions. This coverage captures more than 95 % of the world's Gross Domestic Product (GDP) and population [1]. All quantities are measured in millions of current U.S. dollars for the year of the database. The most recent updates of the GTAP database represent the year 2017 serving as a reference year for the equilibrium of the GTAP-RD model at the beginning year of this study.

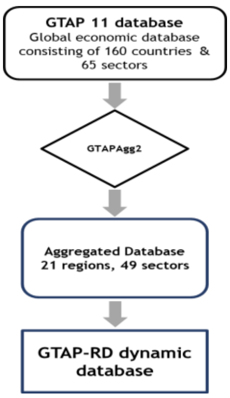


Fig. 1. Regional and Sectors Aggregation flow chart

Since our study focuses on the Republic of Korea, we compiled the top sixteen countries from the GTAP database and selected the top 10 trade export and import partners. In Table 1, we aggregated the regions accordingly using the GTAPAgg2 and the GTAP-RD Data Aggregation Utility [9] as illustrated in Fig. 1. In addition, The 65 sectors in the GTAP Database are combined according to an aggregation scheme of only 24 sectors aggregated into 8 broader categories while retaining the remaining 41 sectors as distinct sectors. The selected aggregation scheme of the 65 GTAP database into the 49-sector aggregation offers a highly granular view of the economy, providing a more detailed breakdown of economic activities.

Table 1. Aggregation of GTAP database vr.11 into 21 region

NO	COUNTRY/REGION	NO.	COUNTRY/REGION	NO.	COUNTRY/REGION
1	South Korea	9	India	17	South Asia
2	China	10	Singapore	18	East Asia
3	United States of America	11	Russia	19	Sub-Saharan and North Africa
4	Japan	12	Southeast Asia	20	Oceania
5	Taiwan	13	Middle East	21	Rest of World
6	Australia	14	Western Europe		
7	Germany	15	Latin America		
8	Saudia Arabia	16	North and Central America		

All tools used in this study are part of the General equilibrium modeling package GEMPACK [7].

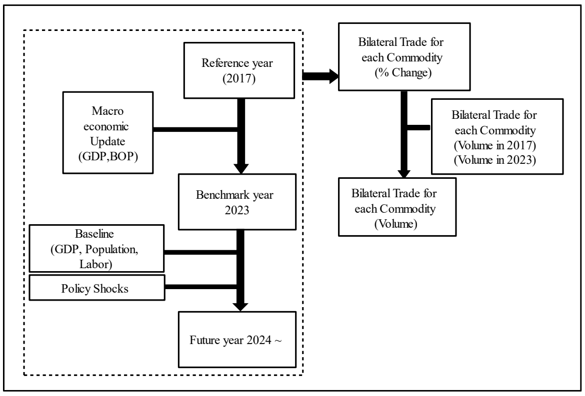


Fig. 2. The Process of Calculating Volume and Value of Bilateral Trade for the Republic of Korea

2.2. Baseline and Policy Scenarios

In the experimental setup, we define closures for different periods of the study to control specific model variables. The closures used are as follows: The baseline projection is run under the default GTAP-RD closure for all regions and to make the GTAP model follow a chosen growth path, real GDP (qgdp) is swapped with the region-wide technological change (afereg). In the policy closure, we swap the slack variable in the closure(cgdslack) with the trade balance as a percentage of the world income (del_tbalry) for the Southeast Asia which includes Vietnam. During the period from 2017 to 2023, the changes in GDP growth are based on data from the International Monetary Fund (IMF) [10]. In the baseline run, we assumed that the changes in the economic variables follow the projections of the SSP2 (Shared Socioeconomic Pathway 2) narrative [6]. This approach aligns the baseline scenario with moderate socioeconomic growth assumptions to represent the business as usual. The economic variables that constitute the shock values are population changes, real GDP, and labor. They are sourced from GTAP-RD Data Aggregation Utility [9] which contains the SSP scenarios after adjustments to align with the GEMPACK environment. This data simulates projections of SSP narratives for different growth assumptions about the global economy as policy scenarios. Considering timeline of the forecasting that shown in Fig. 3, we focused solely on the change in GDP growth rate from 2017 to 2023 to update the economic model. In this case, we assumed a fixed trade balance during this period, with no changes in the trade balance as a percentage of the world income, implying that the trade balance remained constant from 2017 to 2023.

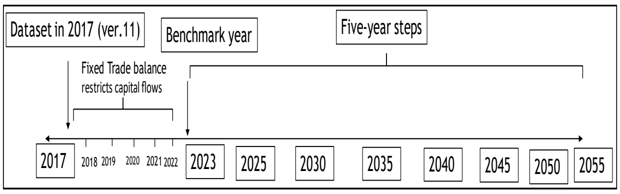


Fig. 3. Temporal Framework of the Forecasting Process

3. Results and Discussion

3.1. Forecasting Trade Values and Volume

The GTAP-RD model ensures that changes in bilateral trade, both exports and imports, for each commodity are reported as nominal values, meaning they reflect changes in both quantity and price. However, our study focuses on reporting the changes in bilateral trade in terms of real quantities, isolating the effect of price and capturing the actual volume changes. Therefore, we use the accumulated percentage change in quantity (qxs) for each commodity and multiply it by the bilateral trade values in free on board prices (VFOB) from the base year (2017). This allows us to express the trade values in base-year prices, specifically 2017 prices reported in millions of USD.

Fig. 4 shows the values of total export from the Republic of Korea to the global economy for all commodities by using SSP narratives assuming the SSP2 is the baseline for comparison to bridge the gap of uncertainty regarding the future development of economies. Similarly, Fig. 5 illustrates the values of imports from the year 2017 till 2055, the final year of the study.

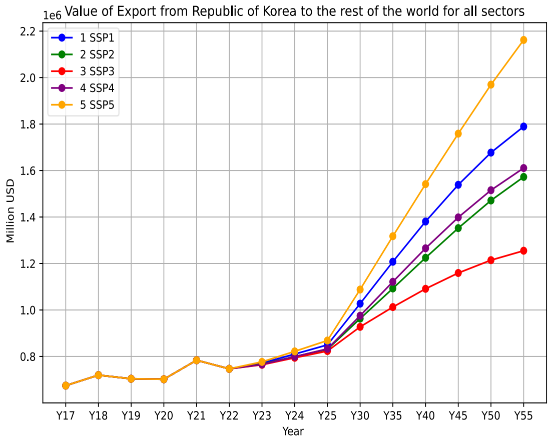


Fig. 4. Export from the Republic of Korea in Million USD

In line with the flowchart Fig. 2, we calculated the bilateral trade volume in tons for all goods (all GTAP database sectors except services) using data from the Korea Transport Database (KTDB) [8].

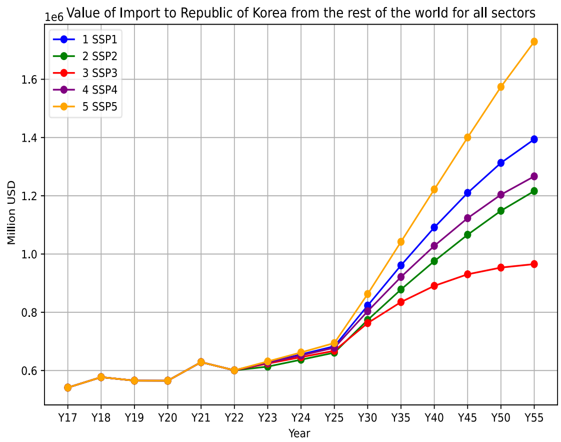


Fig. 5. Value of Import to the Republic of Korea from the rest of the world for all sectors

The following figures Fig. 6 and Fig. 7 compare the trade volume for the Republic of Korea, the calculation is based on the accumulated percentage change in the quantity of bilateral trade from the GTAP-RD model and the bilateral trade volume for each commodity expressed in tons for both 2017 and 2023 as a base year of the calculation.

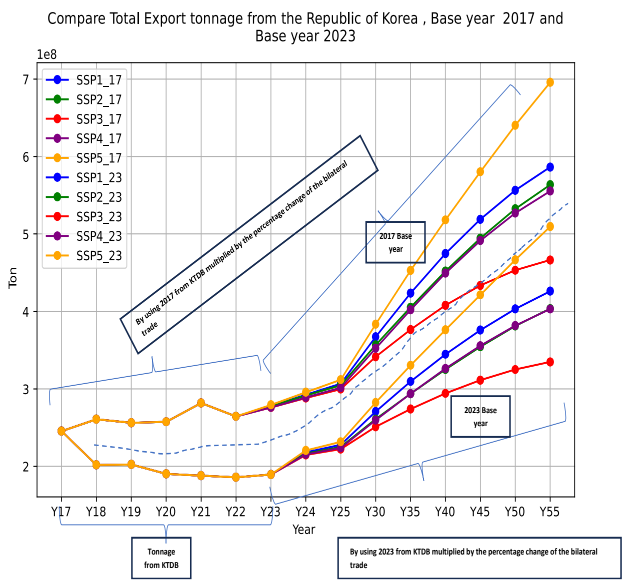


Fig. 6. Compare Total Export tonnage forecasting to the Republic of Korea Base year 2017 and Base year 2023

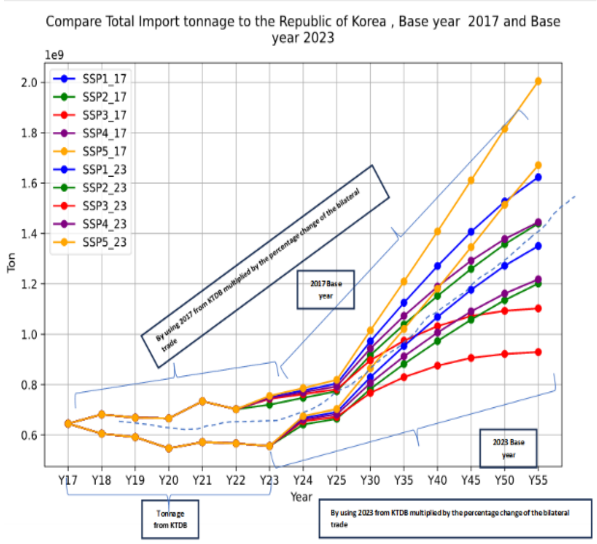


Fig. 7. Compare Total Import tonnage forecasting to the Republic of Korea, Base year 2017 and Base year 2023

4. Conclusion

Analyzing the global trade with the Republic of Korea is highly important for forecasting transportation demands not only for international trades with other countries but also for optimization of the domestic transportation methods. We focused on the top trading partners for the Republic of Korea, which was compiled from the following countries: China, the United States of America, Japan, Vietnam, Australia, Germany, Taiwan Province of China, Saudi Arabia, India, Singapore, and so on. Additionally, the assessment of the aggregation of goods and services was carefully considered and analyzed according to the relationship between the Republic of Korea and those countries. Overall, the GTAP analysis centered on the Republic of Korea was successfully done for the careful consideration of the specific economic dynamics and trade relationships within related countries, as well as the appropriate methodological tools and model adjustments for assessing the potential impacts of trade policy changes.

Acknowledgments

The authors would like to thank Professor Ryuichi Shibasaki for their invaluable suggestions in forecasting international trade amounts with respect to the GTAP framework and Dr Trang Tran for technical support and assistance related to GTAP experimental methods. This work was supported in part by the cooperative research project on digital logistics between the Kyushu Institute of Technology and Logistics Revolution Korea Co., Ltd

References

1. A. Aguiar, M. Chepeliev, E. Corong, and D. Van Der Mensbrugghe, "The Global Trade Analysis Project (GTAP) Data Base: Version 11," J. Glob. Econ. Anal., vol. 7, no. 2, pp. 1–37, Dec. 2022, doi: 10.21642/JGEA.070201AF.
2. Thomas Warren Hertel, Global Trade Analysis. Cambridge University Press, 1997.
3. E. Ianchovichina and T. L. Walmsley, Eds., Dynamic Modeling and Applications for Global Economic Analysis, 1st ed. Cambridge University Press, 2012. doi: 10.1017/CBO9781139059923.
4. A. Aguiar, Corong, Erwin, and van der Mensbrugghe, Dominique, "The GTAP Recursive Dynamic (GTAP-RD) Model: Version 1.0," Present. GTAP-RD Model GTAP Conf. Organ. Sess. On Improving Baseline Proj. GTAP-Recursive Dyn. RD Model Framew, Feb. 2019.
5. M. Li, "The U.S.- China trade war: tariff data and general equilibrium analysis. https://www2.econ.iastate.edu/faculty/balistreri/Papers/crts_trade_war.pdf.
6. K. Riahi et al., "The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview," Glob. Environ. Change, vol. 42, pp. 153–168, Jan. 2017, https://doi.org/10.1016/j.gloenvcha.2016.05.009.
7. Horridge J.M., Jerie M., Mustakinov D. & Schiffmann F. (2018), GEMPACK manual, GEMPACK Software, ISBN 978-1-921654-34-3.
8. KTDB, "교통수요분석네트워크," Ktdb.go.kr, 2020. https://www.ktdb.go.kr/www/contents.do?key=25 (accessed Dec. 10, 2024).
9. Group: GTAP-RD Data Aggregation Utility," Mygeohub.org,2020 . https://mygeohub.org/groups/gtap/dynamic-data-agg-utility (accessed Dec. 10, 2024).
10. Data.imf.org. https://data.imf.org/?sk=9d6028d4f14a464ca2f259b2cd424b85 (accessed Dec. 10, 2024).

Authors Introduction

Ms. BoYoung Lee

She received her Bachelor's degree in Human Relations from Keio University and M.S. in logistics from Inha University in Korea. She is currently a Ph.D. student. She works as a logistics consultant in the Logistics Revolution Korea Co., Ltd.

Mr. Ahmad Altaweel

He received his bachelor’s degree in Electronics and Communication Engineering, from the Faculty of Electrical and Mechanical Engineering, ALBAATH University, Syria. He is currently a Ph.D. student at Kyushu Institute of Technology, Japan.

Mr. Kazuhito Mine

He is a bachelor's student in the School of Computer Science and Systems Engineering, Kyushu Institute of Technology, Japan

Dr. JangSok Yoon

He received his M.S. and Ph.D. degrees from Kyunghee University, in South Korea, in 2004 and 2008. He is the founder, CEO, and Consultant over 23 years of Logistics Revolution Korea Co., Ltd.

Dr. Hiroaki Wagatsuma

He received his M.S., and Ph.D. degrees from Tokyo Denki University, Japan, in 1997 and 2005, respectively. He is currently an Associate Professor at Kyushu Institute of Technology.

Fundamental Research on Athlete Positions Estimation in Indoor Sports at Various View

Iori Iwata
Graduate School of Sustainable of Systems Science, Komatsu University, Komatsu, Ishikawa 923-8511, Japan
Email: 24311001@komatsu-u.ac.jp

Yoshihiro Ueda
Faculty of Production Systems Engineering and Sciences, Komatsu University, Komatsu, Ishikawa 923-8511, Japan
Email: yoshihiro.ueda@komatsu-u.ac.jp

Kazuma Sakamoto
Faculty of Production Systems Engineering and Sciences, Komatsu University, Komatsu, Ishikawa 923-8511, Japan
Email: kazuma.sakamoto@komatsu-u.ac.jp

Riku Kaiba
Graduate School of Sustainable of Systems Science, Komatsu University, Komatsu, Ishikawa 923-8511, Japan
Email: 24211007@komatsu-u.ac.jp

Abstract

In recent years, data collection for tactical analysis in sports has become increasingly prevalent. In sports such as volleyball, basketball, and soccer, where player positioning is closely linked to scoring opportunities, research has been conducted to visualize player positions using various technological approaches. Notably, numerous research has focused on enhancing tactical analysis by estimating player positions through image recognition methodologies. These approaches typically rely on images captured by one or more cameras. From these images, specific reference points on the court are identified and transformed into a bird's-eye view using image transformation algorithms such as projective transformation to visualize player positioning. This process requires the selection of four reference points on the court, preferably encompassing the entire playing area. However, capturing these four reference points from ideal viewing angles is often infeasible in many venues. Additionally, live game images frequently feature zoomed-in views of players and shifts in camera angles as the ball is tracked, creating challenges for consistent analysis. These limitations restrict tactical analysis to videos specifically recorded for such purposes, excluding archival video not originally intended for research and thus limiting data diversity. To address these challenges, this research aims to develop a system that broadens the applicability of tactical analysis by utilizing video captured in stadiums where a part of views is feasible, as well as historical video data. Focusing on volleyball, the proposed approach automatically identifies reference points based on the coordinates of the net and court lines to estimate player positions. This system seeks to enable robust and efficient analysis across diverse video sources, enhancing the scope and utility of tactical insights.

Keywords: volleyball, player position, bird's-eye view, projective transformation

1. Introduction

In recent years, Japan has introduced the Third Basic Plan for Sports [1], which is to be implemented over a five-year period from FY2022 to FY2026. The plan places an emphasis on leveraging advanced technologies and digital transformation (DX) [2] to effect revolutionary change in the domain of sports. In this context, considerable effort has been made in the field of sports to acquire player position data for the purpose of supporting tactical analysis [3], [4], [5]. This practice is particularly prevalent in sports such as soccer, volleyball, and basketball, where player positioning is closely tied to tactical strategies and scoring opportunities. In these sports, a considerable number of teams employ dedicated analysts, alongside coaches and managers, to track player positions at every moment and conduct in-depth tactical analysis.

In volleyball, an indoor sport, data analytics software such as DataVolley [6] is widely used to assist in tactical analysis. This software, which is primarily utilized by professional teams, requires input such as player positions, play details, and back numbers. By accumulating and analyzing this data, the software provides actionable data that contributes to tactical decision-making. However, it is necessary to manually input the data, and as a result, the adoption has been delayed in amateur teams. Recently, advancements in technology have enabled the estimation of player positions using camera images or other sensors, allowing for easier visualization of player movements. Typically, this involves creating a bird's-eye view and plotting player positions after estimating them from camera footage. Projective transformation, a common method used for this purpose, requires at least four reference points. Existing research [1], [2] use the four corners of the court as reference points. Nevertheless, this methodology restricts the scope of investigation to a predefined angle of view, necessitating adjustments whenever the angle undergoes a change. Furthermore, conventional techniques often rely on video footage encompassing the entire court, which limits their scope of application.

Therefore, this research aims to create a player position estimation system that can handle various angles of view.

2. Methodology

(Fig. 1) illustrates the research process, which is represented by a flow diagram. The input is an image of a volleyball player, and the output is a bird's-eye view. The input images are processed using YOLOv9[7] (hereinafter referred to as YOLO), a type of object recognition algorithm, to recognition players. The lines and the net in the image are visually inspected, and their coordinates are manually obtained. The coordinates are obtained by visually locating the lines and nets in the image and tracing them with the mouse. Next, the number of line intersections is calculated, and reference points for projective transformation are determined based on this number. The established reference points are employed to transform the coordinates through a projective transformation, thereby generating a bird's-eye view. In this study, the YOLO algorithm is employed to identify individuals from match videos and ascertain their locations within the images. The individuals are represented by bounding boxes, as illustrated in (Fig. 2), and the upper left and lower right coordinates of the square are extracted. In this case, the upper left coordinate is (0,0), and thus, the coordinate value calculated from Eq. (1) is utilized as the individual's position within the image.

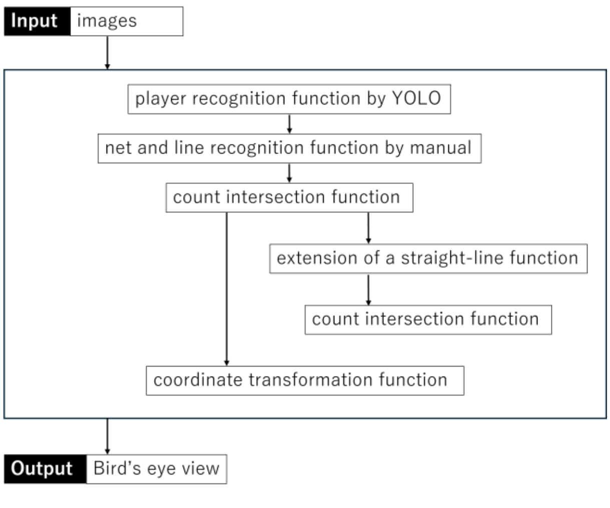


Fig. 1 Flow diagram

$$\begin{cases} x = (x_{min} + x\_max)/2 \\ y = y\_max \end{cases} \quad (1)$$

Fig.2 Person detection with YOLOv9

2.1. *Method of reference points selection for various view*

First, the coordinates of the four reference points for the projective transformation are obtained. In this research, a linear equation of the volleyball court line is obtained. In this case, depending on the angle of view, only a portion of the line may be used, but this is not a problem because the line is extended for use. In such cases, the intersection can be found by extending the lines. After obtaining multiple court lines on the image, the intersections are calculated. After acquiring several court lines in the image, the intersection points are calculated. In general, there can be anywhere from 0 to 6 intersections within an image.

If there are six intersections, as illustrated in (Fig. 3), the entire court is reflected and the coordinates of the four corners of the court can be obtained, so they are used as they are. If there are five or fewer intersections, the straight-line equation of the acquired court lines is extended within the image area as shown in the dotted line in (Fig. 4), and further intersections are searched for. If a new intersection is detected, as indicated by the blue dot in (Fig. 4), the point in question should be selected.

Fig.3 Taken image with virtual court line

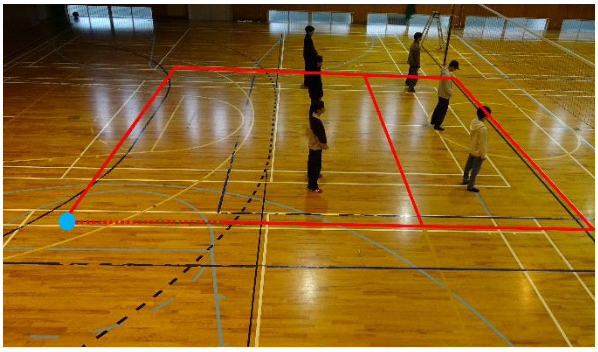


Fig.4 Creating intersection by extension

After this operation is completed, the number of intersections is checked. When the number of intersections is less than four, as illustrate in (Fig. 5), the players are hardly seen in the image and it is difficult to use the image for tactical analysis, so "Please change the angle of view" is represented to encourage the viewer to change the angle of view.

Next, the reference point is established based on the number of intersections identified. In this investigation, the angular perspectives captured within the blue frame in (Fig. 3), do not encompass the net and, thus, were excluded from the tactical analysis. As a result, these angles were excluded from research. When there are four or five intersections, the selected reference points are among A, B, E, and F in (Fig. 3). The selection process is carried out accordingly. The point closest to the straight line of the net and to the left of the net is A, the other is F, the point closest to A that is not F is B, and the point closest to F that is not A is E. When there are six points of intersection, A, B, E, and F are defined in the same way, and the point on the straight-line AB is C and the other is D.

Fig.5 Example with 4 or less intersections

2.2. *Creating a bird's eye view*

Using the coordinates obtained in section 2.1 as a reference point, the projective transformation is used to transform the image into the bird's-eye view shown in (Fig. 6) to estimate the position of the players. The projective transformation is performed with reference to [8]. The number of intersections has three patterns from Section 2.1: four, five, and six. For four or five of them, convert A, B, E, F in (Fig. 3) to A', B', E', F', in (Fig. 6) for six, convert A, C, D, F to A', C', D', F' when there are six.

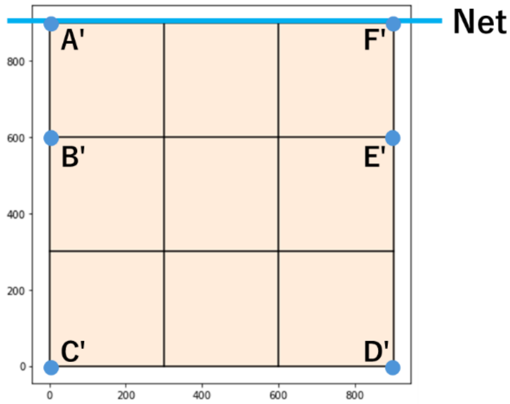


Fig.6 bird's-eye-view

3. Experimental Details

This experiment was conducted in the gymnasium at the Awazu Campus of Komatsu University. As shown in (Fig. 7), multiple cameras were installed to capture images from various angles. Examples of images taken from each angle are presented in (Fig. 8). In this experiment, to obtain the correct position of the players, although videos are typically used to estimate player positions, this experiment utilized still images to evaluate accuracy. The ground truth position of each player on the court was measured using a tape measure and defined as the correct position. The positional arrangement of the players during this experiment is illustrated in (Fig. 9). The error between the player positions estimated by the system and the ground truth positions was calculated and used for evaluation.

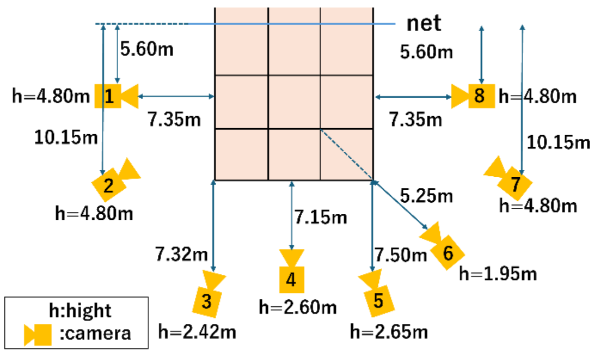


Fig.7 Shooting method

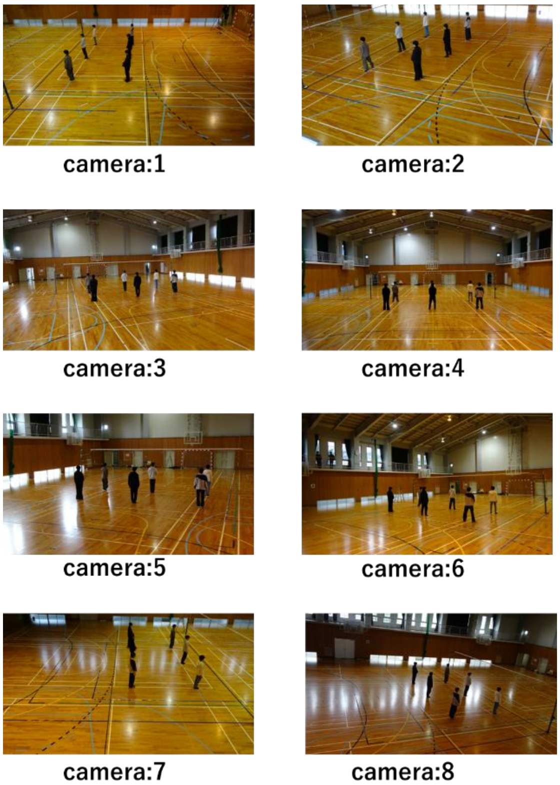


Fig. 8 Images taken from each camera angle of view

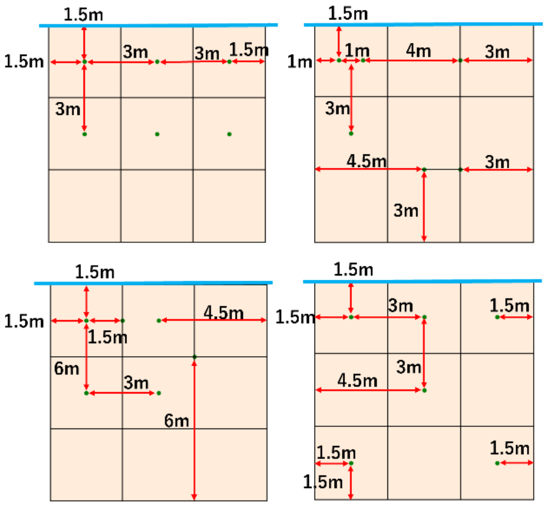


Fig.9 All positions pattern

4. Results and Discussion

The results of this research are summarized in (Table 1). (Table 1) presents a subset of data, displaying the coordinates and ground truth positions of 351 successful plots. (Fig. 10) shows an example of the output bird's-eye view. The red dots (Fig. 10) are the player positions estimated by the system, and the green dots are the correct positions. The line connecting (0,900) and (900,900) is the net position. The x-axis direction in this figure is the net horizontal direction, and the y-axis is the net vertical direction. These plots were derived from a total of 64 images captured by cameras 1 through 8. It is important to note that plots located outside the court were excluded from this experiment. This exclusion was intended to avoid plotting coaches, managers, spectators, or other individuals not directly involved in the game.

Table 1. Error with the correct position

	x1	y1	x2	y2	dx	dy	distance
1	359	209	349	190	0.13	0.34	0.36
2	677	219	676	191	0.02	0.50	0.50
:	:	:	:	:	:	:	:
351	361	383	349	354	0.22	0.52	0.56
avg.					0.21	0.21	0.30

unit: x1, y1, x2, y2(pix) dx, dy(meter)

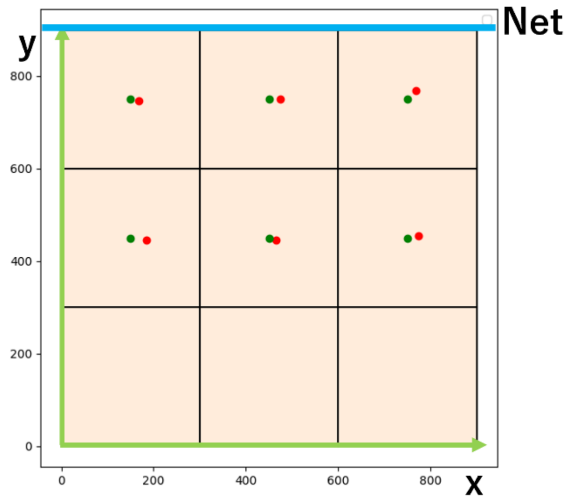


Fig. 10 An example of a bird's-eye view

As (Table 1) shows, the average error of the 351 players in this experiment was 0.21 meters in the x-coordinate, that is, in the direction of transition to the net, and 0.21 meters in the y-coordinate, that is, in the vertical direction of the net. The distance of the error is 0.30 meters. Since the shoulder width of a typical male is between 0.4 meters and 0.45 meters, we consider this accuracy to be sufficient for tactical analysis. As shown in (Fig. 11), the bounding box can accurately surround the frame of the person in the YOLO detection. If a person is surrounded by a frame that is too large, the error is the result of the deviation from the original frame.

(Table 2) summarizes the accuracy from each camera angle. The analysis revealed that the footage captured by cameras 3, 4, and 5, which are positioned near the rear of the court, exhibited larger average errors in the y-direction compared to the x-direction displacement across the entire court. In the image, the x-axis corresponds to the depth in the real world. Since this study uses monocular player position estimation, the error in the y-direction is larger than in the x-direction. Camera 4, which captures the court directly from behind, was found to have the highest error, likely due to the inherent difficulty in accurately estimating player positions along the y axis from this angle. Conversely, accuracy improved as the angle of view approached a top-down perspective, as seen with cameras 1 and 2. However, camera 8, despite being positioned at the same height as cameras 1 and 2, did not achieve similar levels of accuracy. (Fig. 12) shows an image captured by Camera 8, where YOLO algorithm was employed to identify the players. Additionally, (Fig. 13) represents the bird's-eye view at the same time. In this experiment, slight discrepancies in the camera's field of view were observed. For instance, while Camera 1 captured the feet of occluded players, allowing for high-accuracy player positioning, Camera 8 did not capture the feet of the player in the center of the left column, as shown in (Fig. 12). Instead, the bottom edge of the bounding box was displayed around the player's waist. If the bounding box had enclosed the feet, it would have been possible to estimate the player's position more accurately. This discrepancy is thought to have contributed to the error in Camera 8's field of view.

This reduction in YOLO's detection accuracy caused by player occlusion likely contributed to the overall decrease in experimental accuracy. However, when examining images from Camera 4's field of view at the same moment, the two players occluded in Camera 8's field of view were successfully enclosed by bounding boxes. These findings suggest that leveraging multiple fields of view could enable the creation of a system more robust to occlusions.

Table 2. Error of each camera angle of view (meters)

camera's No.	dx	dy	distance
1	0.19	0.10	0.21
2	0.11	0.20	0.23
3	0.08	0.21	0.22
4	0.08	0.40	0.41
5	0.22	0.31	0.38
6	0.14	0.22	0.26
7	0.25	0.08	0.26
8	0.35	0.16	0.38

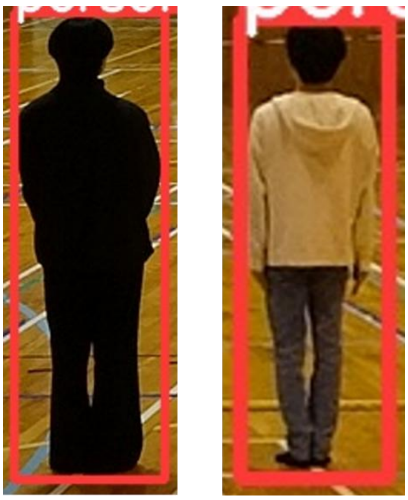


Fig. 11 Example of people detection

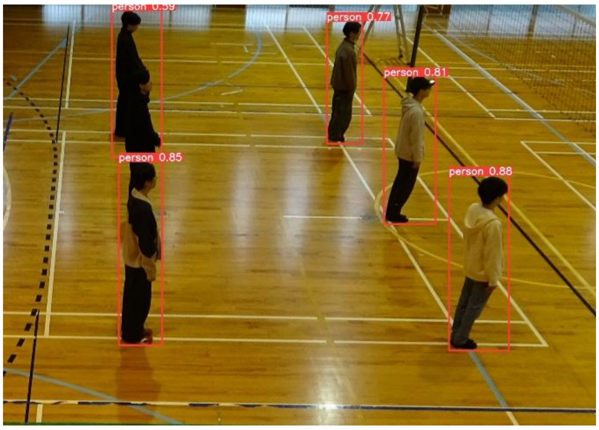


Fig 12 Player Occlusion

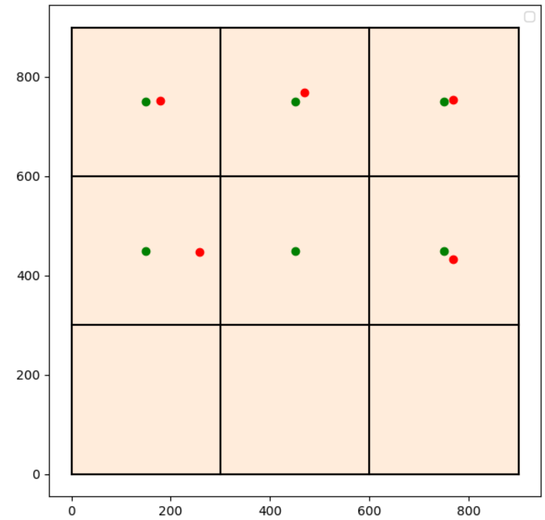


Fig. 13 Estimation of player position at occlusion

Next, we compared the results based on the number of intersections detected in the images. (Table 3) shows the average error when estimating the results of performing a projective transformation using A, B, E, and F as the reference points, and when performing a projective transformation using A, C, D, and F as the reference points. The results indicate that differences in error were observed along the y-axis. Since this comparison does not involve variations in detection accuracy by YOLO, the observed error is considered to arise during the projective transformation process. Generally, when using projective transformation to map shapes, accuracy tends to improve as the area of the quadrilateral formed by the reference points increases. Furthermore, points located inside the quadrilateral defined by the reference points are expected to have higher transformation accuracy compared to those located outside the quadrilateral. These two factors are considered to have contributed to the observed differences in accuracy.

Table 3. Difference in reference point and error

Reference point for the perspective transformation	dx	dy	distance
A, B, E, F	0.21	0.24	0.32
A, C, D, F	0.21	0.15	0.26

5. Conclusion

In this research, we aimed to broaden the scope of tactical analysis by estimating player positions from various viewing angles, rather than relying solely on a fixed angle of view. As detailed in Section 4, the overall error in this experiment was 0.30 meters. Although there is a difference of about 0.2 meters in error depending on the angle of view, it is possible to create a bird's-eye view even if the image was taken at an arbitrary angle of view. This result demonstrates that the proposed method can generate bird's-eye views from diverse angles, thereby contributing to tactical analysis. However, certain viewing angles may result in player occlusion, which can impede accurate player detection by YOLO. Furthermore, this experiment assumed that players remained on the ground; when players are airborne, such as during jumps, the estimation error is expected to increase. To address these limitations, future work will focus on developing a system that integrates multiple viewpoints to enhance robustness against occlusion and enable high-precision position estimation for airborne players. Such advancements would significantly improve the system's applicability for tactical analysis across a broader range of scenarios. Then, the detection of lines and nets is performed by an object recognition algorithm, aiming for full automation.

References

1. Japan Sports Agency, The Third Sport Basic Plan, https://www.mext.go.jp/sports/b_menu/sports/mcatetop01/list/1372413_00001.htm
2. Ministry of Internal Affairs and Communications, 2021 White Paper on Information and Communication Definition of Digital Transformation, https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r03/html/nd112210.html
3. C.-C. Hsu, H.-T. Chen, C.-L. Chou, and S.-Y. Lee, 2D Histogram-based player localization in broadcast volleyball videos, Multimed. Syst., vol. 22, no. 3, Jun (2016), pp. 325–341.
4. G. Tsiotas, DEEP LEARNING MODELS FOR THE AUTOMATIC ANALYSIS OF VOLLEYBALL PLAYS,
5. J. Theiner, W. Gritz, E. Muller-Budack, R. Rein, D. Memmert, and R. Ewerth, Extraction of Positional Player Data from Broadcast Soccer Videos, in 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA: IEEE, Jan (2022), pp. 1463–1473.
6. DataProject, DataVolley,
7. C.-Y. Wang, I.-H. Yeh, and H.-Y. Mark Liao, YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information, arXiv, 2024.
8. Y. Luo et al., A Review of Homography Estimation: Advances and Challenges, Electronics, vol. 12, no. 24, Jan (2023).

Authors Introduction

Mr. Iori Iwata

He received the M.S. degree with the Graduate school of Sustainable Systems Science from Komatsu University, Japan, in 2024, where he is currently pursuing the Ph.D. degree at the Graduate school of Sustainable Systems Science. His research interests include image recognition, tactical analysis, and sports informatics.

Prof. Yoshihiro Ueda

He received the Ph.D. degree in Mathematical Information Science from Kanazawa University, Japan, in 2001. Since 2021, he has been a professor at the Faculty of Production Systems Engineering and Sciences, Komatsu University, Japan. His research interests include productivity improvement, labor saving, automation using data science, and human centric system development.

ASST. Kazuma Sakamoto

He received the B.S., M.S., and Ph.D. degrees in Informatics from Kansai University, Japan, in 2015, 2018, and 2021, respectively. Since 2021, he has been an assistant professor at the Faculty of Production Systems Engineering and Sciences, Komatsu University, Japan. His research interests include web mining, natural language processing, image processing, and sports informatics.

Mr. Riku Kaiba

He received the B.S. degree in Production Systems Engineering and Sciences from Komatsu University, Japan, in 2024, where he is currently pursuing the M.S. degree at the Graduate school of Sustainable Systems Science. His research interests include computer vision and sports informatics.

Overview of the Development of Low Earth Orbit Satellite Navigation Enhancement Technology

Dingcheng Tang [1]*, Jinliang Wang [1], Jianfeng Shan [1], and Guoji Zou [1]
Space Star Technology Co., LTD, Beijing, 100086, China

*Email: *alwaystdc@163.com*

wangjltust@163.com

zougj@spacestar.com.cn

shanjf@spacestar.com.cn

Abstract

Low Earth Orbit (LEO) satellites, with their unique advantages in constellation and signal, are gradually gaining attention and favor in the world's satellite navigation field, and are expected to become a new increment in the development of the new generation of satellite navigation systems. With the low orbit constellations represented by the Starlink program in the United States becoming a new battlefield for global competition for space strategic resources, the development direction of satellite navigation enhancement is also gradually tilting towards the low orbit field, becoming a new growth and empowerment point for the next generation of satellite navigation. This article first introduces two major types of low orbit navigation enhancement technologies, namely information enhancement and signal enhancement; Summarized the current development status of global low orbit satellite navigation; Explained the key technologies for enhancing low orbit navigation; The focus was on discussing the new development opportunities brought by low orbit navigation enhancement technology for building a Global Navigation Satellite Systems (GNSS) global space-based monitoring network, providing global quasi real time high-precision services, and providing global high integrity monitoring services; Finally, the inspiration and suggestions for the development path of low orbit navigation enhancement technology were summarized.

Keywords: Low Earth Orbit Satellite Navigation Enhancement, LEO constellation, GNSS, navigation and positioning

1. Introduction

The Global Navigation Satellite System (GNSS) currently includes China's BeiDou Navigation Satellite System (BDS), the United States' Global Positioning System (GPS), and Russia's GLObal Navigation Satellite System, GLONASS) and the European Union's Galileo system, a total of more than 100 satellites are in orbit, providing global, all-weather, and high-precision services for all walks of life. For the satellite navigation system of medium and high orbit constellation, its satellite signal is very weak when it reaches the earth's surface from 20,000~30,000km space, and it is easy to be shielded and blocked in cities, canyons, jungles and other areas, and cannot serve the underground garage and other environments. At present, all major satellite navigation countries are aiming for higher service accuracy, more diverse functions, and more reliable services, and are starting to build a new generation of systems and a new round of competition.

With its unique advantages of constellation and signal, LEO satellite has gradually attracted the attention and favor of the world's satellite navigation field, and is expected to become a new increment in the development of a new generation of satellite navigation system. LEO satellites can enhance satellite navigation signals as an augmentation and supplement to GNSS (Global Navigation Satellite System); It can also be fused by the communication system and the navigation system to broadcast independent ranging signals to form a backup positioning and navigation capability. At present, in the field of international satellite navigation, the research and development and practice of how to apply low-orbit satellite technology to achieve PNT (positioning, navigation timing) enhancement, backup and supplement are in the ascendant. Iridium and GPS jointly developed and launched a new satellite timing and positioning service (STL), which has become a backup or supplement to the GPS system. The European Galileo system technical team is also actively promoting the research of Kepler system, through the low-orbit constellation composed of 4~6 low-orbit satellites, through the inter-satellite link to monitor and high-precision measurement of medium- and high-orbit satellites, so as to greatly improve the orbit determination accuracy of the Galileo constellation. In addition, the development of domestic low-orbit satellite technology was also in full swing during the same period. Under the promotion of relevant departments, large central enterprises, research institutes, and private enterprises, satellite in-orbit tests have been carried out in low-orbit satellite constellations such as Hongyan, space-ground integrated network, and micro-space.

2. Low-orbit navigation augmentation technology

Satellite navigation augmentation technology was first developed in the 90s of the 20th in response to the US GPS Selective Availability (SA) policy. Scholars have proposed a technology for differential processing between stations, which restores the C/A ranging code of GPS to the system design accuracy by eliminating the common measurement error between stations [1], [2], [3]. In the 21st century, with the abolition of the SA policy, the connotation of navigation augmentation technology has been further expanded, which refers to various technical solutions used to improve the service capability of satellite navigation systems.

As the low-orbit constellation represented by the Starlink program of the United States has become a new battlefield for the global competition for space strategic resources, the development direction of satellite navigation enhancement is also gradually inclined to the low-orbit field, which is becoming a new growth and empowerment point for the next generation of satellite navigation. The low-orbit satellite navigation augmentation technology can improve the accuracy, availability and integrity of satellite navigation and positioning, and it can be divided into two categories: information enhancement and signal enhancement according to the enhancement mode.

Information augmentation refers to a technical way to improve the accuracy and reliability of navigation and positioning by correcting the errors of satellite navigation and positioning systems. Information augmentation does not provide additional distance observations, only corrected information and integrity information to eliminate GNSS errors. Information augmentation usually requires a transmission channel that can broadcast the enhanced information to the user. According to the platform mode of enhanced information transmission, it can be divided into ground-based augmentation and satellite-based augmentation. Network RTK, satellite-based differential and other technologies are typical information augmentation systems.

Signal augmentation refers to the method of transmitting navigation signals through platforms other than navigation satellites, and users can also receive navigation signals from the navigation satellite system itself and other additional navigation signals, thereby improving the accuracy and usability of navigation and positioning. Signal enhancement typically requires a signal transmitter to provide the user with measurement information. The signal augmentation system provides observational measurements and can be used in conjunction with GNSS or independently. According to the position of the transmitting navigation signal platform, signal enhancement can also be divided into ground-based enhancement and satellite-based enhancement.

Information augmentation can improve GNSS positioning performance, but it is not effective in situations where GNSS signals are not received, such as indoors or in occluded environments. Signal enhancement can provide new observations and measurements, and can be used to compensate for scenarios where GNSS is powerless, such as under overpasses, urban canyons, etc., when the number of satellites may not be enough, and even for indoor areas where GNSS signals cannot reach, signal enhancement can also provide a solution.

3. Global development status of low-orbit satellite navigation

The construction of the next-generation Iridium satellite system in the United States was completed in 2019, and the new low-orbit satellite timing and positioning services provided by it have become an effective supplement and backup to GPS. The European Galileo technical team is also actively promoting the research of Kepler system, and 4~6 low-orbit satellites monitor and measure the operation of medium and high orbit satellites, so as to greatly improve the orbit determination accuracy of the system.

In order to improve the capabilities of the GPS system, the United States has integrated the development of the GPS system on the new generation of Iridium system [4], providing users with STL (Satellites Time and Location) service, which is provided by Satelles, which can backup and enhance the GPS capability.

The Iridium system was originally a global mobile communications system designed by the American company Motorola. It consists of 6 orbital planes, each with 11 satellites evenly distributed in each orbit, forming a complete constellation that can cover global regions including the polar regions. At the beginning of the design, it mainly provided users with satellite communication services, but due to the high cost of system construction and the cold reception of the market, the development of the Iridium system was relatively slow. In January 2019, the United States completed the launch and deployment of a new generation of Iridium satellite system. In addition to the communication service, the Iridium constellation can achieve positioning, navigation, and timing in indoor and canyon areas by providing STL services. STL service performance, positioning accuracy of 30~50 meters, timing accuracy of about 200ns, the original signal landing power is 300~2400 times stronger than the GPS L1 C/A code signal (24.8~33.8dB), indoor usability is greatly improved, and navigation availability and safety in complex terrain environment and complex electromagnetic environment are enhanced. With Iridium and GNSS receivers, STL services can be used to augment multi-GNSS navigation services, including GPS, and can also be used as a backup when signals are unavailable or inadequate.

In 2015, SpaceX proposed a plan to build Starlink, which will launch and deploy about 42,000 small satellites in low-earth orbit to achieve all-weather, low-latency high-speed Internet access services around the world. After the construction is completed, the Starlink constellation will undoubtedly be the best performing and largest low-orbit satellite constellation to date.

The European Galileo system technical team proposed to greatly enhance the capabilities of the Galileo system with the Kepler low-orbit system, which can enhance the integrity and accuracy of the Galileo system while reducing the dependence on the ground system. The technical core of the Kepler system is to use 4~6 low-orbit satellites to form a small-scale constellation, and laser inter-satellite link (ISL) to complete the existing constellation system. MEO satellites do not need to be equipped with atomic clocks, and all satellites are connected through laser inter-satellite links, so that navigation satellites can achieve direct synchronization at a very high precision level, and then further provide high-precision distance measurement instead of pseudorange for orbit determination, in order to obtain mm-level orbit determination accuracy and nm-level phase measurement accuracy, so as to achieve a significant increase in capabilities. At the same time, the system observes the navigation signals without ionospheric and tropospheric disturbances through the LEO satellite constellation, which can improve the integrity and accuracy of the MEO system. Under the Kepler system architecture, most of the survey and communication facilities of the ground operation and control system are no longer necessary, and only a small number of ground stations are required to maintain consistency with the earth's coordinate framework and the ability to control the system under special circumstances.

In 2019, the Russian State Space Corporation proposed to add 12 satellites to the existing "Messenger-D1M" low-orbit communication satellite constellation composed of 12 "Messenger-M" satellites, and launched three and six messenger satellites respectively at the end of 2019 and 2020, and launched a total of six satellites in two phases in 2021~2022, completing the networking of 24 messenger satellites. In 2019, Roscosmos also proposed to start building the "Messenger-2" low-orbit narrowband communications satellite constellation for the Internet of Things and mobile satellite communications in 2024, which is planned to consist of 28 satellites with an orbital altitude of 1,500km.

A number of domestic units have also carried out theoretical research, simulation calculation and in-orbit satellite verification of low-orbit satellite augmentation, and proposed corresponding constellation plans. Communication constellations such as "Hongyan" [5], "Hongyun" and "Space-Earth Integrated Information Network" all consider the needs of low-orbit satellite augmentation, and the micro-space and arrow brigade images focus on low-orbit high-precision enhancement. At the same time, the on-orbit technical tests of low-orbit test satellites such as "Hongyan Constellation Test Satellite", "Luojia-1", "Weili Space" and "Netcom-1" have accumulated experimental data for low-orbit satellite navigation signal enhancement technology and precision enhancement technology. At this stage, the "Nebula System" built by China is the second phase of the national satellite Internet project, which mainly provides communication and navigation services for civil and commercial users. Nebula Systems plans to launch 4 test satellites in September 2024, complete in-orbit verification in December 2024, complete the launch network of a total of 1,800 benchmark constellations by 2027, and complete the launch of a total of 3,600 constellations with an expanded configuration by 2030, with an orbital altitude of 508km, and its constellation configuration is shown in the figure below (Fig. 1).

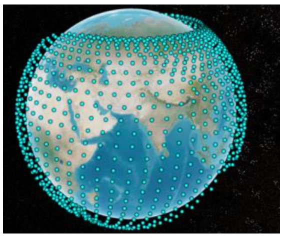


Fig. 1 Constellation diagram of the nebula system.

4. Key technologies for low-orbit navigation enhancement

The low-orbit navigation augmentation system [6] consists of three parts: the space segment, the ground segment and the user segment. The space segment is composed of dozens to hundreds of low-orbit satellites equipped with navigation enhancement payloads, and its main task is to broadcast navigation signals and enhanced information of high, medium and low-orbit navigation satellites to various users, and has the functions of forwarding satellites and navigation satellites. The ground segment includes a ground operation control system and a ground monitoring station to jointly complete the operation management and control of satellites in orbit, and the user segment includes various types of user terminals, modules, chips and supporting equipment. From the current concept stage to the actual operation of the business system in the future, a series of key technologies need to be broken through in the space segment, the ground segment and the user segment.

4.1. *Space Segments*

The design and optimization of constellation configuration is the primary problem that must be solved in the construction of space segment, which directly determines the cost, coverage performance, geometric strength and service capability of the augmentation system. The large number of low-earth orbit satellites is prone to multiple access interference, and due to the widespread use of L-band in multi-frequency GNSS, the signal in this band is becoming increasingly congested, and the problem of signal interference is becoming more and more serious. Therefore, it is necessary to design a new signal system with certain anti-interference performance, develop new navigation frequency bands (such as S-band and C-band), and study the technologies of spread-spectrum code optimization, signal modulation, capture and tracking, channel coding and multiplexing suitable for low-orbit navigation enhancement signals.

How to maximize the advantages of the low-orbit constellation itself, and realize the integration of navigation and communication is also the key. It is now possible to use mobile communication networks to broadcast GNSS augmentation messages. It can be considered that satellite navigation and mobile communications have achieved partial integration. Because the low-orbit navigation enhancement will mainly rely on the Internet constellation in the future to promote, the future will further strengthen the deep integration of navigation and communication at the information level and even the signal level, through the joint design of the navigation enhancement signal and the mobile communication signal, all the power of the navigation signal is allocated to the pilot component, which is conducive to the tracking, acquisition and ranging of the receiver, and the message data can be quickly broadcast through the low-orbit Internet communication.

In terms of the configuration of satellite payloads, it is necessary to rely on the low-orbit Internet constellation platform, and select some satellites to achieve multiple uses of one satellite, and carry out scientific research such as navigation augmentation, remote sensing, satellite gravity, satellite altimetry, occultation detection, and satellite-based GNSS-R by carrying different payloads. Due to the low space ionizing radiation in the low-orbit region, commercial off-the-shelf (COTS) devices and chip-scale atomic clocks (CSACs) can be used to reduce costs.

4.2. *Ground Segments*

The ground operation control part also faces some challenges. Due to the special dynamic characteristics of low-orbit satellites, they are different from traditional medium- and high-orbit GNSS satellites in terms of fitting time, update frequency, and number of parameters of ephemeris and clock parameters. Compared with the basic navigation data, the effective time of the augmentation correction number is short and the timeliness is strong, which is divided into fast and slow correction parameters, and the parameter characteristics of the augmentation information must be analyzed, the relationship between the accuracy index of the fitting parameters and the broadcast delay is obtained, and the real-time requirements of each augmentation parameter are obtained. The low-orbit navigation message information and its orchestration mode, fast and slow message parameters and other information need to be redefined according to the reserved information bits, which mainly depends on the coordination between the system advertising capability and the real-time requirements of parameter broadcasting.

Low-orbit satellites have fast operating speed, short transit time, shortened information transmission time, and small ground coverage, so more ground injection stations or inter-satellite link communication are needed for message transmission. If inter-satellite communication is adopted, it should form a hybrid constellation with medium- and high-orbit satellites, and the optimal choice of communication mode between low-orbit satellites and medium- and high-orbit satellites is a problem to be demonstrated and solved. An important task in the ground segment is to establish and maintain the spatiotemporal datum of the low-orbit navigation augmentation constellation. The fusion and processing of GNSS and low-orbit augmented data needs to be completed under a high-precision and unified spatio-temporal framework. For the time system, it is necessary to give the time definition of the low-orbit navigation augmentation system, and carry out the establishment and maintenance of the system time, the internal time synchronization of the system, the time traceability and the time difference forecast. For the coordinate system, it is necessary to give the method of definition, implementation, update, maintenance and conversion of the coordinate system of the low-earth navigation augmentation system with other coordinate systems. The unification of space-time datums is inseparable from GNSS and low-orbit constellation precision orbit estimation clocks.

In order to obtain the real-time precision orbit of all low-orbit satellites, the ground main control station can adopt a variety of orbit determination strategies, including GNSS and low-orbit constellation joint precision orbit determination and orbit prediction under the condition of regional monitoring stations, orbit determination of low-orbit constellation using only the global tracking station network without relying on GNSS, inter-satellite link orbit determination, etc., due to the large number of satellites and the large amount of calculation, the single station observation arc is short, and there is an unstable and discontinuous situation of satellite entry and exit in local areas. Therefore, it is necessary to analyze and demonstrate the influence of different observed arc lengths on the orbit determination accuracy, and develop a high-efficiency distributed and parallel processing algorithm. The most difficult model of orbit determination and orbit prediction is the solar pressure model, which is mainly affected by the factors such as satellite platform and satellite attitude, and it is necessary to obtain the corresponding light pressure reflection coefficient for different types of low-orbit satellites, study the parameter change characteristics of the dynamic model based on the solar pressure reflection coefficient, and establish a high-precision mathematical model related to satellite attitude and solar position, so as to improve the accuracy of precision orbit determination and medium and long-term forecasting.

In terms of satellite-ground time maintenance and synchronization, due to the influence of factors such as power consumption, size, weight and cost of high-performance spaceborne atomic clocks, it is not suitable for the low-cost requirements of low-orbit satellites, and CASC or OCXO are generally used as substitutes for time maintenance. Or some satellites are equipped with high-performance atomic clocks, and the rest of the satellites achieve time synchronization through inter-satellite and satellite-to-ground communication.

The ground control component is also responsible for low-orbit navigation, enhanced constellation management and satellite control. These include: the development of long-term and medium-term mission plans to ensure global coverage and continuity of signals; maintenance of the overall constellation of satellites; Sufficient replenishment of satellites as planned; Dealing with unexpected events and failed satellites to minimize their impact on services; Monitor and control the status of each satellite in all aspects, ensure the normal operation of the satellite and payload, and deal with some unexpected events when unexpected events occur; Plan and execute track scheduling, platform maintenance, etc.; Support on-orbit software maintenance.

4.3. *User Segments*

The enhancement of low-orbit navigation puts forward new requirements for the software and hardware equipment and data processing methods of ground users. Under the high dynamic conditions of low-orbit satellites, the receiver signal acquisition is more difficult, the Doppler search range is larger, and the acquisition speed is reduced. In a short period of time, the distance in the direction of the station satellite connection changes greatly, which makes the received signal strength change greatly, resulting in uneven changes in pseudorange noise, affecting the accuracy of code observations, and the RF front-end of the receiver must also change. At the same time, the acceleration in the direction of the station-satellite connection line will also change greatly, resulting in a large change in the carrier frequency and code phase, which is easy to cause signal lock-up, greater Doppler prediction uncertainty and shorter coherent integration time requirements, which brings challenges to the high-precision and stable tracking of signals. The number of low-orbit satellites is large, and the observation arc is short, and the satellite switching is frequent, so it is necessary to increase the number of channels, storage capacity, and microprocessor computing power of the receiver, and optimize the antenna unit and the receiving unit. For low-orbit navigation, S or C band enhancement signals may be used, and L+S or L+C integrated antennas need to be designed, and the corresponding RF, analog-to-digital conversion, baseband signal processing, capture, tracking, demodulation and other methods need to be adjusted.

In terms of data processing, more stringent quality control algorithms should be designed and developed to diagnose anomalies in satellite orbits and satellite clock products. Signal lock-up and frequent satellite handover may produce more cycle slips, while low-orbit satellites operate fast and the ionospheric changes between epochs are large, so traditional cycle slip detection algorithms such as the ionospheric residual method will no longer be applicable, and more effective algorithms need to be studied. The mathematical model of HFO satellite fusion positioning needs to be further refined, because the new navigation augmentation signal will introduce more deviations related to the system, orbit type, code type and frequency, etc., and the time and space characteristics of various deviations need to be carefully analyzed and modeled, and they should be corrected or estimated in subsequent data processing. In addition, the reasonable weighting of observations of multi-source heterogeneous constellations is also a problem that needs attention.

5. Opportunities presented by low-orbit navigation augmentation technology

Compared with medium- and high-orbit navigation satellites, including Beidou and GPS, low-orbit satellite navigation signals will bring new development opportunities for the construction of GNSS global space-based monitoring networks, the provision of global quasi-real-time high-precision services, and the provision of global high-integrity monitoring services due to their unique orbit and signal characteristics.

5.1. *LEO constellations and signals have unique advantages*

1) Low-orbit satellites have low orbits and small weights, and the cost of satellites and launches is low. Low-orbit satellites are lighter in weight and lower in orbit than medium- and high-orbit satellites, and can be

launched with multiple satellites with one arrow, and the cost of satellite research and development and rocket launch is lower.

2) The landing signal strength is higher, which can improve the positioning effect and improve the usability under the condition of occlusion

The orbital altitude of low-orbit satellites is generally about 1000km, and compared with medium- and high-orbit navigation satellites with an altitude of more than 20000km, the signal transmission path of low-orbit satellites is shorter, and the signal delay and power loss are smaller. To put it simply, if a low-orbit satellite and a medium- and high-orbit satellite transmit the same signal power, the signal power transmitted by the low-orbit satellite to the earth's surface will be 30dB (i.e., 1000 times) higher than that of the medium- and high-orbit satellite. Stronger landing signal power can improve the positioning effect in complex terrain environments and complex electromagnetic environments, and enhance anti-interference and anti-spoofing capabilities.

3) The low-orbit satellite runs fast, accelerates the high-precision positioning convergence time, and provides a better user experience

The geometric configuration of the satellite constellation in medium and high orbit changes slowly, and the correlation between the observation equations between adjacent epochs is too strong, so it takes a long time to estimate and separate various errors when estimating positioning parameters, so as to fix the carrier phase ambiguity and achieve precise positioning. Therefore, the convergence time of traditional high-precision positioning is generally 15 minutes ~ 30 minutes.

However, the time for low-orbit satellites to orbit the earth is much smaller than that of medium- and high-orbit satellites, and the trajectories of low-orbit satellites in the same period of time are longer and the geometric configuration changes rapidly. Theoretically, the operation of a low-orbit satellite for 1 minute is equivalent to the geometric change of the current medium-orbit satellite for 20 minutes. The orbital characteristics of low-orbit satellites help to accelerate the convergence time of high-precision positioning, reaching 1 minute convergence, and the user experience is more excellent (Fig. 2).

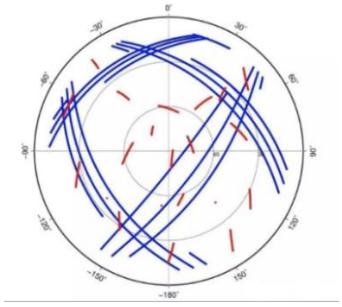


Fig. 2 Trajectories trajectories of low-orbit and medium-high orbit satellites in the same time (red is the trajectory of medium- and high-orbit satellites, and blue is the trajectory of low-orbit satellites).

4) Higher information rate, more precise correction information can be broadcast

Due to the increase of landing signal power, the LEO enhanced signal can carry a higher information rate or a larger signal bandwidth, and can be used as an broadcast channel for satellite navigation basic messages and differential correction messages.

5) The terminal is miniaturized, integrated, low power consumption, and easy for users to use

The improvement of low-orbit enhanced signal power is conducive to the use of smaller terminal equipment by ground users; At the same time, when used as a communication, ground users can be normally received by low-orbit satellites with a smaller signal power.

5.2. *Establish a GNSS global space-based monitoring network*

Generally speaking, GNSS requires global distribution of ground monitoring stations for observation support, and the United States, Russia, and Europe basically adopt the approach of global station construction to meet the requirements of global continuous observation. Most of the GPS monitoring stations in the United States are located near the equator, including five monitoring stations in Colorado, Diego Garcia, Ascension, Cavagalin, and Hawaii. Russia has a large east-west span, which can basically solve the global observation problem of the GLONASS system; The European Galileo system can be used to establish stations in overseas colonies and achieve global observations. At present, China's Beidou system is mainly based on domestic station construction, and realizes global observation and operation support through inter-satellite links.

Based on the low-orbit satellite constellation, it can help build a GNSS global space-based monitoring network.

1) Achieve high-quality global monitoring of satellite navigation signals

Compared with the ground monitoring network, the observation of navigation signals by low-orbit satellites has the characteristics of less influence by ionosphere and troposphere, long tracking arc, more coverage times, and less influence of multipath effect. As a high-precision space-based monitoring station of GNSS, low-orbit satellites can greatly improve the observation geometry, weaken the correlation between tangential orbit and phase ambiguity, improve the accuracy of navigation satellite orbit and clock error, and effectively make up for the shortcomings of ground-based monitoring network in space coverage, so as to achieve global high-quality monitoring.

2) It can effectively improve the accuracy of satellite orbit determination

By selecting 30 navigation satellites, the orbit determination capabilities of only ground monitoring stations, low-orbit satellite-based monitoring and ground joint monitoring were simulated and analyzed. Only the ground monitoring station carries out orbit determination, and the space signal accuracy is about 0.3 meters, and the space-based monitoring station composed of 12 low-orbit satellites has a signal accuracy of about 0.1 meters. Preliminary analysis shows that the ground-based and space-based joint monitoring and orbit determination based on the low-orbit constellation can improve the orbit determination accuracy of the constellation by more than one times.

5.3. *Provide global quasi-real-time high-precision services*

1) Diversification of high-precision positioning service providers and hierarchical service models

At present, high-precision positioning services (precision single-point positioning, PPP) mainly include commercial PPP services and PPP services embedded in satellite navigation systems.

The main commercial services are the Global Differential GPS (GDGPS) system for satellite orbit determination, scientific research and high-end commercial services from the Jet Propulsion Laboratory (JPL) in the United States, the StarFire system from Navcom, the OmniSTAR system and RTX system from Trimble, the StarFix/SeaStar system from Fugro, and the C-Nav system from Oceaneering International. Hexagon's Veripos system and TerraStar system include meter-level, decimeter-level, centimeter-level and other service capabilities.

In recent years, with the increasing demand for high-precision applications in the industry and the public, Galileo in Europe, QZSS in Japan and Beidou in China have also designed and provided PPP services embedded in satellite navigation systems. Galileo provides free PPP services based on E6B signals, with an advertising rate of 500bps, and enhances the GPS and Galileo systems to achieve decimeter-level positioning. QZSS precision positioning is divided into two categories: sub-meter level enhanced service (SLAS) and centimeter-level enhanced service (CLAS), both of which are free services, which are provided by L1 and L6 signals respectively, of which the CLAS service broadcast rate reaches 2000bps, which can enhance the four major GNSS and QZSS five systems at the same time. China's Beidou system uses GEO satellites to provide high-precision positioning free services to users in China and surrounding areas, with a broadcast frequency of B2b and a rate of 500bps, to achieve real-time decimeter-level and post-event centimeter-level positioning.

2) The precision enhancement of low-earth orbit satellites has the characteristics of fast convergence, which can improve the high-precision user experience

At present, the PPP service provided by GNSS through medium and high orbit satellites usually has a convergence time of 15~30 minutes due to its slow orbit change. Although Japan's QZSS system uses PPP-RTK technology to shorten the convergence time of PPP services to less than 1 minute, it requires the support of a large number of highly dense ground monitoring stations, and thousands of monitoring stations need to be built within the territory of Japan alone. The low-orbit satellite has the ability of global GNSS high-precision monitoring, and the geometric change is large, which is easy to achieve high-precision positioning and rapid convergence. Relevant studies show that after the integration of low-orbit satellite constellation, the convergence time of satellite navigation system PPP service can be shortened to less than 1 minute, and users can obtain high-precision services in quasi-real-time.

5.4. *Provide global high integrity monitoring services*

1) At present, each satellite-based augmentation system mainly provides regional integrity services

Integrity service refers to the timely warning to users when navigation satellites fail and risks, so as to improve the user's ability to use safety, which is particularly important for users involved in life safety such as civil aviation.

At present, the satellite-based augmentation systems (SBAS) mainly include the FAA Wide Area Augmentation System (WAAS), the European

Geostationary Satellite Navigation Overlay Service (EGNOS), China's Beidou Satellite-based Augmentation System (BDSBAS), the Russian Differential Correction and Monitoring System (SDCM), the Japanese Multifunctional Satellite Augmentation System (MSAS), the Indian GPS-assisted Geostationary Orbit Augmentation and Navigation System (GAGAN), the South Korean Satellite Augmentation System (KASS), etc. The differential correction number is broadcast, and the GNSS accuracy and integrity of a certain area are enhanced. The coverage of each system is shown in the figure belo w (Fig. 3).

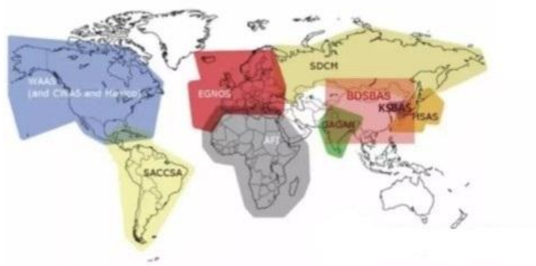


Fig. 3 SBAS Service Providers and Coverage

2) Higher-quality space-based integrity monitoring capabilities

Through the GNSS global space-based monitoring network of the low-orbit constellation, the space-based monitoring of the integrity of navigation satellites can be realized, and the orbital characteristics of the low-orbit constellation make it not affected by the ionosphere and troposphere, and the multipath impact is smaller than that on the ground, which can improve the monitoring ability of the integrity of navigation satellites.

3) Integrity alarm information can be broadcast to the world

In the future, user demand for SBAS services will not only be satisfied with the region, but will be expanded globally. The global coverage of the LEO constellation gives it a natural ability to broadcast global integrity services.

6. Enlightenment and suggestions on the technical path for the enhanced development of low-orbit navigation

In view of the good development prospects of the low-orbit satellite low-orbit navigation augmentation system, the number of low-orbit satellite navigation systems currently planned has exceeded the number of GNSS systems. Each system has different ways to implement navigation enhancements. In the future, the field of low-orbit satellite navigation will face fierce competition, which puts forward high requirements for the construction speed and service level of China's low-orbit satellite navigation system. Foreign low-orbit satellite navigation systems have begun to have service capabilities and have been applied in specific industries and scenarios. At present, most of China's domestic low-orbit satellite navigation systems are in the stage of demonstration and in-orbit verification, and there is still a gap between technology and management and formal service provision.

In view of the construction cost and technical difficulty of the system, the current low-orbit satellite navigation system is mainly enhanced, supplemented by backup mode, and the navigation performance can be improved and backed up with fewer satellites. The enhancement mode adopts the mode of "information + signal enhancement", and comprehensively improves the service capability of the system by a variety of means. Considering the cost and technical risks of system construction, the LEO satellite independent navigation system is still in the stage of academic and technical research, and has the potential to be built simultaneously with the deployment of the giant LEO constellation.

Based on the research on the construction status of low-orbit satellite constellation at home and abroad, combined with China's national conditions, the development route of domestic low-orbit satellite navigation system is proposed.

1) For space segments

Step 1: Achieve single-weight coverage of the earth through 60~150 low-orbit satellites. Low-orbit satellites can be special navigation satellites, or low-orbit communication and remote sensing satellites equipped with standardized navigation enhancement payloads. At present, this stage of construction has been started, and a number of units have put forward construction plans in response.

Step 2: Dual coverage of the earth through 150~250 low-orbit satellites, with independent positioning capabilities, can be used as a supplementary backup system for GNSS. At the same time, it is capable of joint service with GNSS.

Step 3: More than 400 low-orbit satellites achieve multiple coverage of the earth, and the satellites do not rely on the use of low-cost on-board atomic clocks, intelligent large-scale operation and maintenance and other key technologies, reducing system construction and operating costs.

2) For the ground segment

The ground section mainly includes GNSS/low-orbit monitoring stations, GNSS/low-orbit data processing centers, and other data transmission and business control and other reuse with existing ground facilities. The construction of ground stations makes full use of existing resources, and realizes 10 domestic monitoring stations through new or renovated ones, realizes high-precision pseudo-code and carrier phase measurement and collection, and supports high-precision orbit determination and clock deviation determination.

3) For user segments

According to the number of visible Beidou satellites and low-orbit satellites, the terminal adopts the integrated design of conduction, comprehensively collects observation information, independently selects the positioning and calculation method, and supports multi-source fusion positioning, and the solution modes include GNSS independent positioning, low-orbit independent positioning, GNSS and low-orbit combined positioning, and auxiliary Beidou positioning in complex electromagnetic environment.

7. Conclusion

With the full completion of the four major navigation systems, the major countries in satellite navigation applications are also planning and deploying the development of a new generation of navigation systems, and actively looking for new capabilities to grow. With its advantages of high signal power, large geometric dynamics and strong anti-interference ability, the low-orbit constellation can form information enhancement and signal enhancement for GNSS on a global scale, and comprehensively improve the accuracy, integrity, continuity and availability of satellite navigation systems, which will bring new development opportunities for the construction of GNSS global space-based monitoring network, global quasi-real-time high-precision services and global high-integrity monitoring services. It is expected to solve the problems of global coverage, low landing power, long initialization time and low reliability of the current augmented system, and serve high-value security users such as power grids, banks, securities, and military in the future, as well as real-time precision positioning users represented by autonomous driving and drones. With the development of 5G/6G technology, the establishment of mobile communication networks in the air, space, ground and sea, and the improvement of the processing capacity of mobile terminals such as smart phones, the enhancement of low-orbit navigation is finally expected to enter thousands of households and realize mass application. The enhancement of low-orbit navigation will also be an important part of China's comprehensive PNT system. The integration of various satellites with different orbits will bring new changes and new developments to the world's satellite navigation, and the low-orbit constellation will become a new increment in the development of a new generation of satellite navigation systems because of its unique advantages.

References

1. Zeng Tian, Sui Lifen, Jia Xiaolin, et al. Fengyun 3C enhanced Beidou orbit determination test results and analysis[J]. Journal of Surveying and Mapping, 2017, 46(7): 824-833.
2. Wang Lei, Chen Ruizhi, Li Deren, etc Signal Quality Evaluation of Luojia-1 Low Earth Orbit Satellite Navigation Enhancement System [J]. Journal of Wuhan University (Information Science Edition), 2018, 43 (12): 2191-2196.
3. Ge, Haibo, Bofeng Li, Maorong Ge, Nan Zang, Liangwei Nie, Yunzhong Shen and Harald Schuh. Initial Assessment of Precise Point Positioning with LEO Enhanced Global Navigation Satellite Systems (LeGNSS) [J]. Remote. Sens. 10 (2018): 984.
4. Joerger M，Gratton L，Pervan B，etal．Analysis of iridium-augmented GPS for floating carrier phase positioning[J]．Navigation，2010，57(2):137-160．
5. Shi C，Zheng F，Lou Y，etal. National BDS augmentation service system (NBASS) of China: progress and assessment[J]. RemoteSensing，2017，9(8):1-16.
6. Guo Shuren, Liu Cheng, Gao Weiguang, etc Construction and Development of Satellite Navigation Augmentation System [J]. Global Positioning System, 2019, 44 (2): 1-12.

Authors Introduction

Mr. Dingcheng Tang

He received his M.S. degree in electronic science and technology from Huazhong University of Science and Technology, China, in 2008. In 2008, he joined Space Star Technology Co., LTD, Beijing, China, where he is a senior engineer. His research interests include global navigation satellite system, integrated PNT system, and anti-jamming technology of satellite navigation.

Mr. Jinliang Wang

He received his M.S. degree in instrument science and technology in Tianjin University of Science and Technology, China, in 2022. In 2022, he joined Space Star Technology Co., LTD, Beijing, China, where he is an engineer. His research interests include global navigation satellite system, integrated PNT system, and anti-jamming technology of satellite navigation.

Mr. Jianfeng Shan

He received his master's degree in optical engineering from Tianjin University, in 2014. In 2014, he joined Space Star Technology Co., LTD, Beijing, China, where he is a senior engineer. His research interests include satellite navigation signal processing technology, anti-jamming signal processing technology and high-precision navigation technology.

Mr. Guoji Zou

He received his M.E. degrees in software engineering from Beihang University. He is working in Space Star Technology Co., LTD, Beijing, China, where he is a senior engineer. His main research direction is satellite navigation system signal generation and performance evaluation.

Signal Decomposition and Noise Reduction in Single-Channel EEG: A Morphological Component Analysis (MCA) Approach

Kosei Shibata
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Yide Yang
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Rena Kato
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hendry F. Chame
University of Lorraine, CNRS, LORIA, F-54000, France

Laurent Bougrain
University of Lorraine, CNRS, LORIA, F-54000, France
Sorbonne Université, Institut du Cerveau – Paris Brain Institute -ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, 75013 Paris, France

Hiroaki Wagatsuma
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Email: sibata.kousei927@mail.kyutech.jp, yang.yide421@mail.kyutech.jp, kato.rena844@mail.kyutech.jp, hendry.ferreira-chame@loria.fr, laurent.bougrain@loria.fr, waga@brain.kyutech.ac.jp

Abstract

This study applies the Morphological Component Analysis (MCA) to single-channel EEG data obtained during human-to-human interactions in a board game (Hex-game). MCA, a dictionary-based signal decomposition method, separates signals into distinct morphological components. It enables the extraction of plausible brain activity and the removal of noise, such as ocular and muscular artifacts. By focusing on neural dynamics in interactive settings, this approach highlights the relationship between cognitive processes and social behavior. The approach suggests that MCA offers a promising framework for EEG analysis in complex, dynamic environments, combining effective feature extraction with robust artifact removal.

Keywords: EEG, Morphological Component Analysis, MCA, Signal Decomposition, Signal Denoising, Human Interaction

1. Introduction

The analysis of electroencephalogram (EEG) signals in dynamic and complex environments is inherently challenging due to the presence of noise, such as ocular and muscular artifacts, and the need to extract meaningful neural activity [1], [2]. Morphological Component Analysis (MCA) has gained attention as an effective method for addressing these challenges [3]. Utilizing a dictionary-based decomposition approach, MCA separates EEG signals into distinct morphological components, enabling the removal of noise while preserving critical neural dynamics. This dual capability makes MCA a valuable tool for applications requiring precise feature extraction, such as the study of social cognition and interactive behaviors.

This study focuses on exploring the influence of joint attention (JA) on decision-making strategies and engagement within the framework of a board game scenario (Hex-game). JA is a fundamental aspect of human social cognition, allowing individuals to align their attention with others on shared objects or activities [4]. It plays a key role in facilitating shared experiences, coordinated actions, and effective cooperation, serving as a cornerstone of human-to-human interaction [5] , [6], [7].

2. Methodology

2.1. *Experiment setup*

Hex-game is a competitive game between two players on a board of 11x11 hexagonal squares. Each player has colored hexagonal pieces in red (player A) or blue (player B). Each player alternately places a piece anywhere on an unoccupied square of the board. The first player to form a connected path of their pieces linking the opposing sides of the board marked by their color wins (Fig.1).

Eight healthy male adults (aged 21–41 years) participated in this experiment. All participants provided informed consent prior to the experiment, which was approved by the ethics committee of Kyushu Institute of Technology.

Players were fitted with a 16-electrode electroencephalographic headset and Tobii glasses. The experimental setup included simultaneous measurements with EEG, eye-tracking, and a ceiling-mounted camera, enabling the synchronized acquisition of brain activity, gaze patterns, and game progression data (Fig.2).

Players were instructed to synchronize their moves with the strong audible beat of a 50 BPM 4/4 rhythm. Each pair of players played six Hex-game matches. Each match lasted approximately 2–3 minutes. Rest periods of one minute were provided between matches to minimize fatigue(Fig.3).

EEG signals were recorded using a 16-channel system in accordance with the international 10-10 system, with a sampling rate of 512 Hz.

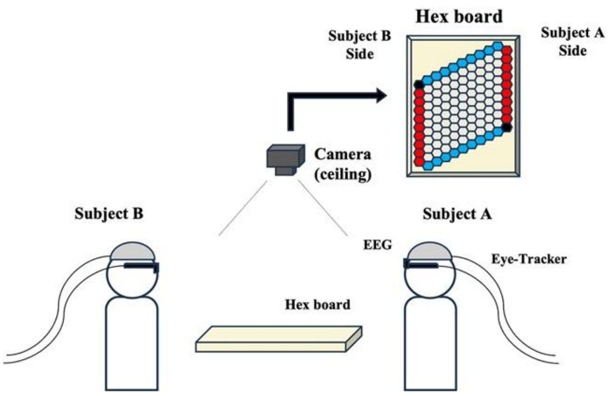


Fig.1 Hex-game Experiment Overview

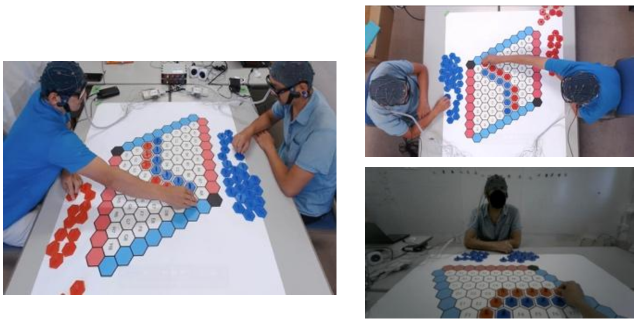


Fig.2 Hex-game experimental setup. Left: Setup overview. Top-right: Ceiling-mounted camera point of view. Bottom-right: Tobii glasses point of view.

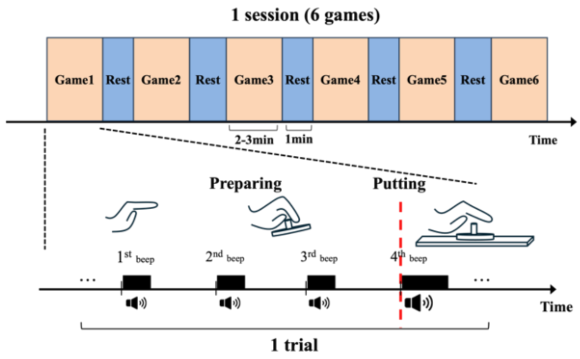


Fig.3 Timing. Top: playing and resting sequence. Bottom: Time sequence of audio instruction to place a piece

2.2. *Morphological Component Analysis*

Morphological Component Analysis (MCA) is a powerful signal decomposition technique that separates signals into distinct morphological components using principles of sparsity and redundant transforms or overcomplete dictionaries [8]. Unlike traditional methods such as Principal Component Analysis (PCA) or Independent Component Analysis (ICA), MCA is particularly well-suited for single-channel data analysis. It achieves effective noise removal while preserving the critical features of the original signal, making it an invaluable tool for applications requiring precise feature extraction and noise mitigation.

EEG signals, characterized by their specific morphological patterns—such as spikes, slow waves, and oscillatory components—are ideal candidates for MCA's sparsity-based decomposition. By employing tailored dictionaries, MCA can effectively isolate and analyze individual signal features (Fig.4):

Undecimated Wavelet Transform (UDWT): Captures smooth, low-frequency components, useful for detecting periodic activity.

Discrete Sine Transform (DST): A mathematical transformation that represents a signal in terms of sine functions. It is particularly effective in capturing spectral features with specific boundary conditions and is commonly used for signal compression and analysis in the mid-to-high frequency range.

Dirac: Identifies sharp spikes, aiding in the detection of transient events such as epileptic discharges.

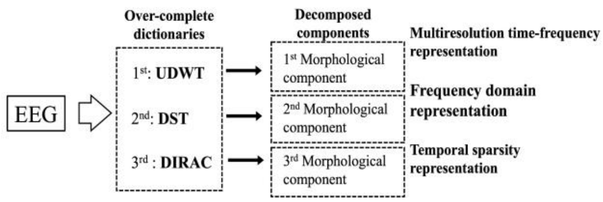


Fig.4 Overview of Morphological Component Analysis (MCA) applied to EEG signal

MCA offers several advantages for EEG analysis, 1.**Noise Resilience:** MCA effectively removes artifacts such as eye movement (EOG) and muscular activity (EMG), enhancing signal quality.
2.**Feature Extraction:** It isolates distinct morphological features, facilitating the classification of EEG states such as ictal, interictal, and non-seizure conditions.
3.**Single-Channel Capability:** Unlike ICA or PCA, MCA does not require multi-channel data, making it applicable to simpler setups.
The utility of MCA extends beyond noise removal; it enables a deeper understanding of the underlying neural dynamics by decomposing EEG signals into their constituent components. Studies have demonstrated that MCA outperforms ICA in both artifact removal and signal decomposition, particularly in single-channel scenarios. Moreover, its application in EEG signal analysis has proven effective in improving classification accuracy for machine learning models, such as Support Vector Machines (SVM), by providing well-separated, high-quality features.

2.3. *EEG data preprocessing*

Prior to processing the EEG data, a high-pass filter with a cutoff frequency of 1 Hz was applied for eliminating low-frequency noise. Then a down sampling from 512 Hz to 128 Hz has been applied to reduce the computational burden associated with EEG data analysis.

2.4. *Sliding window approach*

To ensure that transient features are effectively captured and reduce computation time, a decomposition method has been applied to individual 100-second segments with an overlap of 50 seconds.
After the splitting step, the MCA method has been applied on each 100-second segment (see 2.5) to obtain three 100-second components, one per dictionary
Finally, to preserve signal continuity and integrity, a Hann window has been applied on each 100-second component, and the Overlap-Add (OLA) method was applied to combine overlapping components for seamless signal concatenation.
The complete components extracted from each dictionary were subsequently combined to reconstruct the signal.

2.5. *MCA dictionaries*

Subsequently, MCA was applied to the 100-second segmented C3 channel data within a single game, utilizing three dictionaries (UDWT, DST, Dirac). As all participants are right-handed, C3 contains important information on cortical motor activity during the game.

2.6. *Reconstruction metric*

The reconstructed signal is the sum of the three dictionaries. Part of the original signal not corresponding to one of these three waveforms is eliminated. The Spearman’s correlation between the filtered and reconstructed EEG signal is computed to evaluate the quality of the decomposition.

3. Results and Discussion

Fig.5 shows the decomposition of a filtered EEG signal in C3 during an entire session of 6 Hex-games. Ocular artifacts are mainly captured by the UDWT library. The main cortical activity is with the DST component. The Dirac component also seems to reflect artifact activity. The Spearman’s correlation is high showing that all three selected dictionaries are adequate to represent the information contained in the EEG signal.
Fig.6 shows as an illustration the result of a MCA decomposition (Pair 2, Match 3, Player A, C3). In this figure, we can see that Dictionary 1 (UWDT) perfectly extracts elements correlated with player A's activity phase prior to his movement materialized by the go signal (or 4th strong beat) trigger.

Fig.7 shows that the first dictionary (UWDT) captures mainly activity in the delta band (0-4 Hz) which probably corresponds to eye movement artefact. The second dictionary (DST) shows activity in the alpha band corresponding to movement execution. And the third dictionary (Dirac) shows noise and especially the power activity at 60 Hz.

A more thorough examination of the low-frequency components (Fig.8) reveals that the second component, DST, shows notable activity in the 2-6 Hz range, potentially capturing theta waves (4-7 Hz). In contrast, delta waves (0.5-4 Hz) appear to be absorbed into the first component, mixed with slow movement-related activity. The extraction of alpha waves (8-13 Hz) and beta waves (above 13 Hz), however, remains unconfirmed at this stage [9].

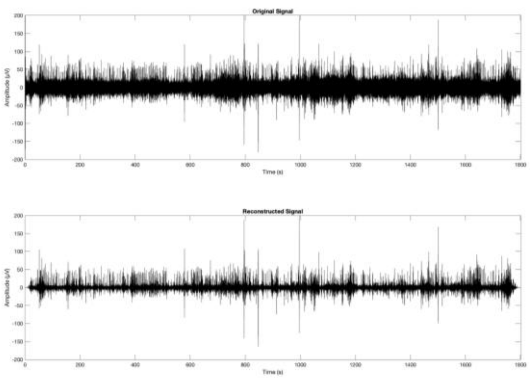


Fig.5 Comparison (corr=0.740) of original and MCA Reconstructed signals using three dictionaries (UDWT, DST, and Dirac) for an entire session of six games (Pair 2, Player A, C3).

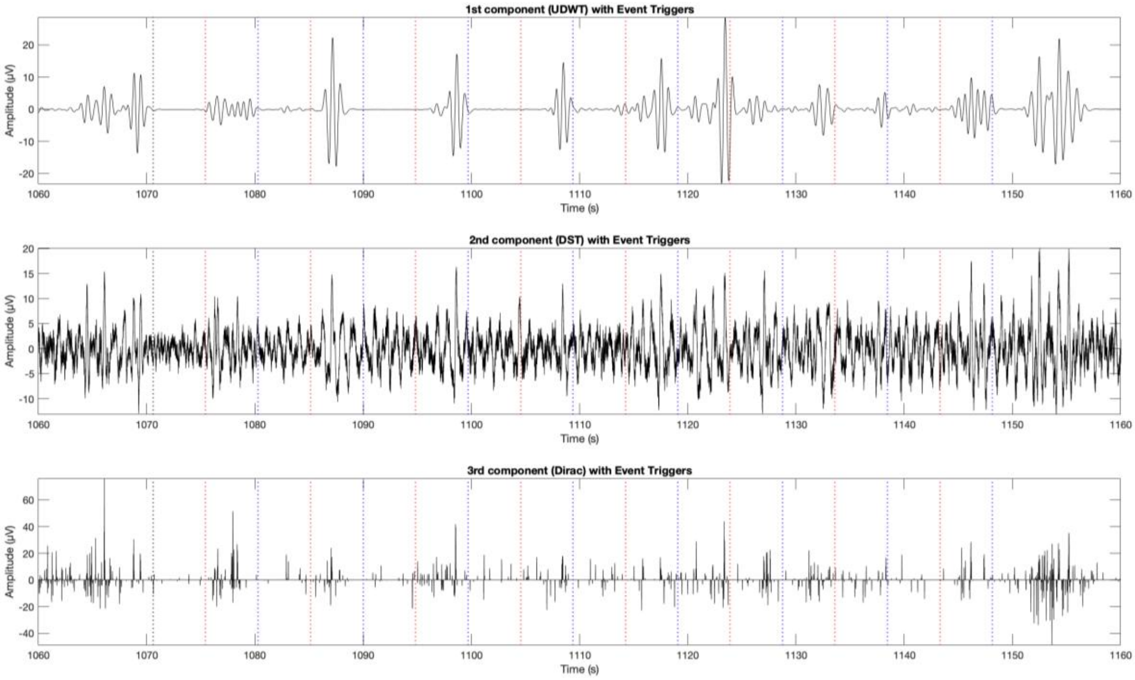


Fig.6 Results of MCA decomposition with event labels (Pair 2, Match 3, Player A, C3). Red and blue events correspond to Go signals to place a piece on board for respectively player A and player B.

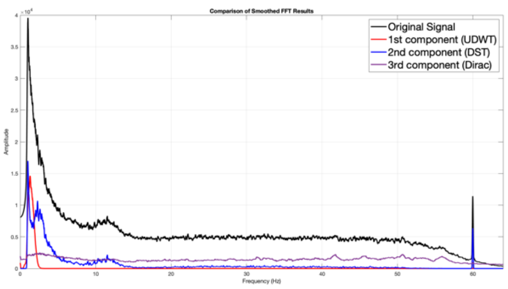


Fig.7 FFT Analysis of the smoothed original signal and components decomposed by MCA-dictionaries (UDWT, DST, and Dirac) for Pair 2/Player A-C3.

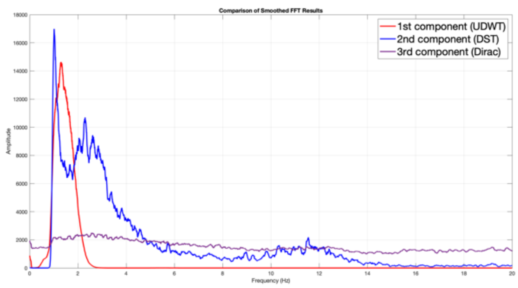


Fig.8 FFT Analysis (0-20 Hz) of the smoothed original signal and components decomposed by MCA-dictionaries (UDWT, DST, and Dirac) for Pair 2/Player A-C3.

4. Conclusion

By applying MCA to the EEG data collected during the Hex-game experiment, the method successfully isolated neural activity associated with JA while effectively removing noise artifacts, thereby enhancing signal quality. The results highlight the potential of MCA as a robust framework for EEG analysis in social and interactive contexts. This work not only deepens our understanding of social cognition in human interactions but also contributes to the advancement of adaptive communication systems in Human-Robot Interaction (HRI), particularly in dynamic and complex environments.

Acknowledgements

This work was supported in part by JSPS KAKENHI (JP17H06383, JP24K07387), and the joint research project: Collaborative project Université de Lorraine – Kyutech "Human/Robot Social Interactions: engagement and affect analysis during gaming tasks."

References

1. R. Sameni and C. Gouy-Pailler, "An iterative subspace denoising algorithm for removing electroencephalogram ocular artifacts," Journal of Neuroscience Methods, vol. 225, pp. 97–105, Mar. 2014.
2. J. A. Urigüen and B. Garcia-Zapirain, "EEG artifact removal—state-of-the-art and guidelines," Journal of Neural Engineering, vol. 12, no. 3, 031001, Apr. 2015.

3. J.-L. Starck, Y. Moudden, J. Bobin, M. Elad, and D. L. Donoho, "Morphological component analysis," in Proceedings of SPIE 5914, Wavelets XI, San Diego, CA, USA, Sep. 2005, vol. 5914, p. 59140Q.
4. G. Butterworth, "The ontogeny and phylogeny of joint visual attention," in Natural Theories of Mind: Evolution, Development, and Simulation of Everyday Mindreading, A. Whiten, Ed. Oxford, UK: Basil Blackwell, 1991, pp. 223–232.
5. D. C. Richardson and R. Dale, "Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension," Cognitive Science, vol. 29, no. 6, pp. 1045–1060, 2005.
6. P. Mundy and M. Sigman, "Joint attention, social competence, and developmental psychopathology," in Developmental Psychopathology, 2nd ed., D. Cicchetti and D. J. Cohen, Eds. Hoboken, NJ, USA: Wiley, 2006, vol. 1, pp. 293–332.
7. J. T. Cheng et al., "Eye gaze and visual attention as a window into leadership and followership: A review of empirical insights and future directions," The Leadership Quarterly, vol. 33, no. 5, p. 101604, Oct. 2022.
8. B. Singh and H. Wagatsuma, "A removal of eye movement and blink artifacts from EEG data using morphological component analysis," Computational and Mathematical Methods in Medicine, vol. 2017, Article ID 1861645, 2017.
9. D. L. Schomer and F. H. Lopes da Silva, Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, 7th ed. Philadelphia, PA, USA: Wolters Kluwer, 2017.

Authors Introduction

Mr. Kosei Shibata

He received his Master's degree in Engineering in 2023 from the Graduate School of Life Science and System Engineering, Kyushu Institute of technology (Kyutech) in Japan,. He is currently a doctoral course student in Kyutech.

Mr. Yide Yang

He received his Bachelor's degree in Intelligent Science and Technology in 2023 from the School of Artificial Intelligence and Data Science, Hebei University of Technology, China. He is currently a Master's student at Kyushu Institute of Technology, Japan, specializing in Neuroinformatics.

Ms. Rena Kato

She received her associate's degree from the Department of Electrical and Electronic Systems Engineering, National Institute of Technology, Toyota College, Japan in 2024. She is currently enrolled in the Department of Systems Design and Informatics, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, Japan.

Dr. Hendry F. Chame

He holds degrees in Psychology and Computer Science (UCV, Venezuela), M.S. degrees in Instructional Psychology and Robotics and a Ph.D degree in Robotics (ECN, France). His postdoctoral research includes bio-inspired underwater robotics (FURG, Brazil), cognitive neurorobotics (OIST, Japan), and cognitive robotics (LAAS, France). In 2021, he became an Associate Professor at the University of Lorraine, focusing on AI, robotics, and complex systems at LORIA.

Dr. Laurent Bougrain

He received his M.S. degree in artificial intelligence in 1995 from university of Paris 6, a B. degree in psychology and Ph.D. degrees in computer science in 2000 from université of Lorraine. He is associate professor at the university of Lorraine (France) since 2001. He is the leader of the NeuroRhythms team of the department of complex system, artificial intelligence and robotics at LORIA laboratory (France).

Dr. Hiroaki Wagatsuma

He received his M.S., and Ph.D. degrees from Tokyo Denki University, Japan, in 1997 and 2005, respectively. In 2009, he joined Kyushu Institute of Technology, where he is currently an Associate Professor of the Department of Human Intelli- gence Systems. His research interests include non- linear dynamics and robotics. He is a member of IEEE.

Variable Selection Methods for Multivariate Time Series Data Using Multivariate Granger Causality

Keita Ohmori
Kyushu Institute of Technology, 680-4 kawazu, Iizuka, Fukuoka, 820-8502, Japan

SUMCO, 1-52 Kubara, Yamashiro, Imari, Saga, 849-4256, Japan

Toshiki Saitoh
Kyushu Institute of Technology, 680-4 kawazu, Iizuka, Fukuoka, 820-8502, Japan

Akiko Fujimoto
Kyushu Institute of Technology, 680-4 kawazu, Iizuka, Fukuoka, 820-8502, Japan

Eiji Miyano
Kyushu Institute of Technology, 680-4 kawazu, Iizuka, Fukuoka, 820-8502, Japan
Email: keita.ohmori468@mail.kyutech.jp, toshikis@ai.kyutech.ac.jp fujimoto@ai.kyutech.ac.jp, miyano@ai.kyutech.ac.jp

Abstract

We study variable selection methods for multivariate time-series data. Hmamouche et al. proposed a method that first constructs a causal graph based on Granger causality among time-series data, and then selects variables from clusters formed by clustering the vertices corresponding to each variable. However, this method only performs pairwise Granger causality tests, which may not fully capture the interactions among variables. To address this issue, we propose a variable selection method that performs multivariate Granger causality tests on all combinations of explanatory variables with respect to the target variable, selecting the combination with the strongest causality. Our method successfully constructs a predictive model with a higher accuracy compared to the previous method.

Keywords: Granger causality, Variable selection, Multivariate time series data

1. Introduction

In recent years, with the proliferation of the IoT, a vast amount of diverse time-series data has been accumulated through sensors in the manufacturing industry. These data are utilized for optimizing manufacturing processes and improving quality, such as analyzing the causes of equipment failures and quality degradation. Machine learning is primarily employed for these analyses, where variable selection is a crucial element, considering computational cost, recognition accuracy, and model interpretability. Moreover, since time-series data inherently involve past data influencing current values, machine learning models for multivariate time-series data must clarify the extent to which the past data of each variable affects the current and future values of the target variable. One method for evaluating such influence is the Granger causality test [1]. The Granger causality test is a statistical technique that examines how well the past information of one variable can explain the future values of another variable. Previous studies have proposed approaches utilizing the Granger causality test for variable selection. For example, in 2015 Sun et al. [2] proposed a method that performs univariate Granger causality tests between the target variable and explanatory variables, selecting all explanatory variables that have a causal relationship with the target variable. In 2018, Hmamouche et al. [3] proposed a method that constructs a causal graph based on Granger causality among time-series data and selects variables by clustering the vertices corresponding to each variable. However, these methods have limitations, as it only performs pairwise Granger causality tests, which makes it challenging to adequately account for interactions among multiple variables.

To address this issue, we propose a new variable selection method that performs multivariate Granger causality tests on all possible combinations of explanatory variables in the dataset and selects the combination with the strongest causality with respect to the target variable. This new method successfully constructs a predictive model with higher accuracy than traditional methods by considering the interactions among variables. Consequently, it overcomes the limitations of the previous method and expands the possibilities for variable selection in time-series data analysis.

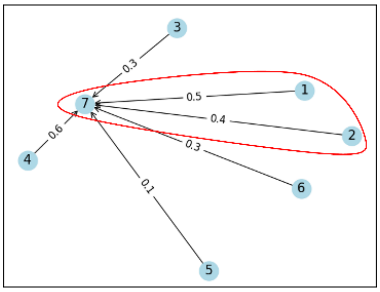


Fig. 1. Causal graph

2. Preliminary

2.1. *Granger Causality*

Granger causality is a method for identifying causal relationships between time series data, evaluating whether a causal variable contributes to predicting the outcome. Specifically, a variable x is said to causally influence another variable y if a regression model using both the past values of x and y significantly outperforms a model using only the past values of y. This relationship is determined by comparing two regression models using statistical methods such as the F-test. In this study, we use Vector Auto Regression (VAR) model [4] for regression, and the two regression models are expressed by the following equations:

Model 1:

$$y_t = \alpha_0 + \sum_{i=1}^{l} A_i y_{t-i} + \varepsilon_t, \quad (1)$$

Model 2:

$$y_t = \alpha_0 + \sum_{i=1}^{l} A_i y_{t-i} + \sum_{i=1}^{l} B_i x_{t-i} + \varepsilon_t. \quad (2)$$

Where α_0 represents the constant term, for each $i \in \{1, \ldots, l\}$, A_i and B_i represent the regression coefficients, l represents the maximum number of lags for x and y, and ε represents the error term (white noise). The accuracy of the two regression models was tested using the F-test. The test statistic is expressed as follows:

$$F = \frac{\frac{RSS_1 - RSS_2}{l}}{\frac{RSS_2}{n - 2l - 1}}. \quad (3)$$

Where RSS_1 and RSS_2 the residual sum of squares for Model 1 and Model 2, respectively, l is the number of lags for x, and n is the sample size. The value F follows an F-distribution with degrees of freedom $(n, n - 2l - 1)$. If the calculated F exceeds significance level, we reject the null hypothesis that x does not cause y and conclude that x causes y. The strength of causality is defined as 1-p, where p is the p -value from the Granger causality test. This value ranges from 0 to 1, with value closer to 1 indicating stronger causality.

2.2. *Hmamouche's method*

Hmamouche's method [2] first constructs a causal graph based on Granger causality among time-series data and then selects variables from clusters formed by clustering the vertices corresponding to each variable. This method is referred to as the GSM (Granger selection method) algorithm. The outline of the GSM algorithm is as follows:

GSM algorithm
Input: A set of explanatory variables $X = \{x^1, x^2, \ldots x^n\}$, Min-Causality threshold T, the selection size k and a target variable y
Output: the selected variables associated to y
Step 1: Perform Granger causality tests distinct pairs of variables x^i and x^j and construct a causality matrix. Each element a_{ij} of the matrix A is the strength of the causality of x^i and x^j. If the strength of causality is smaller than the threshold T, it is considered as no causality.
Step 2: Partition the input variables X from the causality matrix constructed in Step 1 using the PAM (Partitioning Around Medoids) method. The goal is to group variables by minimizing causalities between clusters and maximizing causalities within clusters.
Step 3: Choose one element from each cluster that has the strongest causality on the target variable.

3. Our Method

3.1. *Multivariate Granger causality*

In our method, we use a multivariate Granger causality test, which is an extension of the Granger causality test. The multivariate Granger causality test evaluates not only the causality between two variables but also the causality of a combination of multiple variables on another variable. Fig.1 illustrates a causal graph representing the causal relationships among time-series data. The explanatory variables are labeled 1 through 6, and the target variable is labeled as 7. The values annotated on each edge indicate the strength of causality. In Fig. 1, none of the individual explanatory variables exhibit causality with the target variable. However, when variables 1 and 2 are combined, a causal relationship with the target variable is observed. This demonstrates that, while individual explanatory variables may not exhibit causality with the target variable, combinations of multiple variables exhibit causal relationships. The multivariate Granger causality test is capable of capturing such causal relationships. The two models compared in the multivariable Granger causality test can then be the as follows:

$$\text{Model 1: } y_t = \alpha_0 + \sum_{i=1}^{l} A_i y_{t-i} + \varepsilon_t, \quad (4)$$

$$\text{Model 2: } y_t = \alpha_0 + \sum_{i=1}^{l} A_i y_{t-i} + \sum_{j=1}^{k} \sum_{i=1}^{l} B_i^j x_{t-i}^j + \varepsilon_t. \quad (5)$$

Where a set of explanatory variables $X = \{x^1, x^2, \dots x^k\}$, a target variable y, a constant term α_0, for each $i \in \{1, \dots, l\}$ and $j \in \{1, \dots, k\}$, A_i and B_i^j represents the regression coefficients for y and x, l represents the maximum number of lags for y and x, and ε represents the error term (white noise). The accuracy of the two regression models was tested using the F-test. The test statistic used is the same as in (3).

3.2. *Procedure*

We perform multivariate Granger causality tests on the target variable using all possible combinations of explanatory variables and select the combination that exhibits the strongest causality with respect to the target variable. The algorithm for our method is presented below:

Our Algorithm

Input: A set of explanatory variables $X = \{x^1, x^2, \dots x^n\}$, Min-Causality threshold T, the selection size k and a target variable y

Output: The combination of explanatory variables that exhibits the strongest causal relationship with the target variable.

Step 1: Generate all combinations of k variables from the explanatory variables.

Step 2: For each combination of explanatory variables, perform a multivariable Granger causality test to assess their influence on the target variable.

Step 3: Compute the causality strength s as 1-p, where p is the p-value for multivariate Granger causality test, and consider it significant if $s \geq T$. Retain the combination of explanatory variables X' and the corresponding causality strength as the result.

Step 4: Identify the combination with the highest causality strength and return it as the best explanatory variable set for predicting y.

4. Experiments

4.1. *Evaluation Measure*

We evaluate our method by comparing the variables selected using our method with that of Hmamouche's method. For this, prediction models are constructed using a Vector Error Correction Model (VECM) [5] with the selected variables, and the predictive accuracy of these models are used as the evaluation criterion. For accuracy evaluation, we used the Normalized Root Mean Square Error (NRMSE). VECM model extends the VAR model by considering the non-stationarity of time series and incorporating cointegration equations. If two time series (x_t, y_t) follow a first-order integration process ($I(1)$), VECM can be expressed as:

$$\Delta y_t = \alpha_{0y} - \gamma_y(\beta_0 y_{t-1} - \beta_1 x_{t-1}) + \sum_{i=1}^{p} v_{iy}\, \Delta y_{t-j} + \sum_{i=1}^{p} w_{iy}\, \Delta x_{t-j} + \varepsilon_t, \quad (6)$$

$$\Delta x_t = \alpha_{0x} - \gamma_x(\beta_0 y_{t-1} - \beta_1 x_{t-1}) + \sum_{i=1}^{p} v_{ix}\, \Delta y_{t-j} + \sum_{i=1}^{p} w_{ix}\, \Delta x_{t-j} + \varepsilon_t. \quad (7)$$

Where $(\Delta y_t, \Delta x_t)$ are $(y_t - y_{t-1}, x_t - x_{t-1})$, $(\alpha_{0y}, \alpha_{0x})$ represent the constant term, (v_{iy}, v_{ix}) represents the regression coefficients, the coefficients (β_0, β_1) are the cointegrating parameters, and (γ_y, γ_x) are the error correction parameters. NRMSE is calculated using the following formula:

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)}}{\text{stdev}(y)}. \quad (8)$$

Where $(\hat{y}_1, \hat{y}_2, \dots \hat{y}_n)$ are the forecasts, $(y_1, y_2, \dots y_n)$ are the real values, and n represents number of data points used for accuracy comparison.

In our method, the p-value threshold for the Granger causality test was set at 10%. The lag parameters for the VECM model and Granger causality tests are determined based on Akaike's Information Criterion (AIC) [6] The training data for the VECM model consisted of the first 90% of the dataset in chronological order, while the last 10% was used as test data.

4.2. *Data Sets*

A dataset used in this study is obtained from the Machine Learning Repository website [7]. It includes the returns of the Istanbul Stock Exchange along with seven other international indices: SP, DAX, FTSE, NIKKEI, BOVESPA, EU, and EM, covering the period from June 5, 2009, to February 22, 2011.

4.3. *Results*

Table 1 shows results of variable selection using GSM algorithm. In the GSM algorithm, specifying the number of clusters in the causal graph allows adjustment of the 5 to 7, GSM algorithm outputs BOVESPA, BOVESPA and EU, and BOVESPA, EU, and SP, respectively. The clusters necessarily contain variables with a causality score of 0.9 or higher with respect to the target variable.

Table 1:Variable selection results using the GSM

#Clusters	Selected variables
1	BOVESPA
2~4	BOVESPA, EU
5~7	BOVESPA, EU, SP

Table 2:Variable selection results using our method

# Variables	Selected variables
1	BOVESPA
2	SP, BOVESPA
3	SP, DAX, EU

Table 3:Predictive accuracy for each variable set

Methods	Selected variables	NRMSE
Previous [2], Ours	BOVESPA	1.118
Previous [2]	BOVESPA, EU	1.010
Previous [2]	BOVESPA, EU, SP	1.022
Ours	SP, BOVESPA	1.002
Ours	SP, DAX, EU	0.999

Table 2 shows the combinations of selected variables for each variable count using our method. There is no combinations with four or more variables exceeded a causality score of 0.9 for the target variable. Table 3 summarizes the predictive accuracy (NRMSE) of the models constructed using the selected variables by the methods. These results demonstrate that our method achieves higher predictive accuracy compared to previous studies. Notably, for combinations of three variables, the predictive accuracy improved by 2.3%, i.e., previously 1.022 for {BOVESPA, EU, SP} while newly 0.999 for {SP, DAX, EU}.

4.4. *Discussions*

These results indicates that an advantage of our method is that it enables the selection of variables that improve the accuracy of the predictive model by taking into account the interaction effects among multiple variables.

In contrast, previous methods involve clustering the causal graph to group variables with similar effects on the target variable and then selecting variables with strong causality to the target variable from those groups. However, in this method, the causal graph is constructed based on pairwise causality tests, which are likely insufficient to fully account for the effects of interactions among multiple variables.

5. Conclusion

We proposed a method for variable selection in multivariate time-series forecasting that accounts for variable interactions. Unlike previous methods relying on pairwise Granger causality tests and clustering, our method evaluates all possible combinations of explanatory variables. This enables the identification of variable sets that improve predictive accuracy by considering the combined effects of multiple variables. Experiments demonstrated that our method outperforms traditional approaches, with a 2.3% improvement in NRMSE for three-variable combinations. These results underscore the importance of considering variable interactions in predictive modeling. However, this method requires high computational costs due to testing all possible combinations.

Future efforts will aim to reduce computational cost and extend the method to handle nonlinear interactions. Incorporating nonlinear Granger causality techniques, such as those by Chen et al. [8], could further enhance its effectiveness. In conclusion, our method advances variable selection for time-series forecasting, offering more accurate predictions and deeper insights into variable dependencies. Future developments in efficiency and nonlinearity handling will broaden its applicability across domains.

References

1. C. W. Granger, “Testing for causality: a personalviewpoint”, Journal of Economic Dynamics and control, vol. 2, 1980, pp. 329–352.
2. Y. Sun, J. Li, J. Liu, C. Chow, B. Sun, and R. Wang, “Using causal discovery for feature selection in multivariate numerical time series”, Machine Learning, vol. 101, no. 1-3, 2015, pp. 377–395.
3. Y. Hmamouche, A. Casali, and L. Lakhal “A Causality Based Feature Selection Approach for Multivariate Time Series Forecasting”, HAL Id: hal-01467523, 2018.
4. M. Quenouille, “The analysis of multiple time-series, ser”, Griffin’s statistical monographs & courses. Griffin, 1957.
5. S. Johansen, “Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models”, Econometrica: Journal of the Econometric Society, 1991, pp. 1551–1580.
6. H. Akaike, “A new look at the statistical model identification”, IEEE transactions on automatic control, vol. 19, no. 6, 1974, pp. 716–723.
7. M. Lichman, “UCI machine learning repository”, 2013, [accessed: 2024-11-01].
8. Y. Chen, G. Rangarajan, J. Feng and M. Ding, “Analyzing multiple nonlinear time series with extended Granger causality”, Phys. Lett. A 324(1), 2004, pp. 26–35

Authors Introduction

Mr. Keita Ohmori

He received Bachelor's degree in Science in 2019 and M.S. degree in Science in 2021 from Tohoku University Since 2021, he has been working at SUMCO Corporation. In 2024, he enrolled as a Ph.D. student at Kyushu Institute of Technology, Japan.

Dr. Toshiki Saitoh

He received the B.S.E. degree from Shimane University in 2005, and the M.S. and Ph.D. degrees (Information Science) from Japan Advanced Institute of Science and Technology in 2007 and 2010, respectively. He is a professor at Kyushu Institute of Technology.

Dr. Akiko Fujimoto

She received the B.S., M.S., and Ph.D. in Science degrees from Kyushu University in 2005, 2007, and 2010, respectively. She is an Associate Professor of the Department of Artificial Intelligence, Kyushu Institute of Technology.

Dr. Eiji Miyano

He received the B.Eng., M.Eng., and Dr.Eng., degrees in computer science from Kyushu University in 1991, 1993, and 1995, respectively. He is currently a professor of the Department of Artificial Intelligence, Kyushu Institute of Technology.

A Support System for a Visually Impaired Person Finding Bus Route Numbers Employing MY VISION

Daichi Nanaura
Graduate School of Engineering, Kyushu Institute of Technology, 1-1 Sensuicho, Tobata-ku, Kitakyushu, 804-8550, Japan

Seiji Ishikawa[1*]**, Yui Tanjo**[1]
[1]*Faculty of Engineering, Kyushu Institute of Technology, 1-1 Sensuicho, Tobata-ku, Kitakyushu, 804-8550, Japan*
**Emeritus Professor of Kyushu Institute of Technology*
Email: nanaura.daichi608 @mail.kyutech.jp, tanjo@cntl.kyutech.ac.jp

Abstract

A bus is not a very convenient means for a visually impaired person because of the difficulty in identifying its route number, although it is an economical tool for travel. This paper proposes a method of detecting and identifying a bus route number using the MYVISION system which employs an ego camera worn by a visually impaired person. The method finds a frontal area of a bus approaching a bus stop using the video which the camera provides by using optical flow and random forest employing Haar-like features. It then extracts the upper destination panel area followed by the detection of a route number at the right-hand side of the panel. Finally, the detected route number is identified by template matching. In the experiment, various kinds of videos containing the buses of a bus company were captured at different places and different weather conditions, and the effectiveness of the proposed method was shown. The method is now planned to be applied to the busses of various bus companies.

Keywords: MY VISION, visually impaired person, bus route numbers recognition, Haar-like feature, random forest

1. Introduction

According to 2019 WHO report [1], it is estimated that there are at least 2.2 billion people affected by the conditions that cause severe visual impairment worldwide. People with visual impairment can reach daily familiar places without difficulty, but they have difficulty in unfamiliar places. In particular, travelling by bus is reported to be difficult, because identifying the bus route number is not easy for them [2].

Various studies and developments have been conducted worldwide to address this issue. For example, there is a research on having drivers carry special devices to communicate wirelessly with those people who have difficulty in seeing [3], and a research on using GPS and RFID to obtain bus information [4]. However, they require devices to be installed at all buses and bus stops, which is very costly. Another study [5] uses a mobile application and a smartphone camera to identify buses. Although the cost is lower, since the device only needs to be implemented by the user, he/she cannot operate the smartphone unless the user has a certain level of vision

This paper proposes a method that detects a bus from the images of a single camera MY VISION: (a Magic eYe of a Visually Impaired for Safety and Independent actiON) worn by a person with visual difficulties by video image processing. The proposed procedure is as follows: A small camera is attached to the right arm of the person with visual difficulties and he/she stands at the bus stop. When moving objects such as a bus or a car approaches the bus stop, the method judges which vehicle it is. If the method recognizes the object as a bus, the route number is extracted and identified from the bus frontal image and it is informed to the user The proposed system aims to reduce costs and to improve versatility, which was an issue in the related research [4]. The proposed method uses only a single camera and a computer. In addition, the system can be used regardless of the weakness of the eyesight by automatically performing every procedure from start to recognition without requiring any operation by the user, thereby improving convenience in usage.

The proposed method is improved from the related research [6], which identifies whether nor not a moving vehicle is a bus from the images from a head-mounted camera. The present research detects the movement of a bus or a car when it approaches a bus stop and identifies which one it is. The technique proposed in [6] is applied only to the images taken during the daytime by updating Camshift histogram of each frame to track the movement. This makes the system increase the amount of calculation during execution. On the other hand, a bus number recognition technique proposed in [7] conducts to detect the display on the top of the bus and recognize the bus route number by labelling. However, this method can only be employed in the evening when the contrast of the display is large by the LEDs used in the display.

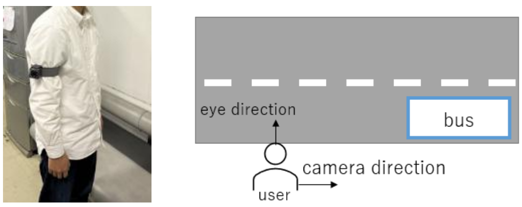
In order to solve the recognition problem of both a bus and its route number, the proposed method combines the problem into a single process and performs the update of the Camshift only once rather than performing it in all frames. Moreover, the proposed method improves the possibility of recognition process even in the daytime when the sunlight is strong using gamma correction to adjust the contrast.

2. Proposed Method

2.1. *Detection of moving objects*

The proposed method replaces the eyes of visually impaired persons with MY VISION. When waiting at a bus stop, a visually impaired person's body faces the road, and the bus approaches from his/her right/left-hand side. Therefore, a MY VISION camera is worn on his/her right upper arm to capture video. A photo of the actual wearing and positioning of a camera is shown in Fig. 1.

The Harris corner detector [8] is applied to the input images to extract feature points. The extracted feature points are tracked using the pyramidal L-K method [9] to acquire optical flows. The acquired optical flows are then clustered using the *k*-means method [10] to extract the candidate regions of moving objects.



(a) (b)

Fig.1. Photographic environment: (a) A camera mounted on the right arm, (b) location of a user and the bus.

2.2. *Recognition of a bus*

Moving objects in the extracted candidate area described in section 2.1 are then recognized as a bus or other objects. Random Forest [11] is used for the recognition, which is trained using a variety of decision trees. Decision trees with low correlation are created by randomly selecting features at the non-terminal nodes of each decision tree. Learning and testing are carried out using the various decision trees created in this way. Haar-like features [12] are calculated from each training image to make teaching data of each class and a set of feature samples. From this set of samples, a random subset is created. Haar-like feature patterns shown in Fig. 2 are calculated by the following equation;

$$H(A,B) = f(A) - f(B) \quad (1)$$

where $f(*)$ is the average luminance in each region *.The learning parameters of the Random Forest are shown in Table 1. The image is scanned from the top left to bottom right to obtain these six kind of Haar-like features. Since we need to identify the bus route number at a close distance, unlike the conventional technique [6,7], we do not trace a bus from a distant location. Instead, we use Random Forest to identify the bus at a distance. If an image is identified as a bus for more than five successive frames, it is judged that the bus is in front of or approaching the user, and then the identified area is automatically tracked. We use a Continuously Adaptive Mean Shift (CAMSHIFT) [13] algorithm for tracking, the size of the rectangle area is varied according to the size of the target (a bus) to be tracked. Initially, a Hue histogram within the rectangular area (a bounding box) is calculated, and the probability distribution of a hue is used to search for the similarity of the area in the successive frames, and the search area is shifted to the area where the similarity is larger.

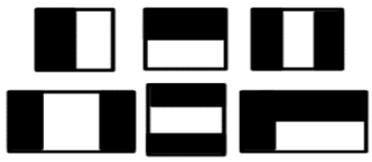


Fig.2. Six Haar-like features used in the method.

2.3. *Application to daytime video*

Table 1 Parameters of Random Forest

data set	Bus ; 50 Background ; 50 Other vehicle ; 50
feature dimensions	5157
Data set size[pixel]	80*48
subset	40
tree	5
Feature selection count	100
Threshold selection count	100

When photographing an LED display of a bus showing the destination and the route number, the camera shutter must be released at a frequency lower than the LED light emission frequency, so that the images do not become dark or do not being reflected. However, if the shutter speed is lowered, the exposure time increases when photographing during daytime, which makes the image overexposed. Overexposure causes the decrease of the recognition accuracy and tracing ability by Camshift. To solve this problem, rather than adjusting the shutter speed, gamma correction is applied to the obtained image. This makes the Haar-like features, *i.e.*, the difference of the brightness values remarkably enhance, and also prevents the color probability distribution from abnormal expansion when applying Camshift to the image for tracing a moving object.

Since the bus is recognized in the distance and moves from the right to the left on the captured video, there is no need to refer to the area on the right of the coordinates that were recognized immediately before.
In addition, since the bus approaches the bus stop at a lower speed and decelerates, the area to be referred can be expanded only on the left-side. The expansion speed of the area to calculate the color probability distribution is determined from the value calculated by assuming the speed range of the bus. This makes the tracing area by Camshift enable to prevent from expanding excessively.

2.4. *Extracting the image of the route number*

In the extraction step of the route number, instead of searching the whole image, we consider and cut out the upper 30% of a rectangular area from the image, and apply HSV transform to the rectangle image. As the route number characteristic is an orange color, we apply the Hue histogram to the rectangle area to obtain a new rectangular area with the orange color distribution. The rectangular area of the bus route number is binarized using the discriminant analysis. Finally, the area of the number is extracted by vertical and horizontal projection of the binary pixels. Template matching is performed to the extracted number to identify it. The process of the route number extraction is illustrated in Fig.3.

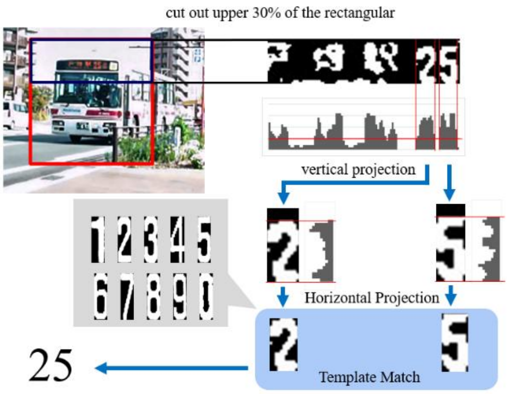


Fig.3. The process of extracting the route number.

3. Experimental Results

3.1. *Bus Recognition*

In the experiment, we used six videos to verify the performance of the proposed method. The evaluation criterion is shown in Fig.4. GT is a ground truth and it stands for the correct area. OA means the detected area. Therefore, the cover value of the GT area with the bus detection area is used for evaluating the bus recognition ability. Here, we set the cover value threshold *T* to 0.8.

The recognition accuracy (*True*) and false recognition (*False*) are calculated as follows;

$$True = \frac{F_t}{F_{GT}} \times 100\ [\%] \tag{2}$$

Fig.4. Evaluation criterion of the cover value.

$$False = \frac{F_f}{F_{GT}} \times 100\ [\%] \tag{3}$$

where, F_{GT} means the total number of Ground Truth frames, F_t means the total number of the frames with cover more than *T*, and F_f means the total number of the frames with cover less than *T*.
As a result, the average recognition accuracy (*True*) was 95.5%. The conventional method [6] sets *T*=0.5, while the proposed method sets *T*=0.8, which is stricter than the previous method. If an approaching bus cannot be correctly enclosed by a bounding box, the route number extraction becomes difficult.

3.2. *Route numbers detection and recognition*

An experiment was conducted to evaluate the accuracy of the route number recognition with six videos. 'True' was considered correct if the result was correct recognition, while 'undetected' or 'false positive' was judged as False. The evaluation and conditions are shown in Table 2. The average percentage of True responses was 67%. The results of two videos out of six were False. This is because the characters' route number of the video 4 were unclear when binarization was performed to the image. In another video, the route number was not detected. This is due to the proposed route number recognition algorithm is only effective for a two-digit number, and the route number of the experimental video 6 was one-digit number. In this case, we evaluated the result as False. Part of the experimental results are shown in Fig.5.

4. Discussion and Conclusion

A system was proposed to detect a bus and its route number to assist a visually impaired person at a bus stop using MY VISION. Experiments showed the effectiveness of the proposed method, *i.e.*, detection and recognition of busses from moving objects and recognition of the route numbers were successfully done. The accuracy of bus recognition was 95.5% in average, which was improved by about 3% compared to the conventional method [6]. It was also successful in recognizing buses over a wider time range than the conventional method [7]. A more robust algorithm needs to be developed to improve the recognition accuracy of the bus and the detection of route numbers including one-digit numbers.

(a)

(b)

Fig.5. Part of the experimental results: (a) Video 1, (b) video 2.

Table 2 Result of the recognition of route numbers and corresponding time zones.

Videos	Recognition	Time zone
1	True	evening
2	True	daytime
3	True	daytime
4	False	daytime
5	True	daytime
6	False	evening
Accuracy	67[%]	

References

1. WHO: World report on vision, https://www.who.int/publications/i/item/9789241516570, 2019, (Accessed 2024-10-23).
2. Kitakyushu City Disability Welfare Information Center: Hikosen Mirai, https://www.ksjc.jp/index.php?id=2293, (Accessed 2024-8-9)
3. P. Herdiansyah, G. I. Hapsari, G. A. Mutiara, L. Meisaroh: "Smart bus arrival notification for visually impaired", Proceedings of International Conference on Electrical Engineering and Informatics (ICEEI), pp.614-618, 2019.
4. M. Z. H. Noor, I. Ismail and M. F. Saaid: "Bus detection device for the blind using RFID application", International Colloquium on Signal Processing & Its Applications (CSPA), pp. 247-249, 2009.
5. P. Wongta, T. H. Chalidabhongse: "Vision-based bus route number reader for visually impaired travelers", International Conference on Imaging, Signal Processing and Communication (ICISPC), pp. 64-69, 2018.
6. K. Ishitobi, J. K. Tan, H. Kim, S. Ishikawa: "Detection of a specific moving object from head-mounted camera images", 2017 IEEE/SICE International Symposium on System Integration (SII), pp. 817-822, 2017.
7. J. K. Tan, Y. Hamasaki, Y. Zhou, K. Ishitobi: "A method of identifying a public bus route number employing MY VISION", Journal of Robotics, Networking and Artificial Life, Volume 8, Issue 3, pp. 224 -228, 2021.
8. C. Harris, M. Stephens: "A combined corner and edge detector", Proc. of the 4th Alvey Vision Conference, pp.147-151, 1988.
9. J. Y. Bouguet: "Pyramidal implementation of the Lucas Kanade feature tracker description of the algorithm", Intel Corporation, Microprocessor Research Labs, OpenCV Documents, 9 pages, 2000.
10. J. MacQueen: "Some methods for classification and analysis of multivariate observation", Proc. of the 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, pp.281-297, 1967.
11. F. Breiman: "Random forests", Machine Learning, Vol. 45, No. 1, pp.5-32, 2001.
12. P. Viola, M. Jones: "Rapid object detection using a boosted cascade of simple features", Proc. of the Conference on Computer Vision and Pattern Recognition, Vol. 1, pp.511- 518, 2001.
13. G.R. Bradski: "Computer vision face tracking for use in a perceptual user interface," Intel Technology Journal Q' 98, 15pages, 1998.

Authors Introduction

Mr. Daichi Nanaura

He received his Bachelor's degree in Engineering in 2023 from the Faculty of Engineering, Kyushu Institute of technology in Japan. He is currently a master student in Kyushu Institute of Technology, Japan

E. Prof. Dr. Seiji Ishikawa

Seiji Ishikawa graduated from Tokyo University and was awarded BE, ME and PhD there. He is now Emeritus Professor of Kyushu Institute of Technology. He was Visiting Researcher of the University of Sheffield, UK, and Visiting Professor of Utrecht University, NL. His research interests include visual sensing & 3-D shape/motion recovery. He was awarded The Best Paper Awards in 2008, 2010, 2013 and 2015 from BMFSA.

Prof. Dr. Yui Tanjo

Dr. Tanjo is currently Professor with the Department of Mechanical and Control Engineering, Kyushu Institute of Technology. Her current research interests include ego-motion analysis by MY VISION, three-dimensional shape/motion recovery, human detection, and its motion analysis from video. She was awarded SICE Kyushu Branch Young Author's Award in 1999, the AROB Young Author's Award in 2004, the Young Author's Award from IPSJ of Kyushu Branch in 2004, and the BMFSA Best Paper Awards in 2008, 2010, 2013 and 2015. She is a member of IEEE, The Information Processing Society, The Institute of Electronics, Information and Communication Engineers of Japan.

Human Pose Estimation from Egocentric Videos

Shunya Egashira
Graduate School of Engineering, Kyushu Institute of Technology, 1-1 Sensuicho, Tobata, Kitakyushu, 804-8550, Japan

Yui Tanjo[1]
[1]*Faculty of Engineering, Kyushu Institute of Technology, 1-1 Sensuicho, Tobata, Kitakyushu, 804-8550, Japan*
Email: Egashira.shunnya602@mail.kyutech.jp, tanjo@cntl.kyutech.ac.jp

Abstract

According to a survey conducted by the Ministry of Health, Labour and Welfare in 2019, about 30% of patients suffer from back pain and stiff shoulders. Although researches on pose estimation have been conducted for a long time, they cannot be used for daily pose estimation, because they need fixed cameras to capture target/subject motion. To solve this problem, the present paper, proposes a novel pose estimation method from egocentric videos using Epipolar Geometry. It computes three rotational angles, i.e., pitch, yaw and roll, from the egocentric motion videos to evaluate differences from his/her normal motion. In the experiment, three egocentric videos were used to verify the performance and effectiveness of the proposed method and reasonable/satisfactory results were obtained.

Keywords: Egocentric, Posture Estimation, Epipolar Geometry, Rotation Angles

1 Introduction

According to the Ministry of Health, Labor and Welfare's 2009 National Survey of Basic Living Conditions [1], the prevalence of back pain and shoulder stiffness is about 30%. [2] [3] Since these problems interfere with daily life of a person, it is important to consciously correct his/her posture in daily life. Therefore, posture estimation, which allows one to check one's posture status, has been used in the research on posture estimation in animation motion production and virtual reality games. One of the posture estimation techniques is motion capture. Motion capture is a method of creating a 3D model by placing a monocular camera[4] or multiple cameras[5][6][7] around a person to be photographed and capturing the subject's movements. Methods for detecting the position of the subject's joints include attaching markers to the subject [8] and deep learning estimation [9].

However, these methods have disadvantages in that they require time for environmental preparation such as equipment and can only be performed in a limited area. There have also been studies on posture estimation using deep learning[10], gait posture estimation using video from a third viewpoint using a fixed camera[11], and gait posture estimation using epipolar geometry[12] as previous researches. However, the problems are that it is difficult to determine the cause of the problem because of the use of deep learning, and it is also impossible to estimate the posture of a human daily life because it can only take pictures in a fixed environment. Moreover, it is difficult to estimate a posture with small movement, because it is normally used to estimate walking posture.

Fig.1. The chest-mounted camera

Based on these backgrounds, this paper proposes a method to acquire images of the surrounding landscape from a chest-mounted camera(Fig.1) and estimate the angle using epipolar geometry using the feature points obtained from the images.

2 Methodology

2.1. *Camera Calibration*

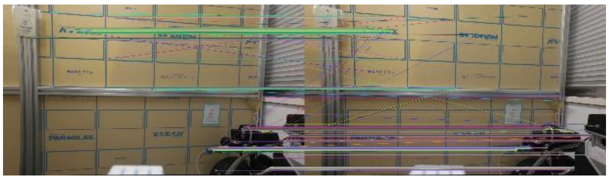
The camera contains lens distortion, which affects the accuracy of the feature point correspondence and makes it impossible to correctly map between the two images. In order to eliminate the distortion, the camera is calibrated. The camera parameters include internal parameters, external parameters, and distortion coefficients. A checkerboard is used to estimate the camera parameters.

2.2. *Feature Point Detection and Matching*

A-KAZE is a robust feature point extraction algorithm invariant to scaling, rotation, and lighting change.
Using A-KAZE, the feature points are extracted and described, and they are matched between two image frames.

While the Lucas-Kanade (LK) tracker is commonly used in target tracking, its use in feature matching is comparatively rare. After detecting feature points in one image, the corresponding matching region in the other image must be identified to establish feature point pairs. (Fig.2) The LK tracker is subsequently employed to locate this matching region.

The removal of the outliers is performed by applying RANSAC to hypothetically matched point pairs.



(a) (b)

Fig.2. Feature point matching: (a) frame1 (b) frame3

2.3. *Modeling with homographic matrices*

A homography matrix is a matrix that represents a projective matrix between two images, as well as a constraint between the corresponding points.
If the x and y coordinates of the point in the image of the previous frame and those of the corresponding point in the image of the next frame are denoted by $\boldsymbol{x_1}$, $\boldsymbol{y_1}$, $\boldsymbol{x_2}$, $\boldsymbol{y_2}$, respectively, the relationship is calculated using a homographic matrix.

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} \quad (1)$$

where $h_{11} \cdots h_{33}$ are the elements of the homography matrix.

When estimating the homography matrix using the corresponding points obtained from the feature point matching described above, RANSAC is used to remove the outliers.

2.4. *Removing Outliers with RANSAC*

The RANSAC algorithm is shown below.

1. Randomly extract a certain number of data from the entire data and perform model estimation.
2. Count and memorize outlier data.
3. Repeat steps 1-2 multiple times.
4. Determine the model with the least outliers as the best model.

In the final projection transformation model, the feature points pairs that correspond to the outlier are removed. If the two images are correlated, the position and the pose of the camera is estimated using epipolar geometry.

2.5. *Estimation of the fundamental matrix*

All correspondences between the two images satisfy Eq. (2).

$$\boldsymbol{\widetilde{m}_{t+1,i}}^{\boldsymbol{T}} \boldsymbol{F} \boldsymbol{\widetilde{m}_{t,i}} = 0 \quad (2)$$

$\boldsymbol{\widetilde{m}_{t,i}}$ and $\boldsymbol{\widetilde{m}_{t+1,i}}$ are the uniform coordinates of the image coordinates of the ith camera at time t and t+1, respectively, and $\boldsymbol{F}$ is the fundamental matrix. They are expressed as shown in Eq. (3).

$$\boldsymbol{F} = \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix}, \ \boldsymbol{\widetilde{m}_{t,i}} = \begin{pmatrix} u_{t,i} \\ v_{t,i} \\ 1 \end{pmatrix}, \ \boldsymbol{\widetilde{m}_{t+1,i}} = \begin{pmatrix} u_{t+1,i} \\ v_{t+1,i} \\ 1 \end{pmatrix} \quad (3)$$

In this study, we use the 8-point algorithm [13] to find the fundamental matrix $\boldsymbol{F}$. Equations (2) and (3) are expanded to $f_{11} \cdots f_{33}$, which is summarized as Eq. (4)

$$\begin{pmatrix} u_{t,i}u_{t+1,i} & v_{t,i}u_{t+1,i} & u_{t+1,i} & u_{t,i}v_{t+1,i} & v_{t,i}v_{t+1,i} & v_{t+1,i} & u_{t,i} & v_{t,i} & 1 \end{pmatrix} \begin{pmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{pmatrix} = 0 \quad (4)$$

Equation (4) has 8 unknown parameters due to the indeterminate scale, and if there are 8 or more sets of corresponding points, the basic matrix $\boldsymbol{F}$ can be estimated. Therefore, Eq. (4) is applied to the corresponding points of N set (N≥8) and is expressed as Eq. (5).

$$\boldsymbol{Mf} = 0 \quad (5)$$

$$\boldsymbol{M} = \begin{pmatrix} u_{t,1}u_{t+1,1} & v_{t,1}u_{t+1,1} & u_{t+1,1} & u_{t,1}v_{t+1,1} & v_{t,1}v_{t+1,1} & v_{t+1,1} & u_{t,1} & v_{t,1} & 1 \\ u_{t,2}u_{t+1,2} & v_{t,2}u_{t+1,2} & u_{t+1,2} & u_{t,2}v_{t+1,2} & v_{t,2}v_{t+1,2} & v_{t+1,2} & u_{t,2} & v_{t,2} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{t,N}u_{t+1,N} & v_{t,N}u_{t+1,N} & u_{t+1,N} & u_{t,N}v_{t+1,N} & v_{t,N}v_{t+1,N} & v_{t+1,N} & u_{t,N} & v_{t,N} & 1 \end{pmatrix} \quad (6)$$

$$\boldsymbol{f} = (f_{11} \quad f_{12} \quad f_{13} \quad f_{21} \quad f_{22} \quad f_{23} \quad f_{31} \quad f_{32} \quad f_{33})^T \quad (7)$$

Since the right-hand side of Eq. (5) is 0, it is considered to have invariance with respect to scale. To do this, the Euclidean norm of $\boldsymbol{f}$ is normalized to 1 and the fundamental matrix $\boldsymbol{F}$ is estimated. However, since the condition that the rank of the fundamental matrix is 2 is not satisfied, the singular value decomposition of the fundamental matrix is performed, so that its rank becomes 2.

First, the obtained fundamental matrix is decomposed using the following equation.

$$\boldsymbol{F} = \boldsymbol{U\Sigma V^T} \tag{8}$$

Here $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices, and $\boldsymbol{\Sigma}$ is an accusative matrix such as Eq. (9), in which the singular values satisfy $\sigma_1 > \sigma_2 > \sigma_3$.

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \tag{9}$$

Of the three diagonal terms, the smallest singular value is 0, and $\boldsymbol{F}$ obtained by recalculating equation (8) is the final fundamental matrix.

2.6. *Estimating the Essential Matrix*

The essential matrix is a matrix containing information about the motion of the camera between two images, and it is expressed by Eq. (10) using the fundamental matrix $\boldsymbol{F}$ and the internal parameter $\boldsymbol{K}$ obtained in Section 2.1.

$$\boldsymbol{E} = \boldsymbol{K^T F K} \tag{10}$$

The essential matrix $\boldsymbol{E}$ is expressed by the following equation using the rotation matrix $\boldsymbol{R}$ of the camera and the translation vector $\boldsymbol{t}$.

$$\boldsymbol{E} = [\boldsymbol{t}]_X \boldsymbol{R} \tag{11}$$

where $[\boldsymbol{t}]_X$is the skewed symmetric matrix of $\boldsymbol{t}$ defined by

$$[\boldsymbol{t}]_X = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} \tag{12}$$

2.7. *Camera position and pose estimation*

The essential matrix is decomposed into a rotation matrix $\boldsymbol{R}$ and a translation vector $\boldsymbol{t}$ as shown above.

The rotation matrix $\boldsymbol{R}$ and the translation vector $\boldsymbol{t}$ are expressed as follows.

$$\boldsymbol{R} = \begin{cases} \boldsymbol{UWV^T} \\ \boldsymbol{UW^T V^T} \end{cases} \tag{13}$$

$$[\boldsymbol{t}]_X = \begin{cases} \boldsymbol{U\Sigma W U^T} \\ \boldsymbol{U\Sigma W^T U^T} \end{cases} \tag{14}$$

where

$$\boldsymbol{W} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{15}$$

From Eq. (13) and Eq. (14), each of $\boldsymbol{R}$ and $[\boldsymbol{t}]_X$ has two sets of solutions. It is therefore necessary to determine the correct combination of $\boldsymbol{R}$ and $[\boldsymbol{t}]_X$ from among them. For this purpose, three-dimensional reconstruction of the corresponding points is performed using those $\boldsymbol{R}$ and $[\boldsymbol{t}]_X$, and a combination of $\boldsymbol{R}$ and $[\boldsymbol{t}]_X$ is chosen which makes the restored points all positive in front of the camera.

2.8. *Estimating Angles*

The rotation matrix $\boldsymbol{R}$ obtained in the previous section is decomposed into three matrices using Eq.(16).

$$\boldsymbol{R} = \boldsymbol{R}_{roll}\boldsymbol{R}_{pitch}\boldsymbol{R}_{yaw} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{16}$$

$$\boldsymbol{R}_{roll} = \begin{bmatrix} \cos\theta_{roll} & -\sin\theta_{roll} & 0 \\ \sin\theta_{roll} & \cos\theta_{roll} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{17}$$

$$\boldsymbol{R}_{yaw} = \begin{bmatrix} \cos\theta_{yaw} & 0 & \sin\theta_{yaw} \\ 0 & 1 & 0 \\ -\sin\theta_{yaw} & 0 & \cos\theta_{yaw} \end{bmatrix} \tag{18}$$

$$\boldsymbol{R}_{pitch} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{pitch} & -\sin\theta_{pitch} \\ 0 & \sin\theta_{pitch} & \cos\theta_{pitch} \end{bmatrix} \tag{19}$$

Here, roll, pitch, and yaw are the axes shown in Fig. 3, and θ_{roll}, θ_{pitch}, and θ_{yaw} are the angles around respective axes.

Using Eqs. (16), (17), (18) and (19), the three angles are computed by

$$\theta_{roll} = \arctan\frac{-r_{12}}{r_{22}}$$

$$\theta_{pitch} = \arcsin(r_{32})$$

$$\theta_{yaw} = \arctan\frac{-r_{31}}{r_{33}}$$

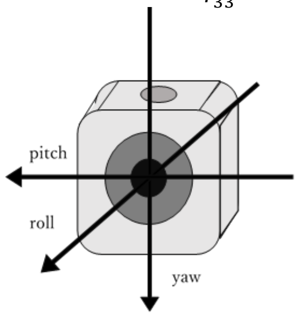


Fig.3. Camera Coordinate System

3 Experimental Results

In the experiment, we estimated walking posture of three types of behavior patterns of four subjects. The input image is obtained by cutting out the video at 3-frame intervals. We find the changes in each axis of the camera coordinate system. Assuming that the attitude angle of the initial state is 0 [deg], the rotation angle of the camera relative to the initial state of the camera is estimated by integrating it. The output obtained from the 9-axis sensor is integrated by the RTQF algorithm, and the value is evaluated comparing with the true value. The following RMSE (Root Mean Squared Error) is used for the evaluation.

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - X_i)^2}$$

where N is the number of data, and x_i and X_i represent the estimated and true values with i=1,2,⋯, N, respectively.

Since the value of the 9-axis sensor is an absolute angle, the initial value of each axis stopped by the sensor is 0 [deg], which is a relative angle.

The experimental results are shown in Table 1, which are the average evaluation results for three types of behavior patterns of four subjects.

Table 1. Indicators of a person's posture by movement

		Types of behavior	Pitch [deg]	Yaw [deg]	Roll [deg]
Average		Sit	3.74	3.98	1.79
		Bend to the side	4.47	8.29	8.36
		Twisting	2.26	11.16	5.77

4 Conclusion

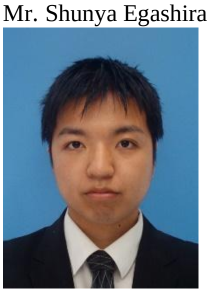
In this paper, we proposed a method of estimating self-posture of a person using MY VISION which employs an ego-camera. In the proposed method, two consecutive images were acquired at 3-frame intervals, and, from these images, the posture of the camera wearer was estimated using epipolar geometry. As the result of the experiment, the average RMSE was 4.74 [deg], and the effectiveness of the proposed method was verified.

5 References

1. Ministry of Health, Labour and Welfare: Overview of the 2019 Basic Survey on National Living Conditions, p.19, 2019.
2. Heisei Medical Association: Posture and Mental Health, 2020.
3. Health Net Sunk. "The way you sit, and the way you walk will extend your healthy life expectancy", 2017.
4. Dushyant. D, Sridhar. S, Sotnychenko. O, Rhodin. H, Shafiei. M, Seidel. H. P, Xu. W, Casas. D, Theobalt. C.: "VNect: real-time 3d human pose estimation with a single rgb camera", ACM Transactions on Graphics, Vol.36, Issue 4, Article No. 44, 2017.
5. S. Ishikawa., J. K. Tan., Kim H.s: "3-D recovery of a non-rigid object from a single camera view employing multiple coordinates representation ", Proceedings 2013 2nd IAPR Asian Conf. on Pattern Recognition: Recent Advances in Computer Vision and Pattern Recognition (RACVPR) , pp. 946−950,2013
6. J. K. Tan, S. Ishikawa.: "Deformable shape recovery by factorization based on a spatiotemporal measurement matrix", Computer Vision and Image Understanding, Vol.82, No.2, pp.101-109, 2001.
7. T. Ohashi, Y. Ikegami, K. Yamamoto, W. Takano, Y. Nakamura: "Video motion capture from the part confidence maps of multi-camera", Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp.4226-4231, 2018.
8. M. Loper, N. Mahmood, J. M. Black: "MoSh: motion and shape capture from sparse markers", ACM Transactions on Graphics (TOG), Article No. 220, pp1-13, 2014.
9. Z. Cao, T. Simon, E. S. Wei, Y. Sheikh: "Realtime multi-person 2d pose estimation using part affinity fields", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.7291-7299, 2017.
10. J. K. Tan, T. Kurosaki.: "Estimation of self-posture of a pedestrian using MY VISION based on depth and motion network", Journal of Robotics, Networking and Artificial Life, Vol.7, No.3, pp.152-155, 2020.
11. M. Ooba., Y. Tanjo.: "A method of recognizing human walk motion from multiple directions", International Journal of Innovative Computing, Information and Control, Vol.20, No.4, pp.1245-1256, 2024.
12. Z. Liu, J. K. Tan.: "Analysis of human walking posture using a wearable camera", International Journal of Innovative Computing, Information and Control, Vol.19, No.3, pp.805-819, 2023.
13. R.I. Hartley: "In Defence of the 8-point Algorithm", Proc. of IEEE Intl. Conf. on Computer Vision, pp.1064-1070,1995

Authors Introduction

Mr. Shunya Egashira



He received his Bachelor's degree in Engineering in 2023 from the Faculty of Engineering, Kyushu Institute of technology in Japan. He is currently a master student in Kyushu Institute of Technology, Japan. He is now interested in human posture analysis using an ego camera.

Prof. Dr. Yui Tanjo

Dr. Tanjo is currently Professor with the Department of Mechanical and Control Engineering, Kyushu Institute of Technology. Her current research interests include ego-motion analysis by MY VISION, three-dimensional shape/motion recovery, human detection, and its motion analysis from video. She was awarded SICE Kyushu Branch Young Author's Award in 1999, the AROB Young Author's Award in 2004, the Young Author's Award from IPSJ of Kyushu Branch in 2004, and the BMFSA Best Paper Awards in 2008, 2010, 2013 and 2015. She is a member of IEEE, The Information Processing Society, The Institute of Electronics, Information and Communication Engineers of Japan.

Analyzing Eye-Tracking Data to Detect Joint Attention in Hexgame Experiments

Yide Yang
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Rena Kato
Kyushu Institute of Technology, 680-4 Kawazu, Iizuka-shi, Fukuoka, 820-8502, Japan

Kosei Shibata
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hendry F. Chame
University of Lorraine, CNRS, LORIA, F-54000, France

Laurent Bougrain
University of Lorraine, CNRS, LORIA, F-54000, France
Sorbonne Université, Institut du Cerveau – Paris Brain Institute -ICM, CNRS, Inria, Inserm, AP-HP, Hopital de la Pitié Salpêtrière, 75013 Paris, France

Hiroaki Wagatsuma
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
Email: yang.yide421@mail.kyutech.jp, katorena.tyt@gmail.com, sibata.kousei927@mail.kyutech.jp, hendry.ferreira-chame@loria.fr, laurent.bougrain@loria.fr, waga@brain.kyutech.ac.jp

Abstract

This study aims to explore the mechanisms of joint attention in a strategic game experiment by analyzing Tobii eye-tracking data. In this experiment, two participants play Hexgame, during which the gaze direction is tracked and projected onto the game's board plane through perspective geometry, in order to track attention sharing correlates at particular stages of the game from behavioral data. The primary focus of this study is the analysis of eye-tracking data to identify attention coordination between players during the progression of the game. Future work will expand this framework to assess win probabilities and predict subsequent moves, providing deeper insights into strategic decision-making.

Keywords: Joint Attention, Hexgame, Eye-Tracking, Perspective Transformation

1. Introduction

Joint attention (JA), where two or more individuals coordinate their gaze towards the same object or event, is fundamental for social interaction and communication [1]. It plays a crucial role in building mutual understanding and effective cooperation, and is also essential in strategic games, as it helps to reveal players' decision-making processes [2].

Previous studies have demonstrated that coordinated gaze between participants often leads to better performance in collaborative contexts, as it enhances the sharing of information and facilitates effective coordination [3]. This ability to synchronize attention is not only important in cooperative settings but also provides insight into the dynamics of shared environments, such as understanding intentions and building shared goals [3].

Joint attention is especially relevant in the context of Autism Spectrum Disorder (ASD), where difficulties in initiating or responding to joint attention can significantly impact social interactions and communication [4]. Research shows that deficits in joint attention hinder social skill development, but targeted interventions can help address these challenges and improve social functioning [4]. Therefore, understanding the mechanisms of joint attention has practical implications for both typical development and enhancing social abilities in individuals with ASD.

Despite extensive research on joint attention in cooperative and developmental settings, its role in competitive environments has received less attention, especially regarding how players anticipate and adapt to each other's strategies during real-time interactions. In competitive games, players must balance their focus between their own actions and their opponent's strategies, creating a dynamic environment where joint attention can become strategically advantageous [5]. Understanding this dynamic can provide deeper insights into player behavior, strategy formation, and decision-making processes.

In this study, we investigate joint attention in the context of competitive games. In particular, we focus on Hexgame, a strategy game that requires players to make rapid decisions. By utilizing Tobii eye-tracking devices, we recorded and analyzed gaze behavior between players. For identifying behavioral correlates of joint attention unfolding in critical phases of the game, we propose a methodology which consists in mapping gaze direction estimations acquired in on-board sensors to a common representation which is the game's board plane. We believe that this method can provide insights into how players adapt to and anticipate each other's moves [4] [5]. Such insights can contribute to broader research on attention dynamics in competitive environments and highlight the importance of shared visual focus in complex, adversarial interactions.

2. Methodology

This study aims to track behavioral correlates of joint attention between two participants by analyzing their eye-tracking data from an Hexgame game experiment. The following section provides a detailed explanation of the experimental setup and the data processing method, including the perspective transformation algorithm.

2.1. *Data Collection*

The experiment setup (Fig.1) is composed of a self-designed 8×8 Hexgame board and pieces. The board followed the standard Hexgame layout, consisting of a regular hexagonal grid with each hexagon's side-to-side distance being 60 mm. The overall board was shaped like a parallelogram, with an 8×8 active game area surrounded by boundary regions. These boundary regions, consisting of one additional row of hexagons on all sides, defined the win conditions for the red and blue sides.

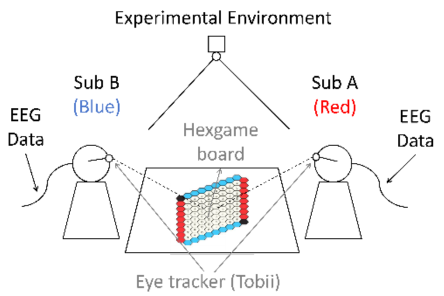


Fig.1 Diagram of Experimental Environment Setup

Hexgame is a two-player strategy game where the goal is to connect two opposite sides of the board with a chain of pieces of the player's color. Players take turns placing pieces, with the red player starting first. The game ends when a connection is made or the board is full. If no winner emerges, the game is a draw. In this experiment, players synchronized their moves with the rhythm to ensure timing consistency in data analysis.

The experiment included 30 red and 30 blue 3D-printed game pieces. A fixed overhead camera recorded the board's movements, capturing the sequence of all piece placements. Two players competed on the board, each wearing a Tobii pro glasses 3 device for eye-tracking, with calibration completed before the experiment. EEG devices were also used, with channels 17 and 34 recording trigger signals (encoded as 2V, 4V, or 8V) for synchronization across devices.

Eight participants were divided into four groups of two, assigned to either the red or blue side. Players acted in sync with a 50 BPM rhythm, completing actions on each down beat. Each game round lasted up to 5 minutes or ended when one player won, with 1-minute breaks between rounds to reduce fatigue. This rhythmic design helps align the timing of in-game events with the EEG data. In the experiment, gaze trajectories were recorded for the participants in real-time. The dataset included images from first-person perspective, timestamp and point coordinates of eye focus on the image, and, as shown in Fig.1, we also recorded from a fixed camera perspective and EEG measures. We hope that this dataset will support future in-depth studies on joint attention mechanisms.

2.2. *Data Processing*

In a preprocessing stage, the recorded videos were converted into individual frames, and typical images containing the chessboard were manually selected. A training consisting of 100 images was constituted. A validation set was also selected, consisting of 20 images. The AI image annotation and dataset creation tool LabelMe was used to annotate the polygons of the chessboard for supervised training of the model.

The data analysis workflow is illustrated in Fig.2. First, gaze data, including timestamps and 2D image coordinates, is extracted from the Tobii data files and formatted into a standardized structure for subsequent processing. By using a convolutional neural network model YOLO [6], the four corners of the Hexgame board are detected in a single pass on images acquired from each participant's perspective and the board's area (i.e. the total pixel count) is calculated. A threshold of 300,000 pixels was chosen for the study, with a criterion that the area must exceed 60% of this threshold to qualify further processing.

The perspective transformation algorithm was implemented by using the OpenCV library. Thus, the *cv2.getPerspectiveTransform* method was employed to calculate a transformation matrix *M* from four pairs of corresponding points between the corners of the board segmented from the ceiling camera's image and the corners detected from the images obtained from the participant's perspective. These transformation matrices are then applied respectively to each participant's perspective coordinates (using *cv2.warpPerspective*), for obtaining their projection onto the chessboard image plane in third-person perspective (as viewed from the ceiling camera). Likewise, the gaze focus location in 2D coordinates (as estimated by the Tobii eye-tracker) are transformed to the third-person perspective for each participant. Resulting from this step, the estimated gaze focus of participants are expressed in a single coordinate frame and board layout, which can facilitate the analysis and tracking of joint attention behavior correlates relative to the game's states.

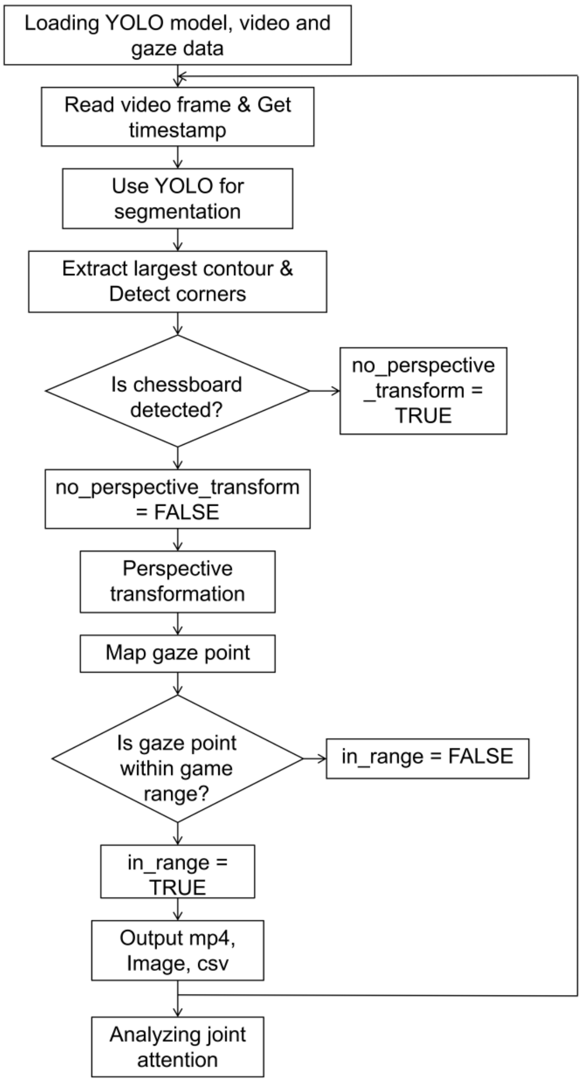


Fig. 2. Data Analysis Flowchart

In cases where the board cannot be detected or its pixel area is considered to be insufficient with respect to the established threshold, the estimation of perspective transformation is skipped for the participant; otherwise, the transformation matrix is calculated and applied to both the image and gaze points. After completing the perspective transformation, the transformed gaze points are further evaluated to determine whether they fall within the board's boundaries.

3. Results and Discussion

As illustrated in Fig.2, our methodological proposal is structured in several stages. Next, we report on the partial results obtained at these stages.

3.1. *Corner Detection Results*

By inspecting the video containing the original gaze points, we observed that transformed gaze points aligned closely with positions on the board (Fig.3), which shows that a reasonably good precision can be obtained from the experiment setup when expressing estimated gaze focus of participants from first-person perspective to a board's third-person perspective.

When analyzing corner segmentation results, we observed that the detection process is influenced by various environmental factors, such as head motions away from the board and participants' hand motions obstructing the board during gameplay.

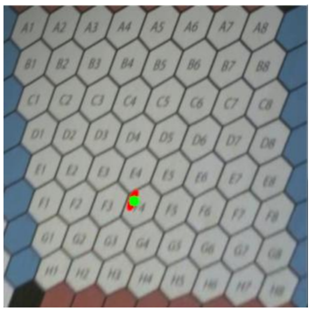
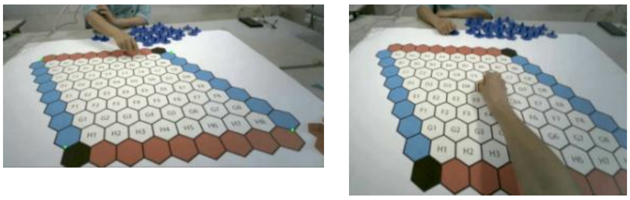


Fig. 3 Comparison of Transformed Gaze with Gaze-Overlaid Video

As shown in Fig.4, momentary occlusions happen during the game situation which compromises the board's segmentation step (Fig.4.b). In the figure it is also shown a situation where small portions of the board's corners occasionally move out of the video frame, leading to detection failures. Frames where the board was not fully detected were discarded. Finally, it is also worth mentioning that for performance reasons, the YOLO architecture was trained with an input size of 640x480 pixels, while the actual resolution of the input video was 1920×1080 pixels. Hence, the detected coordinates were remapped to the original video size which introduced approximation errors estimated to be around 20 pixels.



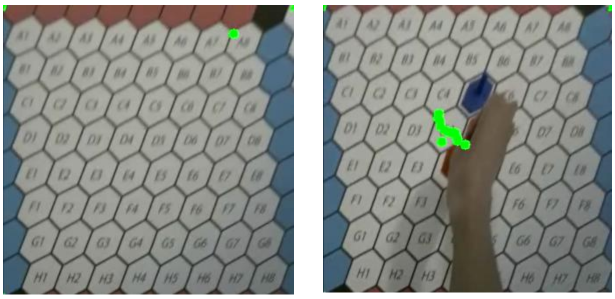
a. Reliable corners detection b. Uncertain corners detection

Fig. 4. Comparison of corners detection results

Despite the limitations addressed above, we observed that corners detection performed reliably when the board was fully contained in the image (Fig.4.a), with results aligning closely with the actual chessboard layout under most game situations.

3.2. *Perspective Transformation Results*

The perspective transformation stage worked well in most cases (Fig.5.a), with successful frames showing a high level of consistency between the mapped chessboard area and gaze points. Thus, most gaze points accurately fall within the chessboard boundaries. As discussed, the perspective transformation stage results are conditioned to successful corners' detection, and the latter was affected by head and arm movements. Consequently, instabilities on gaze focus estimation remapping were observed in cases where small portions of the chessboard corners temporarily moved out of the video frame compromising border detection.



a. Low uncertainty b. Uncertainty when the border is not detected

Fig. 5. Comparison of estimations from good corners detection (left) and the heuristic of reusing the last available transformation when the board is not fully contained in the field of view (right)

As shown in Fig.5b, we tested an heuristic approach for perspective transformation in cases where the segmentation algorithm failed to detect the boards' corners. This heuristic consisted in applying the last available perspective transformation, exploiting spatio-temporal statistical regularities in gaze focus location estimation. This approach showed to be limited since gaze points estimation may fluctuate which can impact further analysis stages and data visualisation, notably, due to distinct behavior style of participants. For instance, it was observed during the game that depending on the player's behavior style, the board can leave the field of view recorded, or be obstructed by arm motions relatively frequently, resulting in fluctuations on estimations. Conversely, when movements are reduced, the heuristic is able to provide more acceptable results.

To address these issues and analyze the impact on further analysis stages and data visualisation, we implemented data preprocessing measures, including visualizing anomalous frames for review and rotating the perspective-transformed images by 180° when necessary to ensure consistent game board orientation. Additionally, we are working on optimizing the robustness of the corner detection model (e.g., using multi-frame fusion detection) and refining code logic to synchronize frame perspective transformation with gaze point transformation to further reduce the occurrence of anomalies. These improvements will enhance the stability of perspective transformation and provide more reliable data for joint attention analysis.

3.3. *Future Developments*

In this ongoing work, we focused on estimating and projecting estimated participants' gaze directions onto the Hexgame board plane, by segmenting the board's corners and applying perspective geometry remapping. We believe that these efforts serve as a foundational step toward the development of more precise methods for tracking joint attention correlated in board grid cells. Preliminary evaluations of our approach revealed that while good results can be achieved when the board is fully visible in each participant's field of view, challenges arise in cases where the board is partially occluded by arms or when participants look away from the board. Consequently, we are working on improving the corner detection stage of our processing pipeline.

Possible improvements include optimizing the corner detection algorithm by considering multiple frame processing and evaluating alternatives or introducing adjustments to the YOLO-based architecture model to suit our research. These improvements will provide a more robust data foundation for future research.

The current results constitute an important basis for subsequent studies. Hence, a comprehensive evaluation of the accuracy of visual focus behavioral tracking in the board plane as an estimate of joint attention correlates is being conducted. This evaluation will involve validating the current outputs, focusing on comparisons with the actual labeled points in the experimental design and conducting simulation tests under known conditions. Future developments will aim at enhancing the performance of existing models and validate them using broader datasets to improve detection stability and robustness.

4. Conclusion

This study proposes an innovative methodological framework that integrates data from eye-tracking devices acquisitions to deep-learning techniques and perspective remapping for tracking joint attention behavior correlates unfolding in an Hexgame game interaction experience. While the partial results reported are limited, they lay a solid foundation for future improvements and applications, and constitute promising directions to study the mechanisms of joint attention in human competitive game interactions.

Acknowledgments

This work was supported in part by JSPS KAKENHI (JP17H06383, JP24K07387), and the joint research project: Collaborative project Université de Lorraine – Kyutech "Human/Robot Social Interactions: engagement and affect analysis during gaming tasks."

References

1. Butterworth, G. (1991). "The ontogeny and phylogeny of joint visual attention." In Natural theories of mind: Evolution, development, and simulation of everyday mindreading, pp. 223-232.
2. Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). "A new look at infant pointing." Child Development, 78(3), 705-722.
3. Richardson, D. C., & Dale, R. (2005). "Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension." Cognitive Science, 29(6), 1045-1060.
4. Mundy, P., & Sigman, M. (2006). "Joint attention, social competence, and developmental psychopathology." In Development and Psychopathology.
5. Cheng, J. T., Gerpott, F. H., Benson, A. J., Bucker, B., Foulsham, T., Lansu, T. A. M., Schülke, O., & Tsuchiya, K. (in press). "Eye gaze and visual attention as a window into leadership and followership: A review of empirical insights and future direction." The Leadership Quarterly, 34(6), 101654.
6. Redmon, J. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition.

Authors Introduction

Mr. Yide Yang

He received his Bachelor's degree in Intelligent Science and Technology in 2023 from the School of Artificial Intelligence and Data Science at Hebei University of Technology, China. He is currently pursuing his Master's degree at the Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Japan.

Ms. Rena Kato

She received her associate's degree from the Department of Electrical and Electronic Systems Engineering, National Institute of Technology, Toyota College, Japan in 2024. She is currently enrolled in the Department of Systems Design and Informatics, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, Japan.

Mr. Kosei Shibata

He received his Master's degree in Engineering in 2023 from the Graduate School of Life Science and System Engineering, Kyushu Institute of technology (Kyutech) in Japan,. He is currently a doctoral course student in Kyutech.

Dr. Hendry F. Chame

He holds degrees in Psychology and Computer Science (UCV, Venezuela), M.S. degrees in Instructional Psychology and Robotics and a Ph.D. degree in Robotics (ECN, France). His postdoctoral research includes bio-inspired (FURG, Brazil), neural (OIST, Japan), and cognitive robotics (LAAS, France). In 2021, he became an Associate Professor at the University of Lorraine, focusing on AI, robotics, and complex systems at LORIA.

Dr. Laurent Bougrain

He received his M.S. degree in artificial intelligence in 1995 from university of Paris 6, a B. degree in psychology and Ph.D. degrees in computer science in 2000 from université of Lorraine. He is an associate professor at the university of Lorraine (France) since 2001. He is the leader of the NeuroRhythms team of the department of the complex system, artificial intelligence and robotics at LORIA laboratory (France).

Dr. Hiroaki Wagatsuma

He received his M.S., and Ph.D. degrees from Tokyo Denki University, Japan, in 1997 and 2005, respectively. In 2009, he joined Kyushu Institute of Technology, where he is currently an Associate Professor of the Department of Human Intelligence Systems. His research interests include non- linear dynamics and robotics. He is a member of IEEE.