ICAROB 2021

PROCEEDINGS OF THE 2021 INTERNATIONAL CONFERENCE ON ARTIFICIAL LIFE AND ROBOTICS

January 21 to 24, 2021
On line
26th AROB International Meeting Series

Editor-in-Chief
Masanori Sugisaka
Editors: Yingmin Jia, Takao Ito, Ju-Jang Lee

ISBN 978-4-9908350-6-4

Contents

1	Organization, etc.	1
2	Messages	12
3	Time Table	17
4	Opening Ceremony	20
5	Technical paper index	21
6	Abstracts	
6-1	PS abstracts	41
6-2	OS abstracts	42
6-3	GS abstracts	89
7	Authors index	100

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)

Higashihiroshima City (Japan)

CO-ORGANIZED BY

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

ADVISORY COMMITTEE CHAIRS

Moshe Kam (New Jersy Institute of Technology, Former IEEE President, USA)
Kensuke Harada (Osaka University, Japan)

ADVISORY COMMITTEE

Adam Grzech (Wroclaw University of Technology, Poland)
Bruce Eisenstein (Drexel University, Former IEEE President, USA)
Hidenori Kimura (RIKEN & Wasada University, Japan)
Jeffrey Johnson (The Open University, UK)
Jerzy Świątek (Wroclaw University of Technology, Poland)
Joshua M. Epstein (The Johns Hopkins University, USA)
Kai-Tai Song (National Chiao Tung University, Taiwan)
Ken-ichi Tanaka (Executive Fellow, Mitsubishi Electric Corporation, Japan)
Masato Nakagawa (Executive Fellow, DENSO CORPORATION, Visiting Professor of Hiroshima University, Japan)
Masayoshi Tomizuka (University of California Berkeley, USA)
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)
Steen Rasmussen (University of Southern Denmark, Denmark)
Zengqi Sun (Tsighua Univesity, P. R. China)

GENERAL CHAIR

Masanori Sugisaka
(Alife Robotics Corporation Ltd., Japan)
(Visiting Professor, The Open University (UK))

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)
Jangmyung Lee (Pusan National University, Korea)
Luigi Pagliarini (Technical University of Denmark, Denmark)
(Academy of Fine Arts of Macerata, Italy)
Mohd Rizon bin Juhari (University of UCSI, Malaysia)

PROGRAM CHAIRMAN

Makoto Sakamoto (University of Miyazaki, Japan)

CO-PROGRAM CHAIR

Marion Oswald (Technische Universität Wien, Austria)

INTERNATIONAL ORGANIZING COMMITTEE

Akira Fukuda (Kyushu University, Japan)
Akira Nakamura (AIST, Japan)
Akinori Abe (Chiba University, Japan)
Andrew Gadsden (University of Guelph, Canada)
Caoli Wang (University of Shanghai for Science and Technology, P. R. China)
Chan Gook Park (Seoul National University, Korea)
Eiji Hayashi (Kyushu Institute of Technology, Japan)
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)
Hideyuki Suzuki (The University of Tokyo, 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)
Kevin Voges (Canterbury University, New Zealand)
Kunikazu Kobayashi (Aichi Prefectural University, Japan)
Kuo-Hsien Hsia (Far East University, Taiwan)
Kuo-Lan Su (National Yunlin University of Science and Technology, Taiwan)
Kyungho Park (U.S. ARMY, Japan and USA)
Masao Kubo (National Defense Academy of Japan, Japan)
Masanao Obayashi (Yamaguchi University, 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)
Shin Jisun (Pusan National University, Korea)
Singo Mabu (Yamaguchi University, Japan)
Takashi Kohno (The University of Tokyo, Japan)
Takashi Ogata (Iwate Prefectural University)
Teruhisa Hochin (Kyoto Prefectural University, Japan)
Tetsuro Hattori (Kagawa University, Japan)
Thi Thi Zin (University of Miyazaki, Japan)
Thomas S. Ray (University of Oklahoma, USA)
Toru Yamamoto (Hiroshima University, Japan)
Toshihro Inukai (DENSO WAVE INCORPORATED, 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)
Zengqiang Chen (Nankai University, P. R. China)

INTERNATIONAL PROGRAM COMMITTEE

Abdul Rahman bin Dullah (Universiti Teknikal Malaysia Melaka, Malaysia)
Akinori Abe (Chiba University, Japan)
Akira Utsumi (The University of Electro-Communications, 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 (Kyushu Institute 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)
Dongmei Ai (Universty of Science and Technology Beijing, 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)
Junping Du (Beijing University of Posts and Telecommunications, 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)

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 (Yamaguchi University, Japan)
Takayoshi Yamada (Gifu University, Japan)
Takuya Fuginaga (Kyushu Institute of Technology, Japan)
Takuya Nanami (The University of Tokyo. Japan)
Taishiro Kishimoto (Keio University, Japan)
Taisuke Akimoto (Kyushu Institute of Technology, Japan)
Takashi Iwamoto (Mitsubishi Electric Corporation, Japan)
Tarik bin Abd Latef (Universiti Malaya, Malaysia)
Taro Shibanoki (Ibaraki University, Japan)
Teruhisa Hochin (Kyoto Prefectural 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 (Tokyo Institute of Technology, 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, Nippon Bunri University (NBU), 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 this journal, International Journal of Robotics, Networking and Artificial Life (JRNAL). From now on, ALife Robotics Corporation Ltd. is in charge of management of both the conference and the journal. 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

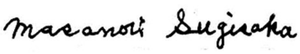
Environment navigation and localization
Evolutionary computations
Facial expression analysis, music recommendation and augmented reality
Foundation of computation and its application
Fuzzy control
Genetic algorithms
Human-welfare robotics
Image processing
Insect-like aero vehicles
Intelligence in biological 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

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 Co., Ltd., Japan)
(Visiting Professor, The Open University, UK)



Masanori Sugisaka

General Chair of ICAROB

It is my great honor to invite you all to The 2021 International Conference on Artificial Life and Robotics (ICAROB 2021). This Conference is changed as the old symposium from the first (1996) to the Eighteenth (2013) annually which were organized by Oita University, Nippon Bunri University(NBU), 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 in **SCOPUS** and **CPCI-Web of Science** now. From now on, ALife Robotics Corporation Ltd. is in charge of management. This year we have The 2021 International Conference on Artificial Life and Robotics (ICAROB2020) (26th 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 hope that fruitful discussions and exchange of ideas between researchers during Conference (ICAROB2021) 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, R .P. China)

Yingmin Jia

Co-General Chair of ICAROB

It is my great pleasure to invite you to The 2021 International Conference on Artificial Life and Robotics (ICAROB 2021), in Higashi Hiroshima Arts & Culture Hall Kurara, Hiroshima, Japan from January 21-24, 2021.

ICAROB develops from the AROB that was created in 1996 by Prof. Masanori Sugisaka and will celebrate her 26th birthday in 2021. 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.

For a successful meeting, many people have contributed their great efforts to 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 meeting you at ICAROB in on line and wishing you enjoy your stay in on line.

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

Takao Ito

Co General Chair of ICAROB

It is my great honor to invite you all to the 2021 International Conference on Artificial Life and Robotics (ICAROB 2021), to the wonderful city of Higashi-Hiroshima, Hiroshima Prefecture, Honshu Island, Japan.

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 AROB. Our conference has brought together many research scholar, faculty members, and graduate students from all over the world, and published many 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. ICAROB has provided a foundation for unifying the exchange of scientific information on the study 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 is on line. I eagerly look forward to personally meeting you in on line, during the ICAROB 2021 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 8th ICAROB 2021 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 8th ICAROB.

I'm looking forward to meeting you at the 8th ICAROB in on line and wishing you all the best.

GENERAL SESSION TOPICS

GS1 Chaos & Application (4)	GS2 Image & Signal Processing 1 (4)
GS3 Image & Signal Processing 2 (3)	GS4 Image & Signal Processing 3 (3)
GS5 Genetic Algorithms & Robotics 1 (3)	GS6 Robotics 1 (4)
GS7 Robotics 2 (4)	GS8 Medical Data Processing (6)

ORGANIZED SESSION TOPICS

OS1 Human-Machine Interface Application (4)	OS2 Media Information Processing, Music Recommendation and Artificial Intelligence (4)
OS3 Robot Control (9)	OS4 Intelligent UAV and Systems (3)
OS5 Advanced Information Processing Applications (6)	OS6 Intelligent Control (6)
OS7 Information Applications and Security (5)	OS8 Intelligent Systems and Control (5)
OS9 Intelligent Systems and Life (7)	OS10 Advanced Control Systems and Signal Processing (7)
OS11 Robotic Manipulation (3)	OS12 System and Control (18)
OS13 Intelligent Systems and Robotics (14)	OS14 Image Processing (4)
OS15 Bridging the Gap Between AI, Cognitive Science, and Narratology (7)	OS16 Software Development Support Method (5)
OS17 Robotics and machine vision (3)	OS18 Natural Computing (3)
OS19 Artificial Intelligence for Embedded Systems and Robotics (4)	OS20 Virtual Reality and Intelligence Interactions (3)
OS21 Educational Application of Control Technology (4)	OS22 Robot Competitions and Education (5)
OS23 Advances in Field Robotics and Their Applications (6)	OS24 Mathematical Informatics (5)

1/21(Thu.)	Group meeting for the conference (ZOOM ID: 984 9796 8944)
1/22(Fri.) – 1/24(Sun.)	ICAROB Secretariat
1/24(Sun)	Group meeting for the next conference (ZOOM ID: 914 3254 0230)

TIME TABLE (1/22)

1/22(Fri.)	**Meeting Room 1** (ZOOM ID: 993 2436 1751)	**Meeting Room 2** (ZOOM ID: 986 1953 6512)	**Meeting Room 3** (ZOOM ID: 984 4367 9770)	**Meeting Room 4** (ZOOM ID: 971 5454 5491)
8:40-	Registration			
9:00-10:00	OS19 Artificial Intelligence for Embedded Systems and Robotics (4) Chair: Hakaru Tamukoh	OS21 Educational Application of Control Technology (4) Chair: Kazuo Kawada	GS2 Image & Signal Processing 1 (4) Chair: Tohru Kamiya	GS5 Genetic Algorithms & Robotics 1 (3) Chair: Hideyuki Tanaka
10:00-10:20	Coffee break			
10:20-10:50	Opening Ceremony (ZOOM ID: 925 6396 7074)			
11:00-12:00	Chair: Takao Ito Plenary Speech PS1 (ZOOM ID: 966 3705 6919) Kazuo Ishii			
12:00-13:00	Lunch			
13:00-14:15	OS2 Media Information Processing, Music Recommendation and Artificial Intelligence (4) Chair: Chair: Yasunari Yoshitomi	OS8 Intelligent Systems and Control (5) Chair Chung-Wen Hung	OS24 Mathematical Informatics (5) Chair: Takao Ito	GS3 Image & Signal Processing 2 (3) Chair: Keiko Sakurai
14:15-15:15	Coffee break			
15:15-17:00	OS9 Intelligent Systems and Life (7) Chair: Kuo-Hsien Hsia	OS10 Advanced Control Systems and Signal Processing (7) Chair: Takuya Kinoshita	OS15 Bridging the Gap Between AI, Cognitive Science, and Narratology (7) Chair: Jumpei Ono	OS23 Advances in Field Robotics and Their Applications (6) Chair: Keisuke Watanabe

TIME TEBLE (1/23)

1/23(Sat.)	Meeting Room 1 (ZOOM ID: 993 2436 1751)	Meeting Room 2 (ZOOM ID: 986 1953 6512)	Meeting Room 3 (ZOOM ID: 984 4367 9770)	Meeting Room 4 (ZOOM ID: 971 5454 5491)
8:40-	Registration			
9:00-10:15	OS14 Image Processing (4) Chair: Joo Kooi Tan	OS16 Software Development Support Method (5) Chair: Tetsuro Katayama	OS22 Robot Competitions and Education (5) Chair: Yasunori Takemura	GS6 Robotics 1 (4) Chair: Noritaka Sato
10:15-11:00	Coffee break			
11:00-12:00	Chair: Masayoshi Tabuse Plenary Speech PS2 (ZOOM ID: 912 8494 3957) Yasunari Yoshitomi			
12:00-13:00	Lunch			
13:00-14:30	OS5 Advanced Information Processing Applications (6) Chair: Toru Hiraoka	OS6 Intelligent Control (6) Chair: Yingmin Jia	OS7 Information Applications and Security (5) Chair: Kuo-Hsien Hsia	GS1 Chaos & Application (4) Chair: Masato Nagayoshi
14:30-15:30	Coffee break			
15:30-16:45	OS11 Robotic Manipulation (3) Chair: Kensuke Harada	OS17 Robotics and machine vision (3) Chair: Jiwu Wang	OS18 Natural Computing (3) Chair: Marion Oswald	GS8 Medical Data Processing (6) Chair: Shingo Mabu

TIME TABLE (1/24)

1/24(Sun.)	Meeting Room 1 (ZOOM ID: 993 2436 1751)	Meeting Room 2 (ZOOM ID: 986 1953 6512)	Meeting Room 3 (ZOOM ID: 984 4367 9770)	Meeting Room 4 (ZOOM ID: 971 5454 5491)
8:40-	Registration			
9:00-9:45	OS4 Intelligent UAV and Systems (3) Chair: Young Im Cho	GS4 Image & Signal Processing 3 (3) Chair: Taro Shibanoki		
9:45-11:00	Coffee break			
11:00-12:00	OS3 Robot Control (9) Chair: Yizhun Peng	OS13 Intelligent Systems and Robotics (14) Chair: Fengzhi Dai	OS20 Virtual Reality and Intelligence Interactions (3) Chair: R.P.C. Janaka Rajapakse	GS7 Robotics 2 (4) Chair: Wisanu Jitviriya
12:00-13:00	Lunch			
13:00-14:15	OS3 Robot Control (9) Chair: Yizhun Peng	OS12 System and Control (18) Chair: Huailin Zhao (18)	OS1 Human-Machine Interface Application (4) Chair: Norrima Mokhtar	
Group meeting for the next conference (15:00-15:30) (ZOOM ID: 914 3254 0230)				

The 2021 International Conference on ARTIFICIAL LIFE AND ROBOTICS (ICAROB2021)

January 21 (Thursday)

17:30-19:30 Group meeting for the conference (ZOOM ID: 984 9796 8944)

January 22 (Friday)

10:20-10:50

Opening Ceremony (ZOOM ID: 925 6396 7074)

Chair: Kuo-Hsien Hsia (National Yunlin University of Science and Technology, Taiwan)

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 (KAIST, Korea)

January 23 (Saturday)

Banquet: on line (ZOOM ID: 971 4311 0144)

18:30-20:30

Chair: Takao Ito (Hiroshima University, Japan)

Welcome Addresses

Prof. Yingmin Jia (Beihang University, China)
Prof. Ju-Jang Lee (KAIST, Korea)
Prof. Jangmyung Lee (Pusan National University, Korea)

TECHNICAL PAPER INDEX

January 22 (Friday)

8:40-Registration

Conference Room:
10:20-10:50 Opening Ceremony (ZOOM ID: 925 6396 7074)
Chair: Tetsuro Katayama (The University of Miyazaki, Japan)

11:00-12:00 (ZOOM ID: 966 3705 6919)
Plenary Speech PS1,
Chair: Takao Ito (Hiroshima University, Japan)

PS1 *Tomato-harvesting-robot competition towards smart agriculture*
Kazuo Ishii (Kyushu Institute of Technology, Japan)

Meeting Room 1 (ZOOM ID: 993 2436 1751)
9:00-10:00 OS19 Artificial Intelligence for Embedded Systems and Robotics (4)
Chair: Hakaru Tamukoh (Kyushu Institute of Technology, Japan)

OS19-1 *Convolutional Network with Sub-Networks*
Ninnart Fuengfusin, Hakaru Tamukoh (Kyushu Institute of Technology, Japan)

OS19-2 *Influence of FPGA implementation methods in High-Level Synthesis*
[1]Yusuke Watanabe, [2]Hakaru Tamukoh
([1]RAFT WORK Co., Ltd, Japan, [2]Kyushu Institute of Technology, Japan)

OS19-3 *A Hardware-Oriented Random Number Generation Method and A Verification System for FPGA*
Sansei Hori, Hakaru Tamukoh (Kyushu Institute of Technology, Japan)

OS19-4 *Synthesis of realistic food dataset using generative adversarial network (GAN) based on RGB and depth images*
Obada Al Aama, Hakaru Tamukoh (Kyushu Institute of Technology, Japan)

13:00-14:00 OS2 Media Information Processing, Music Recommendation and Artificial Intelligence (4)
Chair: Yasunari Yoshitomi (Kyoto Prefectural University, Japan)
Co-Chair：Masayoshi Tabuse (Kyoto Prefectural University, Japan)

OS2-1 *Comparison of Data Augmentation Methods in Pointer-Generator Model Using Various Sentence Ranking Methods*
Tomohito Ouchi, Masayoshi Tabuse (Kyoto Prefectural University, Japan)

OS2-2 *Music Recommendation System Driven by Variations in Fingertip Skin Temperature*
Mayuka Wada[1], Taro Asada[2], Yasunari Yoshitomi[2], Masayoshi Tabuse[2]

([1]Shimadzu System Development Corp., Japan, [2]Kyoto Prefectural University, Japan)

OS2-3 *Music Recommendation System Driven by Interaction between User and Personified Agent Using Speech Recognition, Synthesized Voice and Facial Expression*
Ayumi Matsui[1], Miki Sakurai[2], Taro Asada[3], Yasunari Yoshitomi[3], Masayoshi Tabuse[3]
([1]Sumitomo Mitsui Card Co., Ltd., Japan, [2]TIS Inc., Japan, [3]Kyoto Prefectural University, Japan)

OS2-4 *Wallet Operation Evaluation System Using Deep Learning*
Junichiro Yamawaki[1], Yasunari Yoshitomi[2], Masayoshi Tabuse[2], Taro Asada[2]
([1]SKY Co., Ltd., Japan, [2]Kyoto Prefectural University, Japan)

15:15-17:00 OS9 Intelligent Systems and Life (7)
Chair: Kuo-Hsien Hsia (National Yunlin University of Science and Technology, Taiwan)
Co-Chair: Evgeni Magid (Kazan Federal University, Russia)

OS9-1 *Architecture of a student training computer program for preparing professional outpatient consulting skills within an electronic medical records system during COVID-19 alertness situation*
Sergey Bulatov[1], Evgeni Magid[2], Enzhe Kharisova[1], Roman Lavrenov[2], Vitaly Dudin[1], Artur Khazetdinov[2] ([1]Kazan State Medical University, Russia), ([2]Kazan Federal University, Russia)

OS9-2 *Satisfaction Assessment on the Counseling Service System for Full-Time Teacher-Counselor in Tainan Elementary School*
Hsiu-Hao Liu (Chang Jung Christian University, Taiwan)
Yun-Syuan Jhang (National University of Tainan, Taiwan)

OS9-3 *Assess The Critical Factors for the Counseling Service System Usage Intention*
Hsiu-Hao Liu (Chang Jung Christian University, Taiwan)

OS9-4 *Design and Implementation of EMI Suppression Filter for Electronic Commutation Fans*
Ching-Chun Chuang, Chih-Chiang Hua, Chung-Wen Hung, Chun-Jen Yao
(National Formosa University, Taiwan)

OS9-5 *Application of IOT to Forest Management Taking Fushan Botanical Garden as an Example*
Shuo-Tsung Chen, Chih-Chiang Hua, Ching-Chun Chuang
(National Formosa University, Taiwan)

OS9-6 *Development of the IoT Module using MQTT Protocol and AES*
Jr-Hung Guo, Tzu-Yuan Lin, Kuo-Hsien Hsia
(National Yunlin University of Science and Technology, Taiwan)

OS9-7 *Exploring the Intention to Continuance of Learning Programming at Elementary School of Rural Area by the mBot Robot*
Yung-Hsin Cheng, Jia-Ming Hsiao
(National Yunlin University of Science and Technology, Taiwan)

Meeting Room 2 (ZOOM ID: 986 1953 6512)
9:00-10:00 OS21 Educational Application of Control Technology (4)
Chair: Kazuo Kawada (Hiroshima University, Japan)
Co-Chair: Yoshihiro Ohnishi (Ehime University, Japan)

OS21-1 *Development of Shock Sensitive Tiny Dummy Robot for Junior High School Rescue Robot Challenge*
Kazuo Kawada, Keisuke Iuchi, Keita Murai, Hiroyuki Y. Suzuki
(Hiroshima University, Japan)

OS21-2 *Basic Research on Parameter Tuning Skills Evaluation Based on Sensor Car Behavior Data in Technology Education*
Teruyuki Tamai (Hiroshima University / Ehime University),
Yoshihiro Ohnishi (Ehime University), Kazuo Kawada(Hiroshima University)

OS21-3 *Mini Windmill Generator Kit for homework for Hiroshima Univ. Monozukuri Junior Doctor Special Educational Program*
Hiroyuki Y. Suzuki, Kazuo Kawada, Masayasu Nagamatsu
(Hiroshima University, Japan)

OS21-4 *Development of Teaching Materials to Learn the Efficient Use of Energy*
Yoshihiro Ohnishi[1], Teruyuki Tamai[1], Shinnosuke Mori[1] Kawada Kazuo[2]
([1]Ehime University, [2]Hiroshima University, Japan)

13:00-14:15 OS8 Intelligent Systems and Control (5)
Chair: Chung-Wen Hung (National Yunlin University of Science and Technology, Taiwan)
Co-Chair: Kuo-Hsien Hsia (National Yunlin University of Science and Technology, Taiwan)

OS8-1 *Networking Integration Application of an Intelligent Production Line for Aerospace Precision Manufacturing*
Chau-Chung Song, Chun-Chi Wang, Chen-Pang Chen, Chung-Wen Hung
(National Yunlin University of Science and Technology, Taiwan)

OS8-2 *Low-Cost Indoor Localization Using Sound Spectrum of Light Fingerprints*
Chung-Wen Hung, Hiroyuki Kobayashi, Jun-Rong Wu, Chau-Chung Song
(National Yunlin University of Science and Technology, Taiwan)

OS8-3 *Air Valve Fuzzy Control Combined with Sheet Music Recognition Techniques Applied to Autoplaying Soprano Recorder Machines*
Chun-Chieh Wang, Guang-Ming Jhang (Chienkuo Technology University, Taiwan)

OS8-4 *Image Inpainting Techniques Combined with Isolated Pixel Filtering Applied to Multifunctional Drawing Robots*
Chun-Chieh Wang, Zhan-Xian Ye (Chienkuo Technology University, Taiwan)

OS8-5 *Mobile Robot with Image Recognition -- Using LabVIEW and KNRm*
Kuo-Hsien Hsia, Bo-Jung Yang, Jr-Hung Guo, Chang-Sheng Xiao
(National Yunlin University of Science and Technology, Taiwan)

15:15-17:00 OS10 Advanced Control Systems and Signal Processing (7)

Chair: Takuya Kinoshita (Hiroshima University, Japan)
Co-Chair: Shinichi Imai (Tokyo Gakugei University, Japan)

OS10-1 *Nonlinear Internal Model Controller based on Local Linear Models, and its Application*
Shinichi Imai (Tokyo Gakugei University, Japan)

OS10-2 *Design of a data-driven control system for reference model design using predicted signals*
Yuki Nakatani, Takuya Kinoshita, Toru Yamamoto (Hiroshima University, Japan)

OS10-3 *Design of a Data-driven GMV Controller Using the Nelder-Mead Method*
LiYing Shi, Zhe Guan, Toru Yamamoto (Hiroshima University, Japan)

OS10-4 *Design of a Databased-Driven GPC for Nonlinear Systems*
Zhe Guan, Toru Yamamoto (Hiroshima University, Japan)

OS10-5 *Design of a Data-Driven Controller using Open-Loop Data*
Y. Nishiya, Takuya Kinoshita, Toru Yamamoto (Hiroshima University, Japan)

OS10-6 *Improved Estimation of Sway-Angle for Overhead Crane based on Phase Difference of Acoustic Signals in Frequency Domain*
Hanako Ogawa, Takeshi Yamada, Masayoshi Nakamoto
(Hiroshima University, Japan)

OS10-7 *Study on an Optimal Design Method for Control Systems based on Bayesian Optimization*
Koichi Hirota, Shin Wakitani, Toru Yamamoto (Hiroshima University, Japan)

Meeting Room 3 (ZOOM ID: 984 4367 9770)
9:00-10:00 GS2 Image & Signal Processing 1 (4)
Chair: Tohru Kamiya (Kyushu Institute of Technology, Japan)

GS2-1 *Unconstrained Measurement of Heart Rate Considering Harmonics of a Respiratory Signal by Tactile Sensor*
Kazuya Matsuo (Kyushu Institute of Technology, Japan), Toshiharu Mukai (Meijo University, Japan), Shijie Guo (Hebei University of Technology, China)

GS2-2 *Image Registration Method for Chest MDCT Images Based on 2-D Finite Element Method*
Takuji Ogimoto, Tohru Kamiya (Kyushu Institute of Technology, Japan)
Takatoshi Aoki (University of Occupational and Environmental Health, Japan)

GS2-3 *Detection of Abnormal Shadows in Low-dose CT Images Using CNN*
Hiromu Ikeda, Tohru Kamiya (Kyushu Institute of Technology, Japan)

GS2-4 *EEG signal extraction method based on HHT and CSP*
Lei Gong (Tianjin University of Science and Technology, China)

13:00-14:15 OS24 Mathematical Informatics (5)
Chair: Takao Ito (Hiroshima University, Japan)
Co-Chair: Makoto Sakamoto (University of Miyazaki, Japan)

OS24-1 *Verification of CG character operation by brain wave discrimination*
Kenji Sakoma[1], Kodai Miyamoto[1], Taketo Kamasaka[1], Makoto Sakamoto[1], Amane Takei[1], Tsutomu Ito[2], Takao Ito[3] ([1]University of Miyazaki, Japan)
([2]National Institute of Technology, Ube College, Japan)
([3]Hiroshima University, Japan)

OS24-2 *Approach to tourism support by aerial video using CG*
Kenji Sakoma[1], Noritake Seto[1], Kodai Miyamoto[1], Taketo Kamasaka[1], Makoto Sakamoto[1], Amane Takei[1], Tsutomu Ito[2], Takao Ito[3] ([1]University of Miyazaki, Japan)
([2]National Institute of Technology, Ube College, Japan)
([3]Hiroshima University, Japan)

OS24-3 *Method for detecting eye misalignment based on movement near the center of the pupil*
Uchida Noriyuki[1,2], Kayoko Takatuka[2], Hisaaki Yamaba[2], Masayuki Mukunoki[2], Naonobu Okazaki[2]
([1]Kyushu University of Health and Welfare, Japan), ([2]University of Miyazaki, Japan)

OS24-4 *Basic research on video production for educational support by virtual technology*
Kodai Miyamoto[1], Taketo Kamasaka[1], Kenji Sakoma[1], Makoto Sakamoto[1], Amane Takei[1], Tsutomu Ito[2], Takao Ito[3] ([1]University of Miyazaki, Japan) ([2]National Institute of Technology, Ube College, Japan) ([3]Hiroshima University, Japan)

OS24-5 *There is a movement towards the development of hula costume CAD*
Taketo Kamasaka[1], Kenji Sakoma[1], Kodai Miyamoto[1], Makoto Sakamoto[1], Amane Takei[1], Tsutomu Ito[2], Takao Ito[3] ([1]University of Miyazaki, Japan), ([2]National Institute of Technology, Ube College, Japan) ([3]Hiroshima University, Japan)

15:15-17:00 OS15 Bridging the Gap Between AI, Cognitive Science, and Narratology (7)
Chair: Jumpei Ono (Aomori University, Japan)
Co-Chair: Hiroki Fukushima (Kyushu Womens' University, Japan)
Co-Chair: Takashi Ogata (Iwate Prefectural University, Japan)

OS15-1 *Analysis and Construction of Elements of the Stage Performance Structure in a Kabuki-dance*
Miku Kawai (Iwate Prefectural University, Japan), Jumpei Ono (Aomori University, Japan), Takashi Ogata (Iwate Prefectural University, Japan)

OS15-2 *Unchiku Generation with Moving Illustration Using Kabuki Knowledge*
Jumpei Ono (Aomori University, Japan), Miku Kawai (Iwate Prefectural University, Japan), Takashi Ogata (Iwate Prefectural University, Japan)

OS15-3 *Implementing Story Units of Japanese Folktales with Conceptual Dictionaries*
Takuya Ito (Iwate Prefectural University, Japan), Jumpei Ono (Aomori University, Tokyo) Takashi Ogata (Iwate Prefectural University, Japan)

OS15-4 *The Story Generation Process and Cognitive Biases*
Jun Nakamura (Chuo University, Japan)

OS15-5 *Time in an Aesthetic Experience of a cup of Sake*
Hiroki Fukushima (Kyushu Womens' University, Japan)

OS15-6 *Extension of Clinical/Psychological Approach Using Post-Narratology: Possibility of application on Artificial Intelligence and Robot*
Kai Seino (National Rehabilitation Center for Persons with Disabilities, Japan)

OS15-7 *A relationship between narratology and marketing*
Akinori Abe (Chiba university, Japan)

Meeting Room 4 (ZOOM ID: 971 5454 5491)
9:00-9:45 GS5 Genetic Algorithms & Robotics 1 (3)
Chair: Hideyuki Tanaka

GS5-1 *Layout decision system for multiple production lines using work-flow-line and GA*
Masato Noda[1], Hidehiko Yamamoto[1] (Gifu University, Japan),
Hirohumi Tsuji[2], Yasuhisa Terawa[2], Yoshinori Nakamura[2], Masayuki Tsuchida[2] (Infofarm Corporation, Japan),
Katsuaki Yamada[3], Yukiyasu Kuriyama[3] (Kai Industries Corporation, Japan)

GS5-2 *Autonomous decentralized FMS's AGVs moving control by mind change with deep learning*
Ryunosuke Yamane, Hidehiko Yamamoto (Gifu University)

GS5-3 *Secondary School Robotics Education in Camarin High School: Developments and Challenges for Improvement*
Jeffrey Rivera Galino (Camarin High School, Philippines and Hiroshima University, Japan), Hideyuki Tanaka (Hiroshima University, Japan)

13:00-13:45 GS3 Image & Signal Processing 2 (3)
Chair: Keiko Sakurai (University of Miyazaki, Japan)

GS3-1 *Deep-Learning Based Segmentation Algorithm for Defect Detection in Magnetic Particle Testing*
Akira Ueda, Huimin Lu, Tohru Kamiya (Kyushu Institute of Technology, Japan)

GS3-2 *A Method for Improving Recognition of Lying Postures Using a Measured Signal Intensity of Respiration and Heartbeat by Flexible Tactile Sensor Sheet*
Kazuya Matsuo (Kyushu Institute of Technology, Japan)
Toshiharu Mukai (Meijo University, Japan),
Shijie Guo (Hebei University of Technology, China)

GS3-3 *Basic research for the realization of online MEG using SSD*
Kazuhiro Yagi, Yuta Shibahara, Lindsey Tate, Keiko Sakurai, Hiroki Tamura

(University of Miyazaki, Japan)

15:15-17:00 OS23 Advances in Field Robotics and Their Applications (6)
Chair: Keisuke Watanabe (Tokai University)
Co-Chair: Kazuo Ishii (Kyushu Institute of Technology)

OS23-1 *Three-dimensional Measurement Using Laser Pattern And Its Application to Underwater Scanner*
Yuya Nishida, Tomoya Shinnoki, Shinsuke Yasukawa, Kazuo Ishii
(Kyushu Institute of Technology, Japan)

OS23-2 *Motion control of a cable-restricted underwater vehicle for long-term spot observation*
Yoshiki Tanaka, Yuya Nishida, Kazuo Ishii (Kyushu Institute of Technology, Japan)

OS23-3 *Development of current sensors for digitizing expert knowledge in fish feeding Towards sustainable aquaculture*
Dominic B. Solpico, Yuya Nishida, Kazuo Ishii (Kyushu Institute of Technology, Japan)

OS23-4 *Underwater image reconstruction using convolutional auto-encoder*
Shinsuke Yasukawa, Sreeraman Srinivasa Raghura, Yuya Nishida, Kazuo Ishii
(Kyushu Institute of Technology, Japan)

OS23-5 *Spherical Robot Transfer Problem with Minimal Total Kinetic Energy*
Kenji Kimura[1], Kazuo Ishii[2]
([1]Fukuoka Daiichi High School, Japan, [2]Kyushu Institute of Technology, Japan)

OS23-6 *Acoustic Impedance Measurement through the Modelling of Ultrasonic Wave Transmission*
Ryuugo Mochizuki, Yuya Nishida, Kazuo Ishii (Kyushu Institute of Technology, Japan)

OS23-7 *Installation Method of In-situ Drilling Platform(withdraw)*
Keisuke Watanabe (Tokai University, Japan)

January 23 (Saturday)
8:40-Registration

Conference Room (ZOOM ID: 912 8494 3957)
11:00-12:00
Plenary Speech PS2
Chair: Masayoshi Tabuse (Kyoto Prefectural University, Japan)

PS2 *Human-computer Communication Using Facial Expression*

Yasunari Yoshitomi (Kyoto Prefectural University, Japan)

Meeting Room 1 (ZOOM ID: 993 2436 1751)
9:00-10:15 OS14 Image Processing (5)
Chair: Joo Kooi Tan (Kyushu Institute of Technology, Japan)
Co-Chair: Seiji Ishikawa (Kyushu Institute of Technology, Japan)

OS14-1 *Detection of a Fallen Person and Estimation of Their Head Position from UAV Images*
Haruka Egawa, Seiji Ishikawa, Joo Kooi Tan (Kyushu Institute of Technology, Japan)

OS14-2 *Development of a Pedestrian Crossing Navigation System for Visually Impaired People Using MY VISION*
Kohei Kitagawa, Seiji Ishikawa, JooKooi Tan (Kyushu Institute of Technology, Japan)

OS14-3 *Abnormal Human Action Detection Based on GAN*
Tomoya Sano, Joo Kooi Tan (Kyushu Institute of Technology, Japan)

OS14-4 *3-D Position Measurement of a Cargo Using Epipolar Geometry for Logistics Automation(withdraw)*
Kazuki Fukuda, Joo Kooi Tan (Kyushu Institute of Technology, Japan)

OS14-5 *Bus Line Number Detection Employing MY VISION*
Ye Zhou, Yosiki Hamasaki, Joo Kooi Tan (Kyushu Institute of Technology, Japan)

13:00-14:30 OS5 Advanced Information Processing Applications (6)
Chair: Toru Hiraoka (University of Nagasaki, Japan)
Co-Chair: Masaharu Hirota (Okayama University of Science, Japan)

OS5-1 *A Method for Patterns of Cell-Like Images Based on Distance Transformation*
Toru Hiraoka, Kohei Maeda (University of Nagasaki, Japan)

OS5-2 *Generating Striped Animations by Inverse Line Convergence Index Filter*
Toru Hiraoka, Ryosuke Takaki (University of Nagasaki, Japan)

OS5-3 *A Method for Estimating Home Location of Foreigners in Japan Using Photograph Location*
Masaharu Hirota, Tetsuya Oda (Okayama University of Science, Japan)

OS5-4 *A Proposal of Online Map-matching Based Trajectory Compression Algorithm Using Road Networks*
Shota Iiyama, Tetsuya Oda, Masaharu Hirota (Okayama University of Science, Japan)

OS5-5 *Case Study and Direction of Share Cycle System in Japanese Cities*
Minoru Kumano (University of Miyazaki, Japan)
Toru Hiraoka (University of Nagasaki, Japan)

OS5-6 *The IoT Solution to Play English Word Learning Tool*
Shogo Aizawa, Motohide Yoshimura (University of Nagasaki, Japan)

15:30-16:15 OS11 Robotic Manipulation (3)
Chair: Kensuke Harada (Osaka University, Japan)
Co-Chair: Tokuo Tsuji (Kanazawa University, Japan)
Co-Chair: Akira Nakamura (AIST, Japan)

OS11-1 *Using Various Evaluation Standards to Determine an Error Recovery Process in an Automation Plant*
Akira Nakamura*[1], Natsuki Yamanobe*[1], Ixchel Ramirez Alpizar*[1],
Kensuke Harada*[2], Yukiyasu Domae*[1]
(*[1] National Institute of Advanced Industrial Science and Technology (AIST), Japan
*[2] Osaka University, Japan)

OS11-2 *Robotic Picking for Piled Sushi Topping*
Kenta Matsuura*[1]*[2], Keisuke Koyama*[1], Weiwei Wan*[1], Kensuke Harada*[1]
(*[1] Osaka University, Japan, *[2] Currently with Yaskawa Electric Co. Ltd., Japan)

OS11-3 *Motion Generation by Learning Relationship between Object Shapes and Human Motions*
Tokuo Tsuji[1]*, Sho Tajima*[1], Yosuke Suzuki*[1], Tetsuyou Watanabe*[1],
Shoko Miyauchi*[2], Ken'ichi Morooka*[2], Kensuke Harada*[3], and Hiroaki Seki*[1]
(*[1] Kanazawa University, Japan, *[2] Kyushu University, Japan, *[3] Osaka University, Japan)

Meeting Room 2 (ZOOM ID: 986 1953 6512)
9:00-10:15 OS16 Software Development Support Method (5)
Chair: Tetsuro Katayama (University of Miyazaki, Japan)
Co-Chair: Tomohiko Takagi (Kagawa University, Japan)

OS16-1 *Simulation and Regression Testing for Behavior of Software Models Based on Extended Place/Transition Net with Attributed Tokens*
Tomohiko Takagi, Ryo Kurozumi (Kagawa University, Japan)

OS16-2 *Development of an Early Prototype Tool for Learning Software Modeling Using Extended Place/Transition Net*
Tomohiko Takagi, Akio Usuda (Kagawa University, Japan)

OS16-3 *Test Suite Reusability Measurement Based on Frequency and Coverage of Reused Test Cases*
Mochamad Chandra Saputra*, Tetsuro Katayama*, Yoshihiro Kita†,
Hisaaki Yamaba*, Kentaro Aburada*, Naonobu Okazaki*
(*University of Miyazaki, Japan) (†University of Nagasaki, Japan)

OS16-4 *The Seven Information Features of Class for Blob and Feature Envy Smell Detection in a Class Diagram*
Bayu Priyambadha*, Tetsuro Katayama*, Yoshihiro Kita†,
Hisaaki Yamaba*, Kentaro Aburada*, Naonobu Okazaki*
(*University of Miyazaki, Japan) (†University of Nagasaki, Japan)

OS16-5 *Improvement of RETUSS to Ensure Traceability between Sequence Diagram in UML and Java Source Code in Real Time*
Kaoru Arima*, Tetsuro Katayama*, Yoshihiro Kita†,
Hisaaki Yamaba*, Kentaro Aburada*, Naonobu Okazaki*
(*University of Miyazaki, Japan) (†University of Nagasaki, Japan)

13:00-14:30 OS6 Intelligent Control (6)
Chair: Yingmin Jia (Beihang University, China)
Co-Chair: Weicun Zhang (University of Science and Technology Beijing, China)

OS6-1 *Faster R-CNN Based Defect Detection of Micro-precision Glass Insulated Terminals*
Qunpo Liu[1], Mengke Wang[1], Zonghui Liu[1], Naohiko Hanajima[2], Bo Su[1]
([1]Henan Polytechnic University, P.R.China) ([2]Muroran Institute of Technology, Japan)

OS6-2 *Adaptive Sliding Mode Control for a Constant Tension Suspension System*
Yuxin Jia, Yingmin Jia, Kai Gong (Beihang University, China)

OS6-3 *Weighted Multiple Model ADRC for Uncertain Linear System*
Weicun Zhang, Jing Ge (University of Science and Technology Beijing, China)

OS6-4 *Trajectory Tracking Control of Differential Wheeled Mobile Robot Based on Rhombic Input Constraints*
Kai Gong, Yingmin Jia, Yuxin Jia (Beihang University, China)

OS6-5 *Maneuvering target tracking based on improved interacting multiple model algorithm*
Weicun Zhang, Meiyu Zhu (University of Science and Technology Beijing, China)

OS6-6 *Encapsualted Agents of Hybrid Order Discrete Dyanmics*
Yunzhong Song (Henan Polytechnic University, P.R.China)

15:30-16:15 OS17 Robotics and machine vision (3)
Chair: Jiwu Wang (Beijing Jiaotong University, China)
Co-Chair: Junxiang Xu (Beijing Jiaotong University, China)

OS17-1 *Kinematic modeling and Simulation of humanoid dual-arm robot*
Jiwu Wang, Junxiang Xu (Beijing Jiaotong University, China)

OS17-2 *Kinematics analysis and simulation of 6R Robot Based on MATLAB/Simulink*
Jiwu Wang, Shuo Han (Beijing Jiaotong University, China)

OS17-3 *Research on planar ranging system based on binocular stereo vision*
Jiwu Wang, Xin Pei (Beijing Jiaotong University, China)

Meeting Room 3 (ZOOM ID: 984 4367 9770)
9:00-10:15 OS22 Robot Competitions and Education (5)

Chair: Yasunori Takemura (Nishinippon Institute of Technology, Japan)
Co-Chair: Kazuo Ishii (Kyushu Institute of Technology, Japan)

OS22-1 *Robustness Verification Against Noise of Self-localization Method Using Omni-directional Camera for Soccer Robot*
Yuehang Ma, Kaori Watanabe, Hidekazu Suzuki
(Tokyo Polytechnic University, Japan)

OS22-2 *Tomato-Harvesting Robot Competition: Aims and Developed Robot of 6th Competitions*
Takayuki Matsuo[1], Yasunori Takemura[2], Takashi Sonoda[2], Shinsuke Yasukawa[3], Yuya Nishida[3], Kazuo Ishii[3] ([1]National Institute of Technology, Kitakyushu College, Japan, [2]Nishinippon Institute of Technology, Japan, [3]Kyushu Institute of Technology, Japan)

OS22-3 *Smart Agriculture IoT Education Course in enPiT-everi (Education Network for Practical Information Technologies - Evolving and Empowering Regional Industries)*
Yasunori Takemura[1, 2], Keiji Kamei[1, 2], Atsushi Sanada[1, 2], Kazuo Ishii[1]
([1]Kyushu Institute of Technology, Japan, [2]Nishinippon Institute of Technology, Japan)

OS22-4 *Development of a Handy Autonomous Underwater Vehicle “Kyubic”*
Toshimune Matsumura, Yuuichiro Uemura, Kentaro Yanagise, Yoshiki Tanaka,
Yuya Nishida, Kazuo Ishii (Kyushu Institute of Technology, Japan)

OS22-5 *A Greenhouse Project toward Smart Agriculture*
Kazuo Ishii[1], Yuya Nishida[1], Shinsuke Yasukawa[1], Kanako Shirahashi[1],
Yasunori Takemura[2], Takayuki Matsuo[3]
([1]Kyushu Institute of Technology, Japan, [2]Nishinippon Institute of Technology, Japan, [3]National Institute of Technology, Kitakyushu College, Japan,)

13:00-14:15 OS7 Information Applications and Security (5)
Chair: Kuo-Hsien Hsia (National Yunlin University of Science and Technology, Taiwan)
Co-Chair: I-Hsien Liu (National Cheng Kung University, Taiwan)

OS7-1 *The Security Challenges with the Widespread Use of IT Infrastructure in ICS*
Kuan-Ming Su, I-Hsien Liu, Jung-Shian Li
(National Cheng Kung University, Taiwan)

OS7-2 *A Communication System with Equipment's Characteristics*
Chia-Chun Lai[1], I-Hsien Liu[1], Chi-Che Wu[1], Chuan-Gang Liu[2], Jung-Shian Li[1]
([1]National Cheng Kung University, Taiwan)
([2]Chia Nan University of Pharmacy and Science, Taiwan)

OS7-3 *Application of the Self-Organizing Map (SOM) to Analyze the Multiple Perspectives on Cross-National Culture*
Li-Ming Chuang, Yu-Po Lee, Shu-Tsung Chao
(Chang Jung Christian University, Taiwan)

OS7-4 Threats Hidden in Employee Workstation through Office Files
Tung-Lin Lee, I-Hsien Liu, Jung-Shian Li (National Cheng Kung University, Taiwan)

OS7-5 *Microsatellite Attitude Control Approach: Combined with Generation Adversarial Networks Fault-Detection and Cerebellar Model Articulation Controller Fault-Tolerant Control*

Ho-Nien Shou (Air Force Institute of Technology, Taiwan)

15:30-16:15 OS18 Natural Computing (3)
Chair: Marion Oswald (Vienna University of Technology, Austria)
Co-Chair: Yasuhiro Suzuki (Nagoya University., Japan)

OS18-1 *Contribution to the Theory of Periodic Reaction of Three Bodies Systems*
Yasuhiro Suzuki (Nagoya University Japan)

OS18-2 *Deep Micro Vibrotactile, DMV and its Applications*
Yasuhiro Suzuki (Nagoya Universit, Japan)

OS18-3 *Inter-Induce computation and its Philosophical Foundation*
Yasuhiro Suzuki (Nagoya University, Japan)

Meeting Room 4 (ZOOM ID: 971 5454 5491
9:00-10:00 GS6 Robotics 1 (4)
Chair: Noritaka Sato (Nagoya Institute of Technology, Japan)

GS6-1 *Digital Testing Device for Active Range of Motion of Finger Joints Utilizing Artificial Neural Network*
Huu-Hieu Quang, Yoshifumi Morita (Nagoya Institute of Technology, Japan), Makoto Takekawa (everfine, Japan)

GS6-2 *Position and Force Teaching Method for 6 DoF Manipulator Using Contact Teaching Tool and Teaching Data Editor*
Duy-Do Bui, Hiroki Tanaka, Quang-Trung Chu, Hideki Inuzuka, Yoshifumi Morita (Nagoya Institute of Technology, Japan) Masao Sakai (Aichi Prefecture, Japan)

GS6-3 *Haptic Device that Presents Sensation Corresponding to Palm on Back of Hand for Teleoperation of Robot Hand Report2: Consideration on Decided Specification*
Kyosuke Ushimaru,Noritaka Sato,Yoshifumi Morita (Nagoya Institute of Technology, Japan)

GS6-4 *Robot Assisting Water Serving to Disabilities by Voice Control*
Yang Chunxin, Sakmongkon Chumkamon, Eiji Hayashi (Kyushu Institute of Technology, Japan)

13:00-14:15 GS1 Chaos & Application (4)
Chair: Masato Nagayoshi (Niigata College of Nursing, Japan)

GS1-1 A Method of Role Differentiation Using a State Space Filter with a Waveform Changing Parameter in Multi-agent Reinforcement Learning
Masato Nagayoshi, Simon Elderton (Niigata College of Nursing, Japan)
Hisashi Tamaki (Kobe University, Japan)

GS1-2 *The research about editing system of performance information for player piano. - Make inferences about both handed musical composition by using DP matching system -*

Ryo Kinoshita, Eiji Hayashi (Kyushu Institute of Technology, Japan)

GS1-3 *Development of a LiDAR based Navigation System for Tree Harvester*
Ayumu Tominaga, Ryusuke Fujisawa, Eiji Hayashi (Kyushu Institute of Technology, Japan), Abbe Mowshowitz (The City College of New York, US)

GS1-4 *Numerical study on a class of chaotic financial chaotic systems(withdraw)*
Lei Gong, Minghan Song (Tianjin University of Science and Technology, China)

GS1-5 *Dynamic characteristics analysis of a multi-scroll conservative chaotic system with sinusoidal nonlinearity*
Zhonggao Chen (Tianjin University of Science and Technology, China)

15:30-17:00 GS8 Medical Data Processing (6)
Chair: Shingo Mabu (Yamaguchi University, Japan)

GS8-1 *Development of in-Home Wireless Continuous Temperature Data Logging and Alarming System for Fever Monitoring in Pediatrics*
Ali S. AlMejrad(University of Hail, Kingdom of Saudi Arabia)

GS8-2 *Design and Development of in-home Wireless Crucial Events Logging and Alarming System for Elderly and Disabled People Care*
Ali S. AlMejrad(University of Hail, Kingdom of Saudi Arabia)

GS8-3 *Virtual bird's-eye view for remote operation of unmanned construction machinery*
Noritaka Sato, Akihiro Fukuda (Nagoya Institute of Technology, Japan)

GS8-4 *Domain Transformation of Chest CT Images Using Semi-Supervised Cycle GAN for Opacity Classification of Diffuse Lung Diseases*
Masashi Miyake[1], Shingo Mabu[1] , Shoji Kido[2], Takashi Kuremoto[1]
([1]Yamaguchi University, Japan, [2]Osaka University, Japan)

GS8-5 *Anomaly Detection of Lung Sounds Using DAGMM*
Ryosuke Wakamoto[1], Shingo Mabu[1], Shoji Kido[2], Takashi Kuremoto[1]
([1]Yamaguchi University, Japan, [2]Osaka University, Japan)

GS8-6 *A ground reaction force analysis in walking and running gaits in horse leg model on viscoelastic hoof-ground contact*
Dondogjamts Batbaatar, Hiroaki Wagatsuma (Kyushu Institute of Technology, Japan)

January 24 (Sunday)

8:40-Registration

Meeting Room 1 (ZOOM ID: 993 2436 1751)
9:00-9:45 OS4 Intelligent UAV and Systems (3)
Chair: Young Im Cho (Gachon University, Republic of Korea)

Co-Chair: Jang-Myung Lee (Pusan National University, Republic of Korea)

OS4-1 *Satellite Image-based UAV Localization using Siamese Neural Network*
Seong-Ha Ahn, Ho-Sun Kang, Jang-Myung Lee
(Pusan National University, Republic of Korea)

OS4-2 *Analysis Based on CNN for Automated Vehicle Parking Occupancy*
Muksimova Shakhnoza, Young Im Cho (Gachon University, Republic of Korea)

OS4-3 *On the boundary layer part of the asymptotics of the solution of a singularly perturbed boundary value problem*
Assiya Zhumanazarova, Young Im Cho (Gachon University, Republic of Korea)

11:00-12:00 OS3-1 Robot Control (4)
Chair: Yizhun Peng (Tianjin University of Science and Technology, China)
Co-Chair: Huailin Zhao (Shanghai Institute of Technology, China)

OS3-1 *A Design and Implementation Intelligent Flowerpot*
Songyun Shi, Yizhun Peng (Tianjin University of Science and Technology, China)

OS3-2 *Research of Attention-LSTM Model for Baby Cry Detection Robot*
Tianye Jian, Yizhun Peng, Wanlong Peng, Zhou Yang
(Tianjin University of Science and Technology, China)

OS3-3 *A Design and Implement of an Automatic Intelligent Car*
Chengzhi Liu, Yizhun Peng, Jikai Zhao, Dezhi Yang
(Tianjin University of Science and Technology, China)

OS3-4 *A Target Detection in Remote Sensing Image based on Deep Learning*
Lianchen Zhao, Yizhun Peng, Di Li, Yuheng Zhang
(Tianjin University of Science and Technology, China)

13:00-14:15 OS3-2 Robot Control (5)
Chair: Yizhun Peng (Tianjin University of Science and Technology, China)
Co-Chair: Huailin Zhao (Shanghai Institute of Technology, China)

OS3-5 *A Design of Home Smart Nursing Robot based on Raspberry Pi 4*
Yuqi Zhao, Yizhun Peng, Jie Liu, Xiaowei Wu, Jikai Zhao
(Tianjin University of Science and Technology, China)

OS3-6 *Design and Simulation of Indoor Tour Guide Robot Based on ROS*
Yuheng Zhang, Yizhun Peng, Lianchen Zhao
(Tianjin University of Science and Technology, China

OS3-7 *Research on Path Planning of Manipulator*

Nana Wang, Yizhun Peng, Zhou Yang, Yuheng Zhang
(Tianjin University of Science and Technology, China)

OS3-8 *Research on White-Line-tracking Walking Technique of NAO Robot*
Wanlong peng, Yizhun peng, Yuheng zhang,Tianye jian
(Tianjin University of Science and Technology, China)

OS3-9 *A Design and Implement of Portable Epidemic Detection Device based on STM32*
Yuqi Zhao, Yizhun Peng, Xiaowei Wu, Yusong Zhou
(Tianjin University of Science and Technology, China)

Meeting Room 2 (ZOOM ID: 986 1953 6512)
9:00-9:45 GS4 Image & Signal Processing 3 (3)
Chair: Taro Shibanoki (Ibaraki University, Japan)

GS4-1 *Landslide Area Detection from Synthetic Aperture Radar Images Using Convolutional Adversarial Autoencoder and One-class SVM*
Shingo Mabu, Soichiro Hirata, Takashi Kuremoto (Yamaguchi University, Japan)

GS4-2 *Anomaly Detection in Time Series Data Using Support Vector Machines*
Umaporn Yokkampon[1], Sakmongkon Chumkamon[1], Abbe Mowshowitz[2],
Eiji Hayashi[1] and Ryusuke Fujisawa[1] ([1]Kyushu Institute of Technology, Japan)
([2]The City College of New York, USA)

GS4-3 *Relationship Between Tactile Stimuli and Human Body Sway*
Masaya Tadokoro, Taro Shibanoki (Ibaraki University, Japan)

11:00-12:00 OS13 Intelligent Systems and Robotics (14)
Chair: Fengzhi Dai (Tianjin University of Science and Technology, China)
Co-Chair: Yunzhong Song (Henan Polytechnic University, China)

OS13-1 *Analysis of the Consensus Protocol of Heterogeneous Agents with Time-Delays*
Jichao Zhao[1], Fengzhi Dai[1,3], Yunzhong Song[2]
([1]College of ElTianjin University of Science and Technology, China, [2]Henan Polytechnic University, China, [3]Tianjin Tianke Intelligent and Manufacture Technology CO., LTD, China)

OS13-2 *Burrows-Wheeler Transform Acceleration based on CUDA*
Chang Sheng, Fengzhi Dai (Tianjin University of Science and Technology, China)

OS13-3 *Design of Automatic Water Supply Upper Computer System*
Peng Lu, Fengzhi Dai, Tianyi Zhang
(Tianjin University of Science and Technology, China)

OS13-4 *Analysis of Boiler Water Level System based on the Fuzzy Control*
Tianyi Zhang, Fengzhi Dai, Peng Lu
(Tianjin University of Science and Technology, China)

OS13-5 *Development of the Circuit System for Greenhouse Environment Regulation*
Yuhui Cheng, Fengzhi Dai, Chengxu Ji, Peng Lu

(Tianjin University of Science and Technology, China)

OS13-6 *Design of an Intelligent Car based on MSP430*
Ruming Kang, Fengzhi Dai (Tianjin University of Science and Technology, China)

OS13-7 *Design of a WIFI Video Car*
Qianqian Zhang[1], Fengzhi Dai[1,2], Jichao Zhao[1], Haokang Wen[1], Hongbo Hao[1]
([1]Tianjin University of Science and Technology, China)
([2]Tianjin Tianke Intelligent and Manufacture Technology CO., LTD, China)

OS13-8 *Design of WebGIS Transportation and Distribution System based on the Genetic Algorithm*
Hongbo Hao, Fengzhi Dai (Tianjin University of Science and Technology, China)

OS13-9 *A Study of YOLO Algorithm for Target Detection*
Haokang Wen, Fengzhi Dai, Yasheng Yuan
(Tianjin University of Science and Technology, China)

OS13-10 *Research on Recognition and Application of EEG Signal based on SSVEP-BCI*
Di Yin[1], Fengzhi Dai[1,3], Mengqi Yin[2], Yasheng Yuan[1], Yuxuan Zhu[1]
([1]Tianjin University of Science and Technology, China)
([2]Hebei University of Chinese Medicine, China)
([3]Tianjin Tianke Intelligent and Manufacture Technology CO., LTD, China)

OS13-12 *Research on Crop Image Recognition Technology - Take Daylily as an Example*
Jichao Zhao, Fengzhi Dai (Tianjin University of Science and Technology, China)

OS13-13 *Design of Daylily Agricultural Picking Robot*
Jichao Zhao, Fengzhi Dai (Tianjin University of Science and Technology, China)

OS13-14 *Development of Intelligent Public Trash Can based on Machine Vision and Learning*
Longyu Gao, Fengzhi Dai, Zhiqing Xiao, Jiangyu Wu, Zilong Liu
(Tianjin University of Science and Technology, China)

13:00-14:15 OS12 System and Control (18)
Chair: Huailin Zhao (Shanghai Institute of Technology, China)
Co-Chair: Jichao Zhao (Tianjin University of Science and Technology, China)

OS12-1 *The Optimized Intelligent Algorithms on Face Recognition and Tracking for ROS-based Robots*
Yue Chen, Shuhao Tian, Huailin Zhao, Shengyang Lu
(Shanghai Institute of Technology)

OS12-2 *Control of a Novel 5D Hyperchaotic System*
Qiang Wei[1], Hong Niu[2]
([1]Army Military Transportation University, China)
([2]Tianjin University of Science & echnology, China)

OS12-3 *Design of Multifunctional Vehicle Interior Environment Monitoring System Based on*

Wireless Communication
Yuqi Yan, Jialin Yang, Zhongxu Qin (Wuhan University of Technology, China)

OS12-4 *Robot Structure and Motion control Design Based on UG and Product*
Yuhang Sheng (Tianjin University of Science and Technology, China)

OS12-5 *Design of Brushed Motor Position Loop Control System Based on Incremental PID*
Tianyi Zhang, Peng Lu (Tianjin University of Science and Technology, China)

OS12-6 *Design of a Fire alarm system*
Yuhui Cheng, Tianyi Zhang (Tianjin University of Science and Technology, China)

OS12-7 *Design of intelligent curtain control circuit based on single chip Microcomputer*
Qianqian Zhang, Jichao Zhao, Haokang Wen, Hongbo Hao
(Tianjin University of Science and Technology, China)

OS12-9 *A Survey of Low Visibility Image Enhancement Based on MSRCR Algorithm*
Haokang Wen, Hongbo Hao (Tianjin University of Science and Technology, China)

OS12-10 *Design of a Taxi Meter Based on Single Chip Computer*
Haokang Wen, Min Wang (Tianjin University of Science and Technology, China)

OS12-11 *Design of a High Precision Digital Clock Based on Single Chip Microcomputer*
Haokang Wen, Qiang Zheng (Tianjin University of Science and Technology, China)

OS12-12 *Design of Manchu Image Acquisition System Based on STM32*
Hongbo Hao[1], Fulin Zheng[2]
([1]Tianjin University of Science and Technology, China)
([2] Dalian Minzu University, China)

OS12-13 *Design of timing socket based on single chip microcomputer*
Hongbo Hao[1], Shuailin Chen[2]
([1]Tianjin University of Science and Technology, China)
([2]Qingdao University of Science and Technology, China)

OS12-14 *Study on the effect of physical fitness training on children's cognitive ability*
Jianhua Deng, Lei Ning (Yanshan University, China)

OS12-15 *Effects of high heels on plantar stress in women*
Jianhua Deng, Lei Ning (Yanshan University, China)

OS12-16 *A research on Intelligent Classification of Urban Trash Bins based on Machine Learning*
Longyu Gao, Zilong Liu, Luqi Shen, Songyun Shi, Yongzheng Lv
(Tianjin University of Science and Technology, China)

OS12-17 *A Design on Intelligent Public Trash Can based on Machine Vision and Auxiliary Sensors*
Longyu Gao, Leixin Han, Jiangyu Wu, Mingfei Liu, Ruming Kang
(Tianjin University of Science and Technology, China)

OS12-18 *A research on Front-end Garbage Classification based on Machine Vision*
Longyu Gao, Zhiqing Xiao, Junlong Hao, Luqi Shen, Manqian Hu
(Tianjin University of Science and Technology, Tianjin, China)

Meeting Room 3 (ZOOM ID: 984 4367 9770)
11:00-12:00 OS20 Virtual Reality and Intelligent Interactions (3)
Chair: R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)
Co-Chair: Yoshimasa Tokuyama (Tokyo Polytechnic University, Japan)

OS20-1 *NeuroPhyllotaxis: An Interactive Application for Generative Art Based on EEG Data*
Chien-Tung Lin, R.P.C. Janaka Rajapakse
(Tainan National University of the Arts, Taiwan)
Yoshimasa Tokuyama (Tokyo Polytechnic University, Japan)

OS20-2 *VRMAZU: VR Visualization of Mazu Temple for Passive Interaction with Generated Sound from the ML Technique(withdraw)*
Yi-Li Liang. R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)
Jen-Tun Lee (Japan Advanced Institute of Science and Technology, Japan)
Yoshimasa Tokuyama (Tokyo Polytechnic University, Japan)

OS20-3 *ThoughtMix: Interactive Water Color Generation and Mixing Based on EEG Data*
R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)
Yoshimasa Tokuyama (Tokyo Polytechnic University, Japan)

OS20-4 *HaptWarp: Soft Printable and Motion Sensible Game Controller*
Jen-Tun Lee, Kazunori Miyata
(Japan Advanced Institute of Science and Technology, Japan)
R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)

13:00-14:00 OS1 Human-Machine Interface Application (4)
Chair: Norrima Mokhtar (University of Malaya, Malaysia)
Co-Chair: Heshalini Rajagopal (Manipal International University, Malaysia)

OS1-1 *Gray Level Co-Occurrence Matrix (GLCM) and Gabor Features Based No-Reference Image Quality Assessment for Wood Images*
Heshalini Rajagopal[1], Norrima Mokhtar[1], Anis Salwa Mohd Khairuddin[1], Wan Khairunizam[2], Zuwairie Ibrahim[3], Asrul Bin Adam[3], Wan Amirul Bin Wan Mohd Mahiyidin[1]
([1]University of Malaya, Malaysia) ([2]University of Malaysia Perlis, Malaysia) ([3]University of Malaysia Pahang, Malaysia)

OS1-2 *Design and Development of Automated Seeding and Irrigation System*
Anirban Kumar, Heshalini Rajagopal (Manipal International University, Malaysia)

OS1-3 *Investigation of A Real-Time Driver Eye-Closeness for the Application of Drowsiness Detection*
Muhammad Zubir bin Kamazlan[1], Wan Khairunizam[1], Abdul Hafiz Halin[1], M. Rudzuan M. Nor[1], Azian Azamimi Abdullah[1], Norrima Mokhtar[2]
([1]University Malaysia Perlis, Malaysia) ([2]University of Malaya, Malaysia)

OS1-4 *Towards Establishing Path Planning Strategies For Autonomous UAVs; A Brief Survey-Summary on Recent Technique*
Anees ul Husnain, Norrima Binti Mokhtar, Noraisyah Binti Mohamed Shah, Mahidzal Bin Dahari (University of Malaya, Malaysia)

Meeting Room 4 (ZOOM ID: 971 5454 5491)
11:00-12:00 GS7 Robotics 2 (4)
Chair: Wisanu Jitviriya

GS7-1 *Development of Interactive Robot – Emotion Estimation System Using Speech by 1dCNN -*
Yugo Kawachi, Eiji Hayashi (Kyushu Institute of Technology, Japan)

GS7-2 *Deep Learning Methods for Semantic Segmentation of Dense 3D SLAM Maps*
Pei Yingjian, Sakmongkon Chumkamon, Eiji Hayashi
(Kyushu Institute of Technology, Japan)

GS7-3 *Robot Motion Generation by Hand Demonstration*
Sakmongkon Chumkamon, Umaporn Yokkampon, Eiji Hayashi, Ryusuke Fujisawa
(Kyushu Institute of Technology, Japan)

GS7-4 *Medical Telerobotics: IRAPs SHaRE-aGIVeR*
Noppadol Pudchuen, Jiraphan Inthiam, Wisanu Jitviriya, Amornphun Phunopas, Chirdpong Deelertpaiboon
(King Mongkut's University of Technology North Bangkok, Thailand)
Aran Blattler (Kyushu Institute of Technology, Japan)

Group Meeting

Abstracts
PS abstracts

PS1 Tomato-harvesting-robot competition towards smart agriculture

Kazuo Ishii (Kyushu Institute of Technology, Japan)

In agriculture, the aging and depopulation of farmers cause the shortages of farmers and manpower. Most of commercialized robots are industrial robots for factory automation, and most of robots for the first industry, agriculture, forestry and fisheries are still under developing. The reasons are cost-efficiency of the robotization, safety of the works using robots, difficulty of outdoor operations, and knowledge transfer problem from farmers to computer, etc. As one of solutions for the problems, robot technology into the agriculture is expected to contribute to the laborsaving, improvement of production, production line automation, and also the management toward smart-agriculture. We organize the Tomato-harvesting-robot competition to offer the research field and welcome researchers into agricultural robotics.

PS2 Human-computer Communication Using Facial Expression

Yasunari Yoshitomi (Kyoto Prefectural University, Japan)

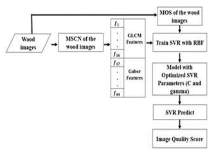
To develop a computer system, such as a robot, that can communicate smoothly with human, it is necessary to equip the system with a function of understanding human emotion. Moreover, it is also necessary to equip the system with a function of expressing emotional signal to human. From both points of view, facial expression is a promising target as a research field. With co-researchers, I have been researching both aspects on facial expressions. For developing a method for recognizing facial expression, we have used infrared-ray images in addition to visible-ray images. For expressing emotional signal to human, we have developed a personified agent. Our challenges for human-computer communication using facial expression will be introduced in my presentation.

OS abstracts
OS1 Human-Machine Interface Application (4)
Chair: Norrima Mokhtar (University of Malaya, Malaysia)
Co-Chair: Heshalini Rajagopal (Manipal International University, Malaysia)

OS1-1 Gray Level Co-Occurrence Matrix (GLCM) and Gabor Features Based No-Reference Image Quality Assessment for Wood Images

Heshalini Rajagopal[1], Norrima Mokhtar[1], Anis Salwa Mohd Khairuddin[1], Wan Khairunizam[2], Zuwairie Ibrahim[3], Asrul Bin Adam[3], Wan Amirul Bin Wan Mohd Mahiyidin[1]
([1]University of Malaya, Malaysia) ([2]University of Malaysia Perlis, Malaysia)
([3]University of Malaysia Pahang, Malaysia)

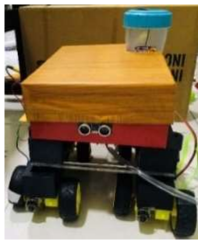
Image Quality Assessment (IQA) is an imperative element in improving the effectiveness of an automatic wood recognition system. There is a need to develop a No-Reference-IQA (NR-IQA) system as a distortion free wood images are impossible to be acquired in the dusty environment in timber factories. Therefore, a Gray Level Co-Occurrence Matrix (GLCM) and Gabor features-based NR-IQA, GGNR-IQA algorithm is proposed to evaluate the quality of wood images. The proposed GGNR-IQA algorithm is compared with a well-known NR-IQA, Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) and Full-Reference-IQA (FR-IQA) algorithms, Structural Similarity Index (SSIM), Multiscale SSIM (MS-SSIM), Feature SIMilarity (FSIM), Information Weighted SSIM (IW-SSIM) and Gradient Magnitude Similarity Deviation (GMSD). Results shows that the GGNR-IQA algorithm outperforms the NR-IQA and FR-IQAs. The GGNR-IQA algorithm is beneficial in wood industry as a distortion free reference image is not required to pre-process wood images.



OS1-2 Design and Development of Automated Seeding and Irrigation System

Anirban Kumar, Heshalini Rajagopal (Manipal International University, Malaysia)

Malaysia, has a very conducive environment for agriculture. Six percent of the population are employed to some of the agriculture activities making agriculture the most curtail industry in Malaysia. Hence, there is a need to develop the agricultural facilities by incorporating latest technological advancements. Conventional seeding process is time consuming and requires additional labor cost. In this project, an automated system is proposed for seeding as well as irrigation process in agriculture which reduces the labor cost. This system aims to increase the efficiency of the seeding process without affecting the nature of soil. The proposed system is equipped with Arduino MEGA and Arduino UNO which acts as the main control unit while ultrasonic and soil moisture sensors are used to detect the obstacles and soil moisture level, respectively. The robot consists of a funnel like arrangement to perform the seeding procedure. The field is equipped with moisture sensors placed at different areas that monitors the moisture level of the soil on a regular interval for irrigation purposes. The proposed system will be of great benefit to the future endeavor of agricultural business as well as it will be able to optimize the seeding and irrigation.



OS1-3 Investigation of A Real-Time Driver Eye-Closeness for the Application of Drowsiness Detection

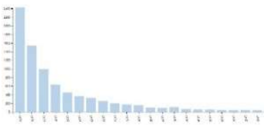
Muhammad Zubir bin Kamazlan[1], Wan Khairunizam[1], Abdul Hafiz Halin[1], M. Rudzuan M. Nor[1], Azian Azamimi Abdullah[1], Norrima Mokhtar[2]
([1]University Malaysia Perlis, Malaysia) ([2]University of Malaya, Malaysia)

The increase in accident and death rates due to drowsiness while driving raises concerns to the community. An efficient solution is vital to ensure the safety of all drivers on the road. Most previous studies have analyzed drowsiness using head tilt, yawning, and eye condition. Face detection applied in drowsiness detection from previous research not included distances between subject and camera. The features used for eye detection required large storage and long-term process which are not applicable in a real-time system. This study uses Haar algorithm and analysis is performed based on the size of the region of interest for face detection. Eye monitoring uses facial landmark features and the evaluation is dependent on the width of the eye. The percentage of eye closure is used to describe the eyes as closed. This study only takes into account the normal rate of blinking eyes while driving because of the long-time constraints required for a person to be in a drowsy state. In this research, the Raspberry Pi 3B+ and Pi cameras are used as processing and vision devices. The highest accuracy of face detection achieved based on the ROI area at a distance of 80 cm is 98.33%. The lowest difference between eye width and the intercanthal distance is 0.36%. The overall normal eye blink rate while driving is in the range of the normal eye blink rate which does not exceed 20 blinks/min as reported by the previous researcher.

OS1-4 Towards Establishing Path Planning Strategies For Autonomous UAVs; A Brief Survey-Summary on Recent Technique

Anees ul Husnain, Norrima Binti Mokhtar, Noraisyah Binti Mohamed Shah, Mahidzal Bin Dahari
(University of Malaya, Malaysia)

The extent of autonomy in path planning for a UAV primarily depends upon the capabilities of its algorithm. The diversity in UAV applications and an abundance of choices in autonomous path planning algorithms are swelling every day, so the selection of most appropriate algorithm gets baffling. The past two decades of research on UAVs revealed that seventy percent of it had been published in the previous three and a half years. Hence, a comprehensive survey study was proposed and conducted to obtain an overview of the recent developments in autonomous path planning applications and their respective algorithms. This article presents a summary of the survey and suggests most suitable path planning algorithms for a UAV application.



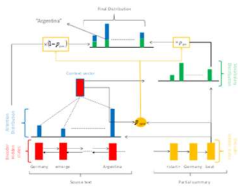
OS2 Media Information Processing, Music Recommendation and Artificial Intelligence (4)

Chair: Yasunari Yoshitomi (Kyoto Prefectural University, Japan)
Co-Chair: Masayoshi Tabuse (Kyoto Prefectural University, Japan)

OS2-1 Comparison of Data Augmentation Methods in Pointer-Generator Model Using Various Sentence Ranking Methods

Tomohito Ouchi, Masayoshi Tabuse (Kyoto Prefectural University, Japan)

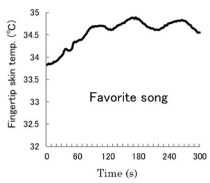
In the existing research, we proposed a data augmentation method using topic model for Pointer-Generator model. In this study, we add to the sentence ranking method in the data augmentation method. Specifically, we add two ranking method using LexRank and Luhn. LexRank is based on Google's search method and Luhn defines sentence features and ranks sentences. We compare three data augmentation method. We considered which method is suitable for data augmentation. Nowadays, it is very difficult to extract information due to the flood of information on the Internet. Therefore, we propose Pointer-Generator model using a data augmentation method using various ranking methods.



OS2-2 Music Recommendation System Driven by Variations in Fingertip Skin Temperature

Mayuka Wada[1], Taro Asada[2], Yasunari Yoshitomi[2], Masayoshi Tabuse[2]
([1]Shimadzu System Development Corp., Japan, [2]Kyoto Prefectural University, Japan)

In recent years, music therapy has been performed to recover cognitive function of elderly people. In the previously reported system, it is necessary to input the user's subjective evaluation to determine the next recommended song. In this study, variations in fingertip skin temperature were used as input instead of subjective evaluation, focusing on evaluation of emotion expressed as physiological response through variations in fingertip skin temperature. Our system can be used with no support for users, such as severe dementia patients or bedridden patients, who cannot respond actively.



OS2-3 Music Recommendation System Driven by Interaction between User and Personified Agent Using Speech Recognition, Synthesized Voice and Facial Expression

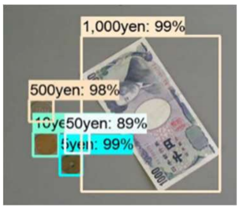
Ayumi Matsui[1], Miki Sakurai[2], Taro Asada[3], Yasunari Yoshitomi[3], Masayoshi Tabuse[3]
([1]Sumitomo Mitsui Card Co., Ltd., Japan, [2]TIS Inc., Japan, [3]Kyoto Prefectural University, Japan)

We propose a system for music recommendation through an interaction between a user and a personified agent using speech recognition, synthesized voice and facial expression. The speech is recognized using a speech recognition system called as Julius, followed by facial expression synthesis of the agent using preset parameters depending on the vowel each. We used MikuMikuDanceAgent (MMDAgent) to create an agent. To produce the agent's voice, we use the speech synthesis function setting built into MMD Agent. We add the agent a new function of changing its facial expression according to user's response to music recommended by our system. The effectiveness of the proposed system is verified.

OS2-4 Wallet Operation Evaluation System Using Deep Learning

Junichiro Yamawaki[1], Yasunari Yoshitomi[2], Masayoshi Tabuse[2], Taro Asada[2]
([1]SKY Co., Ltd., Japan, [2]Kyoto Prefectural University, Japan)

In Japan, the average age of the population has been increasing, and this trend is expected to continue. Because of this trend, the number of older people with dementia is increasing very rapidly. It is necessary to find out persons at the early stage, so called MCI (Mild Cognitive Impairment), of dementia, for taking care of them while suppressing their dementia progression. For investigating MCI, wallet operation test has been receiving considerable attentions. In the present study, we propose a system for wallet operation evaluation using deep learning. By the system, a wallet, bills, and coins are automatically recognized and the ability of correctly picking up bills and coins in the wallet within the reasonable time is estimated.



OS3 Robot Control (9)

Chair: Yizhun Peng (Tianjin University of Science and Technology, China)
Co-Chair: Huailin Zhao (Shanghai Institute of Technology, China)

OS3-1 A Design and Implementation Intelligent Flowerpot

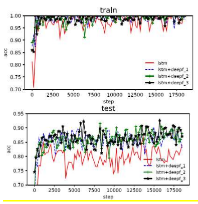
Songyun Shi, Yizhun Peng (Tianjin University of Science and Technology, China)

Aiming at the characteristics that people often neglect the care of potted plants at home, this article designs a smart flowerpot, which can make the potted plants survive and grow better without supervision. This is a smart home product based on the Internet of Things technology. The flowerpot collects data from temperature sensors, humidity sensors, soil humidity sensors, harmful gas sensors, and photosensitive modules through STM32 single-chip microcomputers, and cooperates with smart tracking systems and automatic irrigation systems. To achieve the purpose of unmanned cultivation of potted plants, beautify and improve the living environment.

OS3-2 Research of Attention-LSTM Model for Baby Cry Detection Robot

Tianye Jian, Yizhun Peng, Wanlong Peng, Zhou Yang
(Tianjin University of Science and Technology, China)

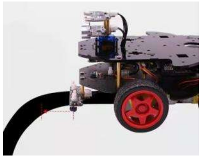
In order to achieve the effective acquisition of frame-level speech features under different emotional needs of baby, a speech emotion recognition model for baby based on improved long-term and short-term memory (LSTM) network is established. The frame-level speech features are used instead of the traditional statistical features to preserve the temporal relationships in the original speech, and the traditional forgetting and input gates are transformed into attention gates by introducing an attention mechanism, in order to improve the performance of speech emotion recognition, the depth attention gate is calculated according to the self-defined depth strategy. The results show that, in Fau Aibo Children's emotional data corpus and baby crying emotional needs database, compared with the traditional LSTM based model, the recall rate and F1 score of this model are 3.14%, 5.50%, 1.84% and 5.49% higher, respectively, compared with the traditional model based on Lstm and GRU, the training time is shorter and the speech emotion recognition rate of baby is higher.



OS3-3 A Design and Implement of an Automatic Intelligent Car

Chengzhi Liu, Yizhun Peng, Jikai Zhao, Dezhi Yang (Tianjin University of Science and Technology, China)

In order to study the situation that some unmanned vehicles need to drive repeatedly on fixed lines, this article introduces an autonomous smart car based on the STM32 platform. Infrared detectors are used to determine whether the vehicle is driving on a prescribed route. Ultrasonic detectors can prevent the vehicle from colliding with other people or vehicles. Inexpensive trolleys can help express companies to transport packages and letters to reduce labor costs. Large vehicles can be used to carry large cargo such as containers in docks and other places.



OS3-4 A Target Detection in Remote Sensing Image based on Deep Learning

Lianchen Zhao, Yizhun Peng, Di Li, Yuheng Zhang (Tianjin University of Science and Technology, China)

For high-resolution optical remote sensing images, there are still many challenges in target detection. In this paper, deep learning algorithm is used to detect the target in remote sensing image. Improve and optimize the deep learning target detection algorithm. When the selected data set is used for target detection, the AP value is improved, which leads to the concept of multi-scale feature fusion feature pyramid and residual network. By improving the selected Yolov3 network model, the detection effect of the two targets of aircraft and ships in remote sensing images has been significantly improved.

OS3-5 A Design of Home Smart Nursing Robot based on Raspberry Pi 4

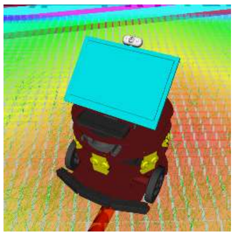
Yuqi Zhao, Yizhun Peng, Jie Liu, Xiaowei Wu, Jikai Zhao
(Tianjin University of Science and Technology, China)

This product is based on Bluetooth, WIFI and other wireless technologies, with the Raspberry Pi 4 as the core, to meet the health, safety and entertainment needs of the serviced users, mainly the elderly, while assisting the leisure and entertainment of users when traveling outdoors. A multi-functional escort robot designed for safety and other issues. It has many methods such as temperature and humidity monitoring, harmful gas monitoring, and noise monitoring to ensure the safety of users. At the same time, it also has TV projection, audio-visual entertainment touch screen interaction, mobile phone projection and other entertainment functions.

OS3-6 Design and Simulation of Indoor Tour Guide Robot Based on ROS

Yuheng Zhang, Yizhun Peng, Lianchen Zhao (Tianjin University of Science and Technology, China

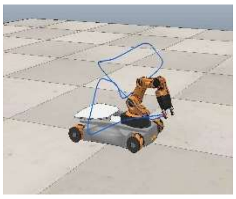
Aiming at the problems of the small number of lecturers in museums and other venues and the large demand for audience consultation, a ros-based indoor guide robot was designed. The robot consists of a mechanical system, a motion control system, and a sensor system. In order to improve the efficiency of development and debugging, and reduce the cost of experiments, it is necessary to test related algorithms in a virtual simulation environment before the robot enters the actual working state. Experiments have proved that the robot can autonomously guide guests to the destination and explain according to a preset path; the robot has multiple sensors to sense obstacles, and can autonomously avoid obstacles during the explanation and continue to move; the robot can accurately and efficiently recognize faces and provide accurate services Fast.



OS3-7 Research on Path Planning of Manipulator

Nana Wang, Yizhun Peng, Zhou Yang, Yuheng Zhang (Tianjin University of Science and Technology, China)

In order to further realize the flexibility, intelligence and human-machine friendliness, the robot must have the ability of flexible motion planning. This paper mainly introduces the general steps of manipulator path planning and common path planning algorithms. First, the path planning of the manipulator is reviewed from three aspects: environment modeling, path search, and path smoothing. In addition, compared various manipulator path planning algorithms, such as RRT algorithm, ant colony algorithm, genetic algorithm, etc. Finally, the development trend of path planning technology is summarized and analyzed.



OS3-8 Research on White-Line-tracking Walking Technique of NAO Robot

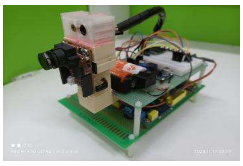
Wanlong peng, Yizhun peng, Yuheng zhang,Tianye jian
(Tianjin University of Science and Technology, China)

Liner-tracking walking technique is an essential part of intelligent robot technology. This paper selected NAO robot as the research platform to study the line-tracking walking problem. Firstly, NAO white shell increased the difficulty of white line recognition, this is different from other robot. A method based on image preprocessing is presented, sorting white edge with cut-point which is non-differential or derivative is zero, then designed the controller based on the improved digital incremental PID(proportion integration differentiation)algorithm and determined the controller parameters through the concise test method on Simulink. Finally, it was realized by programming in Python language.

OS3-9 A Design and Implement of Portable Epidemic Detection Device based on STM32

Yuqi Zhao, Yizhun Peng, Xiaowei Wu,Yusong Zhou (Tianjin University of Science and Technology, China)

The equipment is mainly designed for epidemic detection. It is an intelligent detection box device. It collects and analyzes data through cameras and sensors. It uses LBP algorithm and LAB color model to enable the detection box to identify the temperature and mask of the entered personnel. Wearing conditions, preliminary comparison of epidemic prevention standards for passers-by, whether to wear masks, whether the body temperature exceeds the standard, and at the same time enter the identity of new personnel and store them in the database, and detect the incoming personnel through face recognition to find out in time Those who are not entered shall take corresponding measures in time to improve the efficiency of epidemic prevention and control. The device does not need to be held in hand, can be fixed anywhere, can effectively avoid cross-infection, and supports self-set temperature alarm thresholds, support voice broadcast reminders, fast and accurate non-contact temperature measurement.



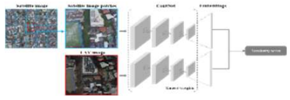
OS4 Intelligent UAV and Systems (3)

Chair: Young Im Cho (Gachon University, Republic of Korea)
Co-Chair: Jang-Myung Lee ((Pusan National University, Republic of Korea))

OS4-1 Satellite Image-based UAV Localization using Siamese Neural Network

Seong-Ha Ahn, Ho-Sun Kang, Jang-Myung Lee (Pusan National University, Republic of Korea)

We present a method for UAV localization using pre-existing satellite images. The use of Unmanned Aerial Vehicles (UAVs) has rapidly increased in several applications such as surveillance, search, and defense. When in GPS-denied situations, however, the onboard GPS signal may be noisy or inaccurate. The proposed method is based on a Siamese Neural Network that contains two instances of the same neural architecture and weights. Siamese Neural Network learns the similarity metric so that can recognize the same place from two raw images. Convolutional Neural Network is used as a backbone in Siamese Neural Network to overcome variation due to differences such as perspective, shadow angle, and presence of vehicles. We describe UAV localization pipeline and a dataset for training and testing our networks. Finally, the performance of the proposed method was shown in accuracy.



OS4-2 Analysis Based on CNN for Automated Vehicle Parking Occupancy

Muksimova Shakhnoza, Young Im Cho (Gachon University, Republic of Korea)

The noteworthy growth of the automotive trade besides an absence of urban development has caused issues like traffic congestion, air contamination, and driving difficulties. To facilitate consumers, most car manufacturers provide a pre-installed or aftermarket navigation system in vehicles. This enables drivers to easily navigate to their destination. However, navigation systems provide a rough estimate of the remaining distance to the destination. With the recent technological advancements in sensing and intelligent technologies, companies are interested to find out whether these advancements can assist in reducing the time spent to search for vacant parking spaces. Locating an unoccupied parking space is a major predicament that drivers face since searching for it is a tedious process. We investigations showed that, in comparison with previous approaches, for the task of classifying given parking spaces as vacant or occupied, the proposed approach is more robust, stable, and well-generalized for unseen images captured from completely different camera viewpoints, which has strong indications that it would generalize effectively to other parking lots.

	Accuracy	Accuracy
Our Method	**97.92%**	**97.87%**
AlexNet[6]	96.99%	97.91%
ResNet[55]	96.51%	97.80%

Table 1. Comparison of training and validation scores for AlexNet, ResNet and Our method on UFPR04.

OS4-3 On the boundary layer part of the asymptotics of the solution of a singularly perturbed boundary value problem

Assiya Zhumanazarova, Young Im Cho (Gachon University, Republic of Korea)

In this paper, we study the boundary layer part of the asymptotics of a singularly perturbed integro-differential boundary value problem with an initial jump. A part of the boundary layer is defined as an expansion in powers of a small parameter, and the coefficients of this expansion are solutions of a third-order differential problem. The purpose of this work is to describe in detail the properties of the boundary layer part of the asymptotics and to prove estimates of their coefficients.

OS5 Advanced Information Processing Applications (6)
Chair: Toru Hiraoka (University of Nagasaki, Japan)
Co-Chair: Masaharu Hirota (Okayama University of Science, Japan)

OS5-1 A Method for Patterns of Cell-Like Images Based on Distance Transformation

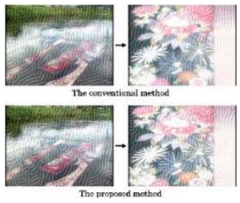
Toru Hiraoka, Kohei Maeda (University of Nagasaki, Japan)

A non-photorealistic rendering method has been proposed to generate cell-like images that cell-like patterns are represented in photographic images. The cell-like patterns are automatically generated according to changes with edges and shades of the photographic images. However, the cell-like patterns are irregularly arranged. In this study, we propose a method to arrange the cell-like patterns along the edges of the photographic images. Our method improves on the conventional method by using Euclidean distance from the edges. We show that the cell-like images with the cell-like patterns arranged along the edges can be generated by our method through experiments using various photographic images.

OS5-2 Generating Striped Animations by Inverse Line Convergence Index Filter

Toru Hiraoka, Ryosuke Takaki (University of Nagasaki, Japan)

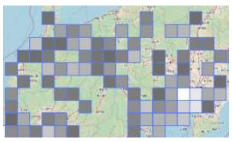
A non-photorealistic rendering method has been proposed for generating a striped image which is overlaid striped patterns in a photograph. The conventional method generates the striped image by an iterative process using an inverse line convergence index filter. When a striped animation is generated by converting each frame of a movie by the conventional method, flickering occurs in the generated striped animation. In this letter, we propose a method for generating a striped animation that has characteristic with less flicker from the movie. The effectiveness of the proposed method is investigated experimentally. As a result of the experiments, the proposed striped animation had less flicker than the conventional striped animation.



OS5-3 A Method for Estimating Home Location of Foreigners in Japan Using Photograph Location

Masaharu Hirota, Tetsuya Oda (Okayama University of Science, Japan)

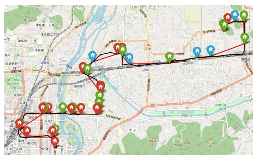
The attributes of travelers such as home location and age could be useful for many applications such as information recommendation and targeted advertisement. However, this information is not accessible in most Web services because users do not reveal them. We propose a method to estimate a foreign tourist's home location based on each country's tendency, in which tourists from certain regions tend to visit certain places when traveling abroad. The feature for the estimation uses the frequency of photograph location in a geohash. In this paper, we use foreigners in Japan as a case study. We evaluate the performance of our proposed method by using photographs obtained from their user accounts on Flickr.



OS5-4 A Proposal of Online Map-matching Based Trajectory Compression Algorithm Using Road Networks

Shota Iiyama, Tetsuya Oda, Masaharu Hirota (Okayama University of Science, Japan)

As the data size of GPS logs increases, the amount of data transferred from mobile devices to the server, and the computational cost of analysis of GPS logs increase. One of the solutions to these problems is to compress the GPS logs. However, it is difficult to compress a sparse GPS log while preserving the feature points in the GPS log, such as the user's movement speed, the shape of the GPS log trajectory, and the direction of movement of the GPS log. In this study, we propose an online compression algorithm for GPS logs that improves the compression ratio while preserving GPS logs' feature points. Our proposed method compresses GPS logs by using information from the road network to identify users' roads. We evaluate the performance of this method using the GPS data of the bus.



OS5-5 Case Study and Direction of Share Cycle System in Japanese Cities

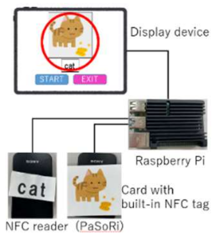
Minoru Kumano (University of Miyazaki, Japan), Toru Hiraoka (University of Nagasaki, Japan)

In the share cycle, people can share a bicycle with others, use and return it using multiple cycle ports, and register and borrow it with just an IC card or smartphone. The share cycle system is still new in Japan, and there are few papers even if you search the National Institute of Informatics. There is no case study. Therefore, in this paper, we selected 16 advanced cases nationwide and conducted a questionnaire survey in 2019. The purpose was to clarify the business purpose, issues, effects, features, numerical values such as the number of ports and the number of bicycles, and the direction. As a result, it was clarified that the purpose is to improve the ease of migration and to supplement public transportation, and to improve the profitability and the number of users are many issues, which contributes to tourism.

OS5-6 The IoT Solution to Play English Word Learning Tool

Shogo Aizawa, Motohide Yoshimura (University of Nagasaki, Japan)

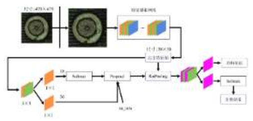
We innovate an IoT solution to play the English word learning tool. It uses a Raspberry Pi and two NFC readers. You can learn English words in three steps by using our tool. First, prepare NFC tags corresponding to English words and illustrations. Next, select a pair of the English word and illustration. Finally, hold them over the NFC reader, then the correct answers are displayed. Our solution has two aspects. The one is an educational tool for children and the other is an IoT toy playing with children. As to the former, the children can learn English words by intuitive operation such as operating the touch panel and holding the NFC tags. As to the latter, you can experience a part of an IoT technology such as NFC. In this paper, we report findings through the construction of the English word learning tool.



OS6 Intelligent Control (6)
Chair: Yingmin Jia (Beihang University, China)
Co-Chair: Weicun Zhang (University of Science and Technology Beijing, China)

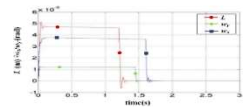
OS6-1 Faster R-CNN Based Defect Detection of Micro-precision Glass Insulated Terminals
Qunpo Liu[1], Mengke Wang [1], Zonghui Liu[1], Naohiko Hanajima[2], Bo Su[1]
([1]Henan Polytechnic University, P.R.China) ([2]Muroran Institute of Technology, Japan)

Micro-precision glass-encapsulated electrical connectors are the core components used in precision electronic equipment. As an electrical connector, its quality has a huge impact on the performance of precision electronic equipment. Due to limitations in materials and production processes, some of the micro-precision glass connectors produced have defects such as missing blocks, bubbles, and cracks. At present, it is difficult to ensure product quality and production efficiency with manual inspection methods. However, the defect characteristics of micro-precision glass connectors are quite different, and it is difficult for traditional defect detection technology to design an ideal feature extractor for detection. Therefore, this paper proposes to use deep learning technology to detect missing blocks. Firstly, the sample pictures of missing defects of electrical connectors are preprocessed, and then be used to train the deep learning network of Faster RCNN. According to the test results, the algorithm has an accuracy of over 80% in detecting missing defects in glass connectors.



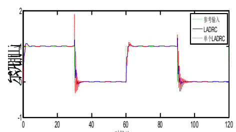
OS6-2 Adaptive Sliding Mode Control for a Constant Tension Suspension System
Yuxin Jia, Yingmin Jia, Kai Gong (Beihang University, China)

The constant tension suspension system can be used to counteract the gravity of the test object, and is widely used in the ground test of spacecraft control schemes. In this paper, the dynamic model of a constant tension suspension system is established, and a new adaptive sliding mode controller is designed. The simulation results show that the transient time and steady-state error of the control system designed in this paper are small, which satisfies the requirements of microgravity simulation.



OS6-3 Weighted Multiple Model ADRC for Uncertain Linear System
Weicun Zhang, Jing Ge (University of Science and Technology Beijing, China)

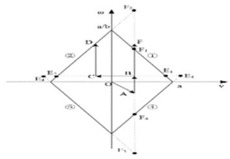
For uncertain linear systems such as parameter jumping, this paper establishes a fixed model set to cover the uncertainties of the real plant to be controlled, and for each local model with minor uncertainties, the corresponding local controller is designed based on Active Disturbance Rejection Controller (ADRC) approach. As a result, a weighted multiple model ADRC adaptive control framework is proposed. Some simulations have been conducted based on MATLAB to verify the effectiveness of the proposed weighted multiple model ADRC adaptive control algorithm.



OS6-4 Trajectory Tracking Control of Differential Wheeled Mobile Robot Based on Rhombic Input Constraints

Kai Gong, Yingmin Jia, Yuxin Jia (Beihang University, China)

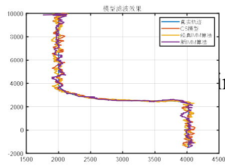
This paper focuses on the trajectory tracking control algorithm for the differential wheeled mobile robot (DWMR) based on rhombic input constraints. The kinematics and dynamics model of the DWMR are Established, and vector analysis method is used to design the time-varying parameters of the controller when the linear velocity and angular velocity of the DWMR were not independent of each other. Through the trajectory tracking simulation of the 8-shaped curve, a good control performance is obtained.



OS6-5 Maneuvering target tracking based on improved interacting multiple model algorithm

Weicun Zhang, Meiyu Zhu (University of Science and Technology Beijing, China)

In recent decades, with the rapid development of artificial intelligence, target tracking technology has become a major aspect of scientific and technological research. Mobile target tracking technology is widely used in military and civilian fields. For example, in air defense and air traffic control, reliable and accurate tracking of targets is always the main purpose of target tracking system design. This paper presents a new interacting multiple model (IMM) algorithm for maneuvering target tracking. Some simulation based on MATLAB have been conducted to verify the effectiveness of the proposed algorithm.

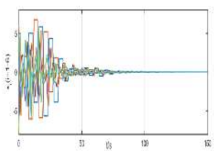


ng

OS6-6 Encapsualted Agents of Hybrid Order Discrete Dyanmics

Yunzhong Song (Henan Polytechnic University, P.R. China)

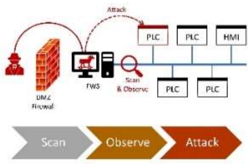
This paper is targeting to investigate the encapsulated cell realization of the hybrid order agents in discrete time, where it is composed by two different dynamic order agents, one is the first order and the another one is the second order, the work is the continued case of its continuous counter part of the encapsulated agents. To be specific, the exploration of the sampling speed is made distinguished from each other, where the fist order section is assigned at low speed, and the second order section is selected at comparatively high speed. To compensate the distinct sampling speed, the lifting technology is taken to analyze the encapsulated system. Theoretical analysis and simulation results are made available for the further reference.



OS7 Information Applications and Security (5)
Chair: Kuo-Hsien Hsia (National Yunlin University of Science and Technology, Taiwan)
Co-Chair: I-Hsien Liu (National Cheng Kung University, Taiwan)

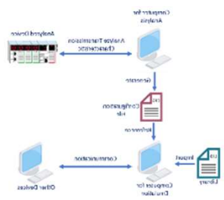
OS7-1 The Security Challenges with the Widespread Use of IT Infrastructure in ICS
Kuan-Ming Su, I-Hsien Liu, Jung-Shian Li (National Cheng Kung University, Taiwan)

The communication established by Ethernet is becoming more and more common in the industrial control systems (ICS), and it brings not only pros but also cons like vulnerabilities. To hijack or get access to industrial devices, the attacker must know the information about the devices first, but most of the devices, equipment, and systems in ICS are not public, which is like a black box to the attacker. Therefore, we generalized a standard procedure to attack ICS with Ethernet-enabled, which is able to get the information and access of the devices in an unknown ICS like identifying the manufacturer of programmable logic controllers (PLC) and overwriting the configuration of PLCs. Also, we will implement the procedure to demonstrate the feasibility of this procedure with the system obtained from the real industrial control field whose devices' manufacturer is Schneider and protocol is Modbus TCP.



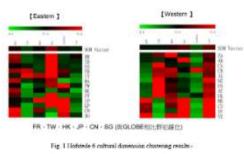
OS7-2 A Communication System with Equipment's Characteristics
Chia-Chun Lai[1], I-Hsien Liu[1], Chi-Che Wu[1], Chuan-Gang Liu[2], Jung-Shian Li[1]
([1]National Cheng Kung University, Taiwan) ([2]Chia Nan University of Pharmacy and Science, Taiwan)

Over the past few years, the application on internet has grown rapidly and the devices that can connect to internet also have an explosive increase. In the past, some information such as IP address, MAC address and communication port are commonly used to identify a specific device on the internet. However, using this kind of information is not enough for the purpose to identify precisely because the information can simply be imitated by any other devices. In our research, we focus on the communicational characteristics of device. And we also provide a mechanism that can analyze these characteristics and generate configuration references which other devices can utilize. With the customized communication module and the references mentioned above, we can make other device imitate the behavior of analyzed device and provide a better effect on device emulation.



OS7-3 Application of the Self-Organizing Map (SOM) to Analyze the Multiple Perspectives on Cross-National Culture
Li-Ming Chuang, Yu-Po Lee, Shu-Tsung Chao (Chang Jung Christian University, Taiwan)

This study integrates the previous cross-cultural literature and aims to construct an analysis model of cross-national culture with multiple dimensions from three important cultural dimension theoretical models commonly used in cross-cultural studies: Hofstede, Global Leadership and Organizational Effectiveness (GLOBE) and World Values Survey (WVS). This study uses a self-organizing map (SOM) as an analysis method to integrate 17 cultural variables from this multicultural dimension for cluster analysis and explains the cultural types in 26 countries based on the results. This study explores the differences and similarities of different countries in different cultural dimension analyses and provides a comparative model of multicultural analysis. This study takes samples from three cross-cultural analysis databases as data sources and employs the self-organizing map for analysis based on a neural network algorithm. The results identify the cross-cultural groups of 26 countries and reveal their key cultural similarities and differences. We also elaborate upon the findings of these cultural characteristics and multi-cultural dimensions.



OS7-4 Threats Hidden in Employee Workstation through Office Files

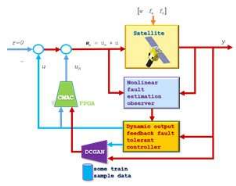
Tung-Lin Lee, I-Hsien Liu, Jung-Shian Li (National Cheng Kung University, Taiwan)

Due to the single sign-on implementation of windows operating systems, users could save a lot of time without having to enter their account and password again and again. However, this setting also allows hackers to Use some dedicated constructed malicious files to steal the user's identity verification information (NTLM hash), and then launch attacks like pass the hash, and then log in to the user's computer remotely, which becomes a way to gain access to the organization's intranet. Although there are many ways to cause an NTLM hash leak, fortunately, most of them cannot be exploited by hackers. For example, this paper will meet various common NTLM hash leak methods, and focus on the most commonly used method of hacking: Office files, and discuss them The possibility of extended application and under what circumstances will the existing protection mechanism be bypassed.

OS7-5 Microsatellite Attitude Control Approach: Combined with Generation Adversarial Networks Fault-Detection and Cerebellar Model Articulation Controller Fault-Tolerant Control

Ho-Nien Shou (Air Force Institute of Technology, Taiwan)

A new attitude control architecture for microsatellite is proposed. Based on the deep learning fault-detection method, the Cerebellar Model Articulation Controller(CMAC) is used as the fault-tolerant control. With the function of Generation Adversarial Networks(GAN) to recognize images, the microsatellite attitude fault wavelet spectrum is used as a guide for the training of generators and discriminators, and for system real-time fault diagnosis and classification. When the system fault diagnosis determines that the fault occurs, the cerebellar neural network is involved in fault-tolerant control. The GAN learning ability of the generative confrontation network is used to solve the problem of insufficient sample data and sample labeling respectively. The CMAC is used as a local learning network, which has generalization ability, strong convergence speed, and easy software and hardware implementation. The simulation results show that compared with other methods, the GAN method of fault-detection combined with CMAC can achieve higher accuracy and robustness.



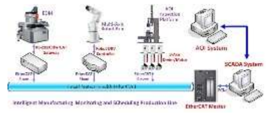
OS8 Intelligent Systems and Control (5)

Chair: Chung-Wen Hung (National Yunlin University of Science and Technology, Taiwan)
Co-Chair: Kuo-Hsien Hsia (National Yunlin University of Science and Technology, Taiwan)

OS8-1 Networking Integration Application of an Intelligent Production Line for Aerospace Precision Manufacturing

Chau-Chung Song, Chun-Chi Wang, Chen-Pang Chen, Chung-Wen Hung
(National Yunlin University of Science and Technology, Taiwan)

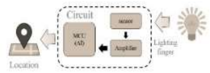
In this paper, the networking integration application of an intelligent production line is studied and developed for aerospace precision manufacturing. The technical development of the performance equalization and control for EDM machine is also introduced with the related technologies of the machine network and EtherCAT network. In addition, the development of intelligent production lines is integrated to improve the system stability and production efficiency of production line systems for aerospace component manufacturing. The key technologies of Cyber-Physical System (CPS) and intelligent networking is also integrated to constructs the real-time intelligent system monitoring system (SCADA) for data record and management. The key technologies of through-hole process are developed for intelligent diagnosis and through-hole inspection of related production processes. The integrated robot arm is responsible for the loading and unloading of the material, and the entire production line is connected by EtherCAT network. The automatic optical inspection (AOI) is introduced at the rear end of the production line to realize the quality judgment of the finished product. Furthermore, the virtual cloud network and intelligent factory is achieved with the information collection and intelligent monitoring system.



OS8-2 Low-Cost Indoor Localization Using Sound Spectrum of Light Fingerprints

Chung-Wen Hung, Hiroyuki Kobayashi, Jun-Rong Wu, Chau-Chung Song
(National Yunlin University of Science and Technology, Taiwan)

A low-cost indoor localization system using sound spectrum of light fingerprint is proposed in this paper, an artificial intelligence, AI, algorithm will be implemented in a low-cost micro-control unit, MCU, to perform the localization. The unit, complex and tiny differences of the light fingerprints are caused by the different characteristics of the discrete components used in lighting devices. To reduce the memory size for low-cost MCU, only sound spectrum of light fingerprint is used to identification the lighting device. Moreover, the grid search also is used to optimize the hyperparameters to compact the AI model. The system architecture and algorithm development are discussed in this paper, and the experimental results will be present to show the proposed system workable.



OS8-3 Air Valve Fuzzy Control Combined with Sheet Music Recognition Techniques Applied to Autoplaying Soprano Recorder Machines

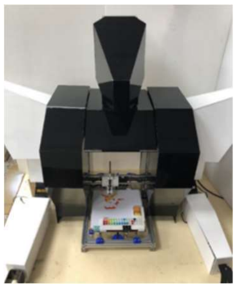
Chun-Chieh Wang, Guang-Ming Jhang (Chienkuo Technology University, Taiwan)

In the past research, there are many disadvantages to score recognition and flute performance. Therefore, this paper would improve on these two points. For the part of music score recognition, we use the y-axis projection method to detect the staff position and remove it to replace the erosion and expansion in morphology. After research and observation, it was found that the notes have a specific writing style on the staff. Therefore, we use this feature to distinguish notes. Detect note stem by x-axis projection to distinguish them in the first stage, then remove the stems to distinguish them in the second stage. After the notes are distinguished, the pixel clustering method is used to identify the scales. The center point of the rune head is compared with the staff position. The scale codes are arranged in order from bass to treble. Moreover, the performance is arranged according to basic music theory. The soprano recorder performance part, using the motors to control the finger to press the blow hole will not keep up the tempo. So we changed the motor to a solenoid valve to control the pneumatic cylinder to press the blow hole. In addition, the recorder will have different pressures depending on the pitch. So we increased the part of the control from one air valve to three air valves. Not only that, we have divided the range of the recorder into three different ranges of bass, midrange and treble, which greatly improves the situation that the original sound is generated when playing. Finally, we use fuzzy control theory to control the air valve. Experiments prove that the air valve fuzzy control combined with sheet music recognition techniques can fully realize the functions of autoplaying soprano recorder machines.

OS8-4 Image Inpainting Techniques Combined with Isolated Pixel Filtering Applied to Multifunctional Drawing Robots

Chun-Chieh Wang, Zhan-Xian Ye (Chienkuo Technology University, Taiwan)

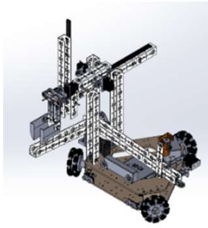
The purpose of this thesis is to assist street painters to improve their work efficiency. That is, the simpler part of the drawing is drawn by the robot. Once the initial picture is completed, the artist will take over the painting. In this paper, HSI color space is used to improve the effect of color simplification such that the recognition of image processing results is enhanced. Moreover, the isolated pixel filtering is used to replace the less-affected isolated point color with the surrounding color. Furthermore, Image Inpainting Techniques are utilized to reduce the distortion caused by the isolated pixel filtering. Besides, we adjusted the path planning as well as reduced isolated points to dramatically reduce drawing time. At the same time, LabVIEW can issue commands directly to control the robot by adding the communication function. In addition, in order to achieve the multifunctional drawing robotic mode, this system adds the sketch function. Through the image resolution adjustment as well as the shortening of the spacing of the drawing lines, the robot can draw more detailed pictures in the same size of drawing space. Simultaneously, we reserved some space in the organization to provide the use of subsequent development of other functions. The measured results confirm that the application of the technology in this paper can shorten the drawing time by about 55% on the multifunctional drawing robot system.



OS8-5 Mobile Robot with Image Recognition -- Using LabVIEW and KNRm

Kuo-Hsien Hsia, Bo-Jung Yang, Jr-Hung Guo, Chang-Sheng Xiao
(National Yunlin University of Science and Technology, Taiwan)

The main purpose of this paper is to use the image recognition of LabVIEW software to construct mobile robots with various functions, and make the robots applicable to the industry and have web monitoring applications. The core of the robot is mainly the KNRm controller. This controller is suitable for beginners, and can be connected to DC servo motor, RC servo motor, infrared, ultrasonic and camera to achieve various functions of the robot. The structure of the robot uses metal parts sold by Studica company, which can be in accordance with the desired function to assemble the robot. Since the company is a designated equipment sponsor company for World Skills competitions, it can also be in line with international standards. Finally, PID control and sensors are added to make the robot movement and position more accurately.



OS9 Intelligent Systems and Life (7)

Chair: Kuo-Hsien Hsia (National Yunlin University of Science and Technology, Taiwan)
Co-Chair: Evgeni Magid (Kazan Federal University, Russia)

OS9-1 Architecture of a student training computer program for preparing professional outpatient consulting skills within an electronic medical records system during COVID-19 alertness situation

Sergey Bulatov[1], Evgeni Magid[2], Enzhe Kharisova[1], Roman Lavrenov[2], Vitaly Dudin[1], Artur Khazetdinov[2] ([1]Kazan State Medical University, Russia), ([2]Kazan Federal University, Russia)

One of the important goals of a medical education in the context of the COVID-19 pandemic is a broad introduction of simulated-based teaching methods, which might include elements of robotics and artificial intelligence. We analyzed computer programs that are currently used for maintaining medical records of patients by various polyclinics of Kazan. Based on these results, we summarized requirements for a training computer program that could provide students with medical records maintaining skills and developed program's preliminary architecture. Our main attention focused on modeling situations that are associated with pre-hospital stage processing of medical records for patients with suspected or confirmed COVID-19.

OS9-2 Satisfaction Assessment on the Counseling Service System for Full-Time Teacher-Counselor in Tainan Elementary School

Hsiu-Hao Liu (Chang Jung Christian University, Taiwan)
Yun-Syuan Jhang (National University of Tainan, Taiwan)

As the most important case management system in the school counseling works, the counseling service system must meet the needs and expectations of the school teacher-counselor (the main users). This investigation conducted a questionnaire among teacher-counselor to survey their satisfaction when using the counseling service system in Tainan City. The results of satisfaction items are as follows: "The items are provided abundant and complete by the system" was the highest of all items. "The items are provided to meet the needs to be collected in the counseling works by the system" was secondly. The least satisfactory item was "Using the system can help me communicate and interact with others in the counseling works". Additionally, system usage intention was effectively explained by perceived ease of use and usage attitudes. The conclusion was to discuss the implications of this research and suggest several future research issues.

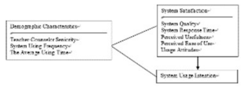
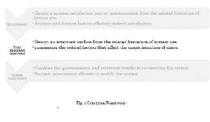


Fig. 1 Conceptual Framework

OS9-3 Assess The Critical Factors for the Counseling Service System Usage Intention

Hsiu-Hao Liu (Chang Jung Christian University, Taiwan)

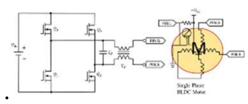
As the most important case management system in the school counseling works, the counseling service system must meet the needs and expectations of the school teacher-counselor (the main users). Semi-structured interviews were used to interview teacher-counselor who have experience in using systems in Tainan. The data are recorded and semantic analyzed after interviewing, summarize the critical factors that affect the usage intention of the teacher-counselor. The conclusion was to discuss the implications of this research and suggest several future research issues.



OS9-4 Design and Implementation of EMI Suppression Filter for Electronic Commutation Fans

Ching-Chun Chuang, Chih-Chiang Hua, Chung-Wen Hung, Chun-Jen Yao
(National Formosa University, Taiwan)

This paper presents an electronically commutated fan has a complex digital control compared to conventional alternating current fans. Since serious electromagnetic interferences occur in the electronic commutation fan, a filter is required and its mechanical design is very important. A digital control for an electronically commutated fan is used to reduce the components of the EMI filter, and the suppression of conducted and radiated electromagnetic interference is discussed. Moreover, the fan system is integrated with proper mechanical design to reduce the radiated electromagnetic interference. Fig.1 shows a single-phase BLDC motor driver for the proposed fan. The electromagnetic interference produced by the electronically commutated fan is analyzed and the suppression strategy is provided. The experimental verification for the electronically commutated fan will be made available to the public.



OS9-5 Application of IOT to Forest Management Taking Fushan Botanical Garden as an Example

Shuo-Tsung Chen, Chih-Chiang Hua, Ching-Chun Chuang
(National Formosa University, Taiwan)

In recent years, the Internet of Things (IoT) technology has developed rapidly and has been successfully applied in different fields, and has begun to expand the application context of the Internet of Things. This work aims to study how to apply the IoT technology to forestry management, including: 1. Forestry management using wireless network communication technology of Low Power Wide Area Network (LPWAN) such as LoRa and NB-IoT; 2. Apply different sensing technologies to survey resource of forest and monitor the microclimate changes in forest. In order to actually verify the proposed feasible communication technology and sensor arrangement, we chose the Fushan Botanical Garden with the most diverse and complex terrain in Taiwan as the experimental site. Fig. 1 shown that the evaluation of a situation and setting for Forest Resources Survey. We actually built LoRa and NB-IoT communication equipment (including relay equipment) and various sensors to test the influence of terrain, climate, and trees on signal transmission efficiency and equipment installation. The returned data also verifies the successful operation of various communication devices and sensors.

OS9-6 Development of the IoT Module using MQTT Protocol and AES

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

The efficiency and safety of the Internet of Things have always been the focus of the development of IoT devices. Because the chips used in IoT devices generally have poor computing power, they cannot transmit data quickly and in large amounts, and use more complex security algorithms. Therefore, this thesis is to develop an IoT module with STM32 chip as the main controller. This module uses MQTT Protocol and AES encryption technology, and this IoT module can be operated directly with a browser. MQTT is a communication protocol for the Internet of Things, which was developed by IBM and Eurotech, and officially became an OASIS international standard in 2014. The purpose of development is to send and receive processing messages under narrow bandwidth and low energy consumption conditions. To ensure the security of IoT communications, we use AES encryption technology. Through this design, the communication of the entire IoT module is more efficient and safe. Finally, we applied this IoT module to the security system, and the overall efficiency and safety have been verified. In the future, we will continue to improve related software and hardware so that this IoT module can be used in different fields.

OS9-7 Exploring the Intention to Continuance of Learning Programming at Elementary School of Rural Area by the mBot Robot

Yung-Hsin Cheng, Jia-Ming Hsiao (National Yunlin University of Science and Technology, Taiwan)

Since Curriculum Guidelines of 12-Year Basic Education implemented by the Ministry of Education in 2018, the program learning courses have been added to junior high school education. However, there are no programming course in the elementary school. This study is proposed to explore the continuity and intentions of the rural area students in the programming course with mBot robot and mBlock programming tool through the Post-Acceptance Model of IS Continuance. It is indicated that enlightenment education of computational thinking should be implemented during the elementary school by means of graphical programming software and robot practice. Through the graphical software and teaching robots, we can cultivate problem solving skills for students' logic, creative thinking and communication, as well as through the task-guided way to train the students' concentration and perseverance.

OS10 Advanced Control Systems and Signal Processing (7)

Chair: Takuya Kinoshita (Hiroshima University, Japan)
Co-Chair: Shinichi Imai (Tokyo Gakugei University, Japan)

OS10-1 Nonlinear Internal Model Controller based on Local Linear Models, and its Application

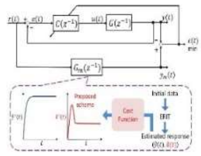
Shinichi Imai (Tokyo Gakugei University, Japan)

In this paper, nonlinear internal model controller based on local linear models, and its Application. The internal model control has a simple structure and has a high robustness for system uncertainties. However, there are few studies of internal model control schemes for nonlinear systems. On the other hand, many controlled systems have the nonlinearity. The effectiveness of the newly proposed control scheme is numerically evaluated on experiment examples in comparison with the conventional control methods for nonlinear systems.

OS10-2 Design of a data-driven control system for reference model design using predicted signals

Yuki Nakatani, Takuya Kinoshita, Toru Yamamoto (Hiroshima University, Japan)

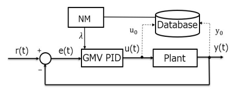
In recent years, data-driven control schemes that do not require system identification have been actively studied. Generally, it is easy to give a reference model focusing only on the output response. In contrast, it is difficult to give a reference trajectory considering the input signals based on the controlled system's characteristics. Furthermore, it is necessary to consider the output signal and the input signal since there is a limit of the actuator performance in the control design of the actual machine. This paper proposes a data-driven control system that can predict the input/output response of an unknown system in offline using operation data. The effectiveness of the proposed scheme is numerically verified.



OS10-3 Design of a Data-driven GMV Controller Using the Nelder-Mead Method

LiYing Shi, Zhe Guan, Toru Yamamoto (Hiroshima University, Japan)

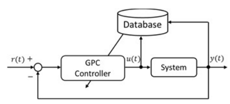
This paper presents a design scheme that can obtain the optimized generalized minimum variance (GMV) control parameters by applying the Nelder-Mead (NM) method based on proportional-integral-derivative (PID) controller for linear systems. The NM method is used to find the most suitable parameter λ of GMV controller, then the PID parameters can be obtained. In the previous GMV controller, the PID parameters are calculated by simply changing λ manually. Therefore, it is hard to get desirable control performance. The application of NM method can optimize the calculation of the most suitable λ and get better PID parameters. Furthermore, the estimation of closed-loop response uses data-driven approach. The effectiveness of the proposed scheme is verified by using a simulation example.



OS10-4 Design of a Databased-Driven GPC for Nonlinear Systems

Zhe Guan, Toru Yamamoto (Hiroshima University, Japan)

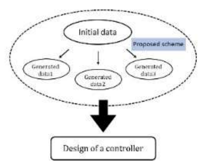
This paper presents a Generalized Predictive Control (GPC) design scheme considering databased-driven approach for nonlinear systems. In several conventional design methods, model parameters are required to calculate control parameters. However, it is time- and cost-consuming to identify the model in real practical systems, especially for chemical or thermal industries with unknown delay time and strong nonlinearity. Recently the database-driven approach has been attracted considerable attention to tackle the model identification issue. It is widely applied in nonlinear systems. On the other hand, GPC is considered in the cases with unknown delay time. As a result, the controller is designed based on GPC by using databased driven approach. The proposed scheme is verified by a numerical simulation which demonstrates the effectiveness.



OS10-5 Design of a Data-Driven Controller using Open-Loop Data

Y. Nishiya, Takuya Kinoshita, Toru Yamamoto (Hiroshima University, Japan)

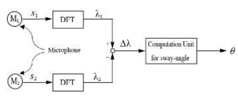
In recent years, data-driven control has been proposed as a control system design that does not require system modeling. Furthermore, it is extended to a nonlinear system using a database. It is necessary to collect several data to obtain good control performance. However, collecting sufficient data incurs both time and human costs. This paper proposes an offline scheme for generating several data from a set of open-loop data is proposed. Additionally, the controller is designed considering the input signal to prevent a heavy burden on the actuator. In this paper, the numerical examples verify the effectiveness of the proposed scheme.



OS10-6 Improved Estimation of Sway-Angle for Overhead Crane based on Phase Difference of Acoustic Signals in Frequency Domain

Hanako Ogawa, Takeshi Yamada, Masayoshi Nakamoto (Hiroshima University, Japan)

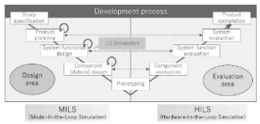
When operating crane, sway of the payload is an issue that should be controlled. For the anti-sway control, it is essential to detect the sway-angle θ. However, it is difficult to measure the sway-angle directly. In this paper, we propose a method to obtain the phase difference (λ1 - λ2) of the acoustic signals s1 and s2 in the frequency domain by using the discrete Fourier transform (DFT). Based on the phase difference, we compute the arrival time difference and convert that to the sway-angle of deflection. To improve the accuracy, we discuss a method for generating the frequencies of the sound sources and the signal processing with filters. In the experiments, the sway-angle measured by the camera are compared with the proposed method.



OS10-7 Study on an Optimal Design Method for Control Systems based on Bayesian Optimization

Koichi Hirota, Shin Wakitani, and Toru Yamamoto (Hiroshima University, Japan)

Most product systems consist of a combination of multiple subsystems. Appropriate setting and design of operating goals (functional goals) of these subsystems are important to achieve and maintain the target performance of the system. In recent years, model-based development (MBD) has been attracting attention in the industrial world. In the MBD process, a user can evaluate the subsystem's performance and easily improve their design. In many situations of control system design, plant design is completed in advance. In subsystem design, a plant and a controller are required to be optimized at the same time for the functional goals determined upstream design process. Based on the background, this study considers the simultaneous optimum design of the controller and the plant.



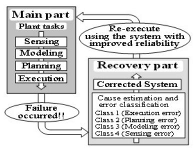
OS11 Robotic Manipulation (3)

Chair: Kensuke Harada (Osaka University, Japan)
Co-Chair: Tokuo Tsuji (Kanazawa University, Japan)
Co-Chair: Akira Nakamura (AIST, Japan)

OS11-1 Using Various Evaluation Standards to Determine an Error Recovery Process in an Automation Plant

Akira Nakamura[*1], Natsuki Yamanobe[*1], Ixchel Ramirez Alpizar[*1], Kensuke Harada[*2], Yukiyasu Domae[*1]
([*1] National Institute of Advanced Industrial Science and Technology (AIST), Japan
[*2] Osaka University, Japan)

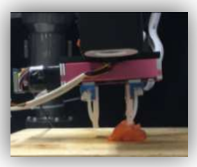
In an automation plant, not only simple repetition tasks but also complicated tasks are carried out. An error is more likely to occur in such difficult work, so the improvement of the technique to perform recovery of an error is needed. The task often is re-executed after returning to previous step, in the case of a big error. Therefore, it becomes the important problem to decide both the past step that it should return to and the recovery planning after returning. In this paper, error recovery planning taking these two factors is proposed by using various evaluation standards.



OS11-2 Robotic Picking for Piled Sushi Topping

Kenta Matsuura[*1*2], Keisuke Koyama[*1], Weiwei Wan[*1], Kensuke Harada[*1]
([*1] Osaka University, Japan, [*2] Currently with Yaskawa Electric Co. Ltd., Japan)

This paper proposes a method for picking piled sushi topping. By observing a human motion picking a sushi topping, we propose two picking strategies where one is to insert a finger into the separation among two toppings with shaking the finger, and the other is to insert a finger into the separation between a topping and a table. Along with two segmentation method, i.e., plane segmentation and LCCP segmentation methods, we experimentally verify the effectiveness of the proposed approach.



OS11-3 Motion Generation by Learning Relationship between Object Shapes and Human Motions

Tokuo Tsuji[1*], Sho Tajima[*1], Yosuke Suzuki[*1], Tetsuyou Watanabe[*1], Shoko Miyauchi[*2], Ken'ichi Morooka[*2], Kensuke Harada[*3], and Hiroaki Seki[*1]
([*1] Kanazawa University, Japan, [*2] Kyushu University, Japan, [*3] Osaka University, Japan)

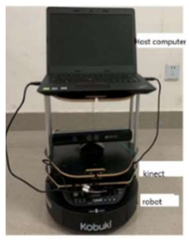
(This paper presents a method for planning a robot motion of daily tasks by learning the relationship between object shapes and human motions. Robots are required to be able to deal with multifarious objects in various categories. However, it is difficult for robots to plan motions automatically for performing a task because objects even in the same category have different shapes. In our method, the motions are estimated by learning the relationship between object shapes and human motions using linear regression analysis. We evaluate the estimated motions and the experimental results of tasks which are performed by a robot to verify the effectiveness of our proposed method.

OS12 System and Control (18)
Chair: Huailin Zhao (Shanghai Institute of Technology, China)
Co-Chair: Jichao Zhao (Tianjin University of Science and Technology, China)

OS12-1 The Optimized Intelligent Algorithms on Face Recognition and Tracking for ROS-based Robots

Yue Chen, Shuhao Tian, Huailin Zhao, Shengyang Lu (Shanghai Institute of Technology)

With the development of artificial intelligence, face recognition and tracking technology have been widely used in many fields such as target positioning, automatic driving, and human-computer interaction. Recently, a large number of face detection, recognition and tracking algorithms have emerged, but there are still many shortcomings in practical applications, such as slow face detection, low detection accuracy, and face recognition and tracking for ROS robots Algorithms are rare. This paper improves the traditional Haar-like algorithm and LK optical flow tracking algorithm, and designs a ROS robot platform based on the improved algorithm. By comparing the accuracy and timeliness of face detection and tracking between the improved algorithm and the traditional algorithm, the superiority of this design algorithm is obtained.



OS12-2 Control of a Novel 5D Hyperchaotic System

Qiang Wei[1], Hong Niu[2]
([1]Army Military Transportation University, China)
([2]Tianjin University of Science & echnology, China)

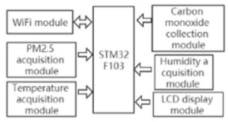
In this paper, a novel five-dimensional (5D) autonomous hyperchaotic system is presented, and the characteristics of the 5D system are given in brief. For control of the 5D hyperchaotic system, a linear feedback controller is designed via the Lyapunov stability theory, so that the 5D system is no longer hyperchaotic but globally asymptotically converges to the equilibrium point at the origin. The numerical simulation results are given to illustrate the feasibility and effectiveness of the method.

$$\begin{aligned}
\dot{\tilde{x}} &= a\left(\tilde{y}-\tilde{x}\right)+u_{c1}\\
\dot{\tilde{y}} &= \left(c-a\right)\tilde{x}+c\tilde{y}+\tilde{w}-\tilde{x}\tilde{z}+u_{c2}\\
\dot{\tilde{z}} &= -b\tilde{z}+\tilde{x}\tilde{y}+u_{c3}\\
\dot{\tilde{v}} &= m\tilde{w}+u_{c4}\\
\dot{\tilde{w}} &= -\tilde{y}-h\tilde{v}+u_{c5}
\end{aligned}$$

OS12-3 Design of Multifunctional Vehicle Interior Environment Monitoring System Based on Wireless Communication

Yuqi Yan, Jialin Yang, Zhongxu Qin (Wuhan University of Technology, China)

The interior environment of the car affects the driver's mental state to a certain extent, so it is necessary to design a multi-functional interior environment monitoring system. In this paper, STM32103 single chip microcomputer is used as the core controller, which integrates multiple sensors to collect various environmental information in the vehicle, and transmits the data to the monitoring platform through WiFi wireless communication. When the detected data exceeds the preset threshold, the monitoring system will generate an alarm to remind the user to carry out relevant operations. The system has the advantages of low cost, convenient use and high precision, and has certain commercial value and market prospect.



OS12-4 Robot Structure and Motion control Design Based on UG and Product

Yuhang Sheng (Tianjin University of Science and Technology, China)

The paper is to design a six-degree-of-freedom biped robot by the research on the humanoid characteristics of the biped robot. Our biped robot chooses the steering gear ASMC-03B as its power unit, we restricts the model parameters by analyzing the function relationship between the steering gear torque and the volume, and regulates the size of every parts, and create a motion analysis model. In the hardware part, Arduino UNO, which is used as the main control chip, realize the communication between the main control chip and the servo drive module PAC9685 through the IIC bus protocol, which saves the main control chip resources and ensures the execution efficiency. In the simulation part, it is to simulate the steering angle of the steering gear, and output it in the form of a waveform.

OS12-5 Design of Brushed Motor Position Loop Control System Based on Incremental PID

Tianyi Zhang, Peng Lu (Tianjin University of Science and Technology, China)

Motors are an important part of the transmission and control system. With the development of technology, the focus of motors in practical applications has shifted to precise control of speed, position and torque. This paper focuses on the control system design of the brush motor position loops, using STM32 microcontroller and incremental PID to achieve accurate control of the motor stopping moment position. The software is designed in C programming. It is divided into four parts: System Clock Configuration Functions, Principal Functions, System Tick Timer Interrupt Callback Function, Position closed loop PID Control Design Function. The hardware part is based on the STM32F103VET6 core board.

OS12-6 Design of a Fire alarm system

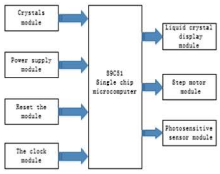
Yuhui Cheng, Tianyi Zhang (Tianjin University of Science and Technology, China)

This paper introduces the hardware and software design of the fire alarm system, and analyzes its working principle. This paper uses a distributed sensor network based on ZigBee to achieve fire alarm and real-time evacuation. Select infrared flame sensor and MQ2 smoke sensor as the information source. The core control board is the Arduino Mage2560 development board. Through the ESP8266 module and the GPRS module, the multi-channel information transmission function can be realized. The alarm can be classified according to the fire situation. System status monitoring software, written in C#, can view the status of each node in real time.

OS12-7 Design of intelligent curtain control circuit based on single chip Microcomputer

Qianqian Zhang, Jichao Zhao, Haokang Wen, Hongbo Hao
(Tianjin University of Science and Technology, China)

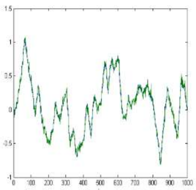
Since entering the 21st century, high technology has promoted the development of human beings and artificial intelligence has been popularized gradually. In this paper, the structure and principle of crystal oscillator circuit, reset circuit and photosensitive sensor circuit are introduced with 89C51 single chip microcomputer as the main control unit. The whole circuit system is analyzed in this paper. The circuit module of photosensitive sensor can be used to detect the external light intensity and automatically control the curtain. The user can set the temperature threshold with the remote control.



OS12-8 Research on the algorithm of flue gas desulfurization system

Hongbo Hao (Tianjin University of Science and Technology, China)

In this paper, the process of flue gas desulfurization and denitrification, which is a nonlinear, time-varying, large lag and strong coupling complex variable process, is analyzed and studied in depth. Based on the analysis of the coupling characteristics of the adsorption tower, the control model of the desulfurization and denitrification process of the adsorption tower is established, and the system identification adaptive PID Decoupling control algorithm is adopted. The Simulink toolbox corresponding to Matlab software is used to simulate and verify the effectiveness of the algorithm. The RBF neural network identifier is used to identify the plant model on-line to improve the adaptive ability of the controller.



OS12-9 A Survey of Low Visibility Image Enhancement Based on MSRCR Algorithm

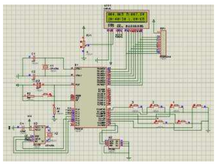
Haokang Wen, Hongbo Hao (Tianjin University of Science and Technology, China)

With the development of computer vision systems in the fields of traffic and safety monitoring, image enhancement has become an important research direction of computer vision. After development, the Retinex algorithm has a better enhancement effect, and the MSRCR algorithm has now developed into one of the important methods in image enhancement. This article introduces the development of the Retinex algorithm, and focuses on the main process and steps of the MSRCR algorithm, and uses the algorithm to enhance the low-visibility images of haze and night in multiple scenes. The actual results show that the MSRCR algorithm has a better enhancement effect and has a wide range of application values.

OS12-10 Design of a Taxi Meter Based on Single Chip Computer

Haokang Wen, Min Wang (Tianjin University of Science and Technology, China)

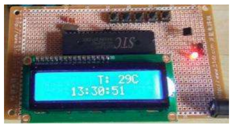
As a convenient means of transportation, taxis greatly facilitate people's daily travel. Based on such a background, this article designed a simple taxi meter based on STC89C52 microcontroller. This taximeter uses a single-chip microcomputer as the core, combines display module, clock module, storage module, key module, and combines software and hardware to build the required taxi meter model. The taxi meter introduced in this article has the functions of displaying time, starting price, and real-time display of the cost according to vehicle mileage. Theoretically, the meter has higher accuracy and has a wider practical value in daily life.



OS12-11 Design of a High Precision Digital Clock Based on Single Chip Microcomputer

Haokang Wen, Qiang Zheng (Tianjin University of Science and Technology, China)

Compared with traditional mechanical clocks, digital clocks have higher accuracy and durability and are widely used in people's daily lives. This paper designs a digital clock which uses AT89S52 single-chip microcomputer as the main control chip. It can count hours, minutes, and seconds, and can calibrate the time. It can switch between 24-hour and 12-hour systems. As a smart clock, while displaying the time, it also adds the function of temperature display. The intelligent digital clock designed in this paper has stable performance in theory and has certain practical value.



OS12-12 Design of Manchu Image Acquisition System Based on STM32

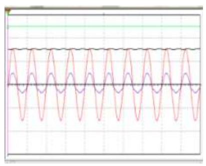
Hongbo Hao[1], Fulin Zheng[2]
([1]Tianjin University of Science and Technology, China)
([2] Dalian Minzu University, China)

This paper designs a Manchu image acquisition system based on STM32 to protect and utilize Manchu ancient books. The system uses stm32f407 as the core controller, with the help of image sensor, LCD screen, SD storage module to complete the function of image acquisition, storage and upload of Manchu ancient books. This paper describes in detail the hardware circuit design, software programming and overall function debugging of the image acquisition system. In order to achieve the purpose of digital protection of Manchu ancient books, we can collect, display, save and upload images in different situations.

OS12-13 Design of timing socket based on single chip microcomputer

Hongbo Hao[1], Shuailin Chen[2]
([1]Tianjin University of Science and Technology, China)
([2]Qingdao University of Science and Technology, China)

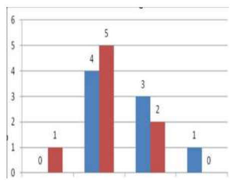
In this paper, the design of timing socket is mainly used for intelligent timing control of household appliances, and the working time of timing is set by pressing the key. In the working period, the external connection device and the power supply are in the conduction state. Outside the set time period, the external connector is disconnected from the power supply. This design uses STC89C52 as the driving chip. DS1302 clock module provides accurate timing. LCD1602 LCD module displays time and operation interface. The relay module controls the connection between the socket and the power supply. DC power supply module is converted to DC power supply module.



OS12-14 Study on the effect of physical fitness training on children's cognitive ability

Jianhua Deng, Lei Ning (Yanshan University, China)

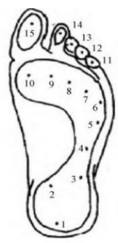
With the word physical fitness getting more and more attention in our country, physical fitness has been widely developed in our country. Therefore, in recent years, the students' physique has received attention from all aspects of society, and improving the students' physique has also become a new topic for physical education workers. I try to discuss the effect of physical fitness training on children's cognitive ability from the perspective of physical fitness training. To provide theoretical basis and reference for intervention of children's cognitive ability training methods. So as to promote the improvement of children's cognitive ability and provide guarantee. This paper studies the influence of physical fitness training on children's cognitive ability by using the methods of literature, observation and experiment. Finally, the conclusions are drawn.



OS12-15 Effects of high heels on plantar stress in women

Jianhua Deng, Lei Ning (Yanshan University, China)

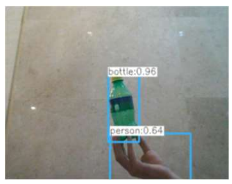
OBJECTIVE: to investigate the changes of Plantar pressure and gait in women wearing high-heeled shoes for different years. METHODS: from March 2019 to April 2019, a total of 30 female faculty members in Yanshan University were surveyed by questionnaires and plantar pressure was measured by static and dynamic tests. RESULT: 1. The peak value of plantar pressure of the arch of foot increased with the increase of wearing time (P & Lt; 0.05). 2. The plantar pressure and impulse of the second metatarsal and the fourth metatarsal increased with the increase of wearing time (P & Lt; 0.05). CONCLUSION: 1. Long-term wearing high-heeled shoes will cause foot discomfort, easy to cause flat foot and Hallux Valgus Lesions; 2. The distribution of plantar pressure and the change of arch shape and gait may be the main reasons for wearing high-heel shoes.



OS12-16 A research on Intelligent Classification of Urban Trash Bins based on Machine Learning

Longyu Gao, Zilong Liu, Luqi Shen, Songyun Shi, Yongzheng Lv
(Tianjin University of Science and Technology, China)

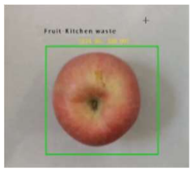
Aiming at the problems of inaccurate, insensitive and general performance of the current intelligent garbage sorting bins, an intelligent garbage sorting bin based on machine vision is proposed. The trash can is mainly divided into five parts: the main control module, machine vision module, classification module, overflow reminder module, and Wi-Fi Internet of Things module. The trash can uses convolutional neural networks to build an intelligent garbage classification model and classification algorithm to achieve rapid and accurate garbage classification. The experiment will be done based on the identification of waste bottles, analyze the recognition characteristics of machine vision, and then propose methods to improve the accuracy of recognition.



OS12-17 A Design on Intelligent Public Trash Can based on Machine Vision and Auxiliary Sensors

Longyu Gao, Leixin Han, Jiangyu Wu, Mingfei Liu, Ruming Kang
(Tianjin University of Science and Technology, China)

In order to improve the correct rate of front-end recognition in the garbage classification process, based on the machine vision technology, the automatic garbage classification system is designed and it has a significant improvement in recognition accuracy compared to traditional smart garbage cans. But in the case of identifying irregular garbage, the recognition accuracy is greatly reduced. In order to solve this kind of problem, four types of auxiliary sensors are added to the trash can, through the mutual cooperation between the sensors, combined with the results of machine vision recognition, comprehensive judgment, greatly improved the recognition accuracy of irregular garbage.



OS12-18 A research on Front-end Garbage Classification based on Machine Vision

Longyu Gao, Zhiqing Xiao, Junlong Hao, Luqi Shen, Manqian Hu
(Tianjin University of Science and Technology, Tianjin, China)

Adding a machine vision recognition module to the traditional smart trash can can effectively improve the efficiency of trash recognition. The intelligent garbage classification model constructed by the convolutional neural network can accurately identify the types of garbage, with an average accuracy rate of 0.87. Deploy the trained model on openMV and test it on the produced physical trash can. After the system is stable, the average time to complete a sorting and recovery is 2s. Experiments show that the system can effectively identify the types of garbage and complete garbage classification and recycling.

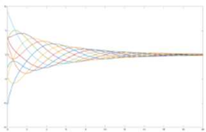
OS13 Intelligent Systems and Robotics (14)

Chair: Fengzhi Dai (Tianjin University of Science and Technology, China)
Co-Chair: Yunzhong Song (Henan Polytechnic University, China)

OS13-1 Analysis of the Consensus Protocol of Heterogeneous Agents with Time-Delays

Jichao Zhao[1], Fengzhi Dai[1,3], Yunzhong Song[2]
([1]College of ElTianjin University of Science and Technology, China, [2]Henan Polytechnic University, China, [3]Tianjin Tianke Intelligent and Manufacture Technology CO., LTD, China)

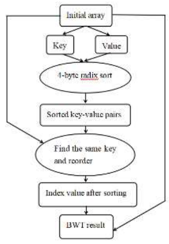
In practical engineering applications, the research on the consensus of heterogeneous multi-agents is of great significance. The consensus of multi-agents mainly includes average consensus, maximum consensus, and minimum consensus. The paper studies the average consensus of heterogeneous multi-agents, including continuous-time consensus protocol, discrete-time consensus protocol, consensus with time delay, and consensus of switching topology. The knowledge of graph theory is used to describe the system, and the results are simulated and analyzed for the consensus of time delay and switching topology to verify the correctness and effectiveness of the consensus protocol.



OS13-2 Burrows-Wheeler Transform Acceleration based on CUDA

Chang Sheng, Fengzhi Dai (Tianjin University of Science and Technology, China)

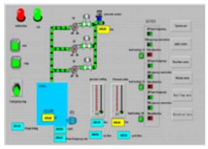
Burrows-Wheeler transform (BWT) is a commonly used transform in compression or text comparison. For example, in bzip2, BWT is used to preprocess the original data, then the same characters in the original data are close to each other, which improves the compression rate. Because the prefix tree of the original string can be easily obtained from the result of the BWT, BWT is also applied to the search and comparison of strings. For instance, the comparison of DNA sequences uses the BWT algorithm. However, BWT is not a fast algorithm, only Tens of megabytes per second on CPU. This article uses the GPU to sort the original string by the base of the 4-byte key size radix sort. After radix sort, the part with insufficient length is sorted again to complete the BWT algorithm.



OS13-3 Design of Automatic Water Supply Upper Computer System

Peng Lu, Fengzhi Dai, Tianyi Zhang (Tianjin University of Science and Technology, China)

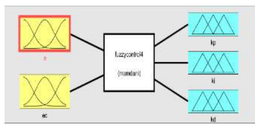
Aiming at the unstable water supply and fluctuating water pressure in the water supply system, an upper computer system of automatic water supply system based on Siemens PLC and KingView software was proposed. The pressure sensor in the water supply pipeline is used to detect the pressure in the pipeline, and the liquid level sensor monitors the liquid level in the tank. The sensor transmits the data to the PLC, and the PLC issues the control instruction after the computation processing. KingView software can realize real-time monitoring and fault alarm of the system. The system can effectively improve the stability of water supply, and can avoid human entering the dangerous environment to search for unknown faults.



OS13-4 Analysis of Boiler Water Level System based on the Fuzzy Control

Tianyi Zhang, Fengzhi Dai, Peng Lu (Tianjin University of Science and Technology, China)

Based on the three-stroke water supply system, this paper analyzes the performance characteristics of the boiler water level control system in reality, so as to reason out the appropriate fuzzy control rules, design fuzzy controller, and applied to the control system, so that the system for self-adjustment of PID parameters, constitute a fuzzy PID control system. On this basis, this paper analyzes the performance, advantages and characteristics of two control systems: the traditional PID control system and the fuzzy PID control system, and simulates the parameters of the input variables for comparison and analysis.



OS13-5 Development of the Circuit System for Greenhouse Environment Regulation

Yuhui Cheng, Fengzhi Dai, Chengxu Ji, Peng Lu
(Tianjin University of Science and Technology, China)

Aiming at the demand of agricultural modernization, this paper proposes a greenhouse environment regulation system based on single-chip microcomputer. For the greenhouse, due to the closed space and the requirements of the crops for the stability of the growth environment, the cheap single-chip control system can be used to complete this work well. The system can obtain the current environmental data in the greenhouse through the DHT11 temperature and humidity sensor and the light sensor. The system can keep the environment in the greenhouse in a proper state without people, and is suitable for large-scale greenhouse planting.

OS13-6 Design of an Intelligent Car based on MSP430

Ruming Kang, Fengzhi Dai (Tianjin University of Science and Technology, China)

The system takes the MSP430 single chip microcomputer as the control core, uses the reflective photoelectric sensor TCRT5000 module to track the line, and realizes the automatic tracking. In the experiment, the black-and-white line is used as the route. The system is driven by the L298N module and controlled by the PWM DC motor. This paper introduces the principle of the reflective photoelectric sensor and the circuit diagram of the tracking module. It also explains how to realize the automatic tracking based on the MSP430 single chip microcomputer. The technology can be used in warehouse, unmanned production line, intelligent service robot and other fields.

OS13-7 Design of a WIFI Video Car

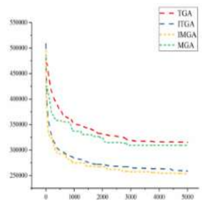
Qianqian Zhang[1], Fengzhi Dai[1,2], Jichao Zhao[1], Haokang Wen[1], Hongbo Hao[1]
([1]Tianjin University of Science and Technology, China)
([2]Tianjin Tianke Intelligent and Manufacture Technology CO., LTD, China)

This subject is mainly composed of motor, frame, STC89C51 single-chip microcomputer and other auxiliary modules. Through the android client broadcast video car cameras, wireless video signal acquisition by the client to send commands to the car at the same time, wireless router transmission instruction by wireless router, single-chip microcomputer for processing, then driven by single chip microcomputer control motor, drive motor rotation, steering device, so as to realize the control of motor sports and video acquisition.

OS13-8 Design of WebGIS Transportation and Distribution System based on the Genetic Algorithm

Hongbo Hao, Fengzhi Dai (Tianjin University of Science and Technology, China)

With the concept of "smart logistics" put forward, the transformation of the logistics industry facing information technology is facing great challenges. At present, China's logistics industry is facing problems such as extensive resource allocation, low management level and high cost. Aiming at the transportation problem in logistics industry, this paper designs a solution method of balanced / unbalanced transportation problem based on Monte Carlo similarity and the genetic algorithm. Two coding methods, Prufer number and matrix are adopted. On this basis, dynamic mutation rate and random mutation strategy are designed, and Monte Carlo similarity receiving method is introduced. Finally, from the perspective of system requirements, the WebGIS transportation and distribution system based on the genetic algorithm is designed and developed.



OS13-9 A Study of YOLO Algorithm for Target Detection

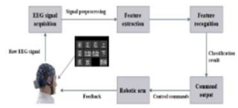
Haokang Wen, Fengzhi Dai, Yasheng Yuan
(Tianjin University of Science and Technology, China)

With the development of deep learning, target detection has become one of the research directions of many scholars. As one of the more mature algorithms, the YOLO series of algorithms have been widely used in real life. Combining the development history of the YOLO algorithm, this article focuses on the main framework and main content of the current latest YOLOv5 algorithm, and uses the YOLOv5 model to identify and detect footballs. This article evaluates its detection effect. The test results show that YOLOv5 has a wider application meaning in real life.

OS13-10 Research on Recognition and Application of EEG Signal based on SSVEP-BCI

Di Yin[1], Fengzhi Dai[1,3], Mengqi Yin[2], Yasheng Yuan[1], Yuxuan Zhu[1]
([1]Tianjin University of Science and Technology, China)
([2]Hebei University of Chinese Medicine, China)
([3]Tianjin Tianke Intelligent and Manufacture Technology CO., LTD, China)

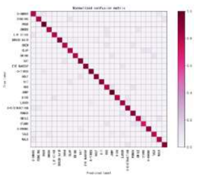
In recent years, brain-computer interface systems based on steady-state visual evoked potentials (SSVEP) have attracted attention due to their high information transfer rate (ITR) and more and more targets. The current mainstream algorithms for SSVEP recognition have greatly improved the accuracy and target detection time. This paper designs a robotic arm application system based on the eCCA-Y method for multi-target recognition. The phase characteristics of CCA's sine and cosine signals are added to the EEG signal. Compared with mainstream algorithms, research shows that this method can improve the SSVEP-based BCI performance. And choose a six-degree-of-freedom manipulator as the actuator of the brain-computer interface, and use a phase-encoded stimulation paradigm for multi-target recognition to conduct experiments on the application of the proposed method.



OS13-11 Research on Bad Driving Detection based on Behavior Recognition

Yasheng Yuan, Fengzhi Dai, Di Yin, Yuxuan Zhu
(Tianjin University of Science and Technology, China)

Dangerous driving behavior is considered to be the direct or indirect reason of road accidents. Although artificial video surveillance is good to prevent bad driving, it wastes too much time and manpower. How to effectively identify behavior become the focus of the research. In recent years, deep learning showed the huge advantage in the field of computer vision. This paper adopt a number of deep learning network models, mining video integration of space and time features, introduction of analogy in human visual attention mechanism, improve the model deeply, using the LSTM to accurate and efficient video behavior analysis technology.



OS13-12 Research on Crop Image Recognition Technology - Take Daylily as an Example

Jichao Zhao, Fengzhi Dai (Tianjin University of Science and Technology, China)

Digital image recognition technology is the core technology of agricultural robots. The key research content is to judge the maturity of crops through image recognition. In this paper, the daylily is used as the crop to be identified, and the digital image processing technology is used to separate the image of daylily from the background, and then send the processed information to the picking actuator of the robot to assist the actuator in picking tasks. This method can be used not only in the recognition of daylily, but also in the recognition of tomatoes, cucumbers and other crops by adjusting the recognition pictures and recognition parameters. It has high scalability.

OS13-13 Design of Daylily Agricultural Picking Robot

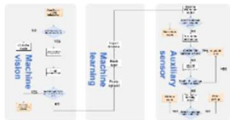
Jichao Zhao, Fengzhi Dai (Tianjin University of Science and Technology, China)

Agricultural picking activities occupy an important role in daily life, and the research and development of agricultural picking robots play a more crucial role and increasingly become an advantageous tool to improve agricultural productivity. Daylily, as a crop for daily consumption, can significantly reduce serum cholesterol and has extremely high nutritional value. However, the picking conditions of day lily are harsh, and long-term picking is likely to cause extremely high harm to the human body, especially the hands. This paper designs an agricultural picking robot that can pick daylily automatically, which can greatly free hands and improve labor productivity. Especially in health, it can significantly reduce the harm to the human body.

OS13-14 Development of Intelligent Public Trash Can based on Machine Vision and Learning

Longyu Gao, Fengzhi Dai, Zhiqing Xiao, Jiangyu Wu, Zilong Liu
(Tianjin University of Science and Technology, China)

At present, with the maturity of machine vision technology and the continuous expansion of application fields, there have been many intelligent trash cans based on machine vision, which can realize certain garbage identification and automatic classification. However, due to certain technical limitations of machine vision, it is impossible to identify all garbage. In this paper, a smart public trash can based on machine vision and auxiliary sensors is proposed. In addition to realizing machine vision to identify and automatically classify garbage, sensors will also be used to assist in identifying garbage to solve problems such as the same garbage classification of different shapes. At the same time, enhanced learning will be added to realize the self-learning of the trash can, so as to achieve the goal of continuously increasing identifiable types.



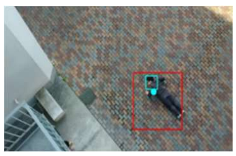
OS14 Image Processing (5)

Chair: Joo Kooi Tan (Kyushu Institute of Technology, Japan)
Co-Chair: Seiji Ishikawa (Kyushu Institute of Technology, Japan)

OS14-1 Detection of a Fallen Person and Estimation of Their Head Position from UAV Images

Haruka Egawa, Seiji Ishikawa, Joo Kooi Tan (Kyushu Institute of Technology, Japan)

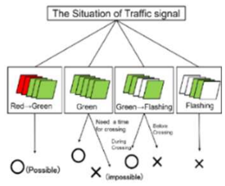
In recent years, aerial photography has often been used for finding victims in the event of a disaster. Searching from the sky enables quick rescue activities in places that are difficult for human to approach. In this paper, we propose a method of detecting a person fallen on the ground from the images taken by a camera mounted on a UAV(multicopter). Unlike pedestrians, the head orientation of a fallen person in an image is not identical. Therefore, in the proposes method, combination of Ri-HOG features and Ri-LBP features are employed for representing a fallen person, and the fallen person is detected by a classifier constructed using Random Forest. In addition, a head position of a fallen person is estimated by using the peak of the gradient direction histogram. The effectiveness of the proposed method was verified by experiments.



OS14-2 Development of a Pedestrian Crossing Navigation System for Visually Impaired People Using MY VISION

Kohei Kitagawa, Seiji Ishikawa, JooKooi Tan (Kyushu Institute of Technology, Japan)

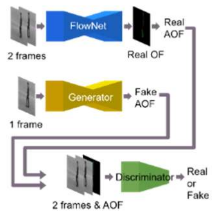
In this paper, we propose a system for a visually impaired person to cross a pedestrian crossing safely by the employment of the state of a traffic light and the remaining distance to the other side obtained from MY VISION images. The traffic light at a pedestrian crossing is detected by a discriminator using HOG (Histograms of Oriented Gradients) features and Random Forest. The possible actions for a visually impaired person at a pedestrian crossing are classified into six types of action patterns by analyzing the traffic lights and the distance information. The proposed system chooses one of the six patterns so that the visually impaired person may cross the pedestrian crossing safely. The effectiveness of the proposed method was verified by experiments.



OS14-3 Abnormal Human Action Detection Based on GAN

Tomoya Sano, Joo Kooi Tan (Kyushu Institute of Technology, Japan)

Recently, surveillance systems using cameras have been widely used, according to frequent outbreak of crimes. One of the important roles of such a camera surveillance system is to detect abnormal human actions or events. In this paper, we propose a method of abnormal human actions/events detection using Generative Adversarial Nets (GAN). In anomaly action detection, the main problem is that the image data of abnormal human actions/events is more difficult to obtain than normal human actions/events. To solve this difficulty, we use only normal human action data in the employed training networks. The GAN architecture trains two deep networks which compete with each other; a generator that performs conversion of Pixel-to-Pixel and a discriminator that supports it. In the experiment, a single class classification using KTH dataset and anomaly detection using Avenue dataset were performed for detecting abnormal human actions and anomaly events, respectively. Our experimental results show the effectiveness of the proposed method in detecting abnormal human actions and events.



OS14-4 3-D Position Measurement of a Cargo Using Epipolar Geometry for Logistics Automation(withdraw)

Kazuki Fukuda, Joo Kooi Tan (Kyushu Institute of Technology, Japan)

In recent years, with the expansion of the sales market such as online shopping, labor shortage has become a problem in logistics industries. Logistics is roughly composed of five tasks, in which transport vehicles such as forklifts are mainly used for cargo storage and cargo handling work. Since the operations of forklifts are still performed manually in many factories, automating the operations is strongly requested. This problem can be solved by the acquisition of the 3-D position information of the cargo. In this paper, we propose a method of estimating the 3-D position of a cargo from the images obtained from (i) two fixed cameras and (ii) a single camera mounted on a transport vehicle for measuring the relative distance between the transport vehicle and the cargo and for position control of the vehicle. The effectiveness of the proposed method is shown by experiments.

OS14-5 Bus Line Number Detection Employing MY VISION

Ye Zhou, Yosiki Hamasaki, Joo Kooi Tan (Kyushu Institute of Technology, Japan)

In response to the requests from visually impaired people that they wish higher QoL in their daily activities, and to the facts that they find difficulty in using public transportation, we propose we propose a bus boarding support system using MY VISION. For the support, we detect a bus approaching a bus stop from MY VISION images employing the characteristics of Haar-like filters, and extract the frontal upper area displaying bus line number by random forest. After having obtained the bus line number area, we use Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance the LED light of the area, and then recognize the line number using template matching. The effectiveness of the proposed method is shown by experiments.

OS15 Bridging the Gap Between AI, Cognitive Science, and Narratology (7)

Chair: Jumpei Ono (Aomori University, Japan)
Co-Chair: Hiroki Fukushima (Kyushu Womens' University, Japan)
Co-Chair: Takashi Ogata (Iwate Prefectural University, Japan)

OS15-1 Analysis and Construction of Elements of the Stage Performance Structure in a Kabuki-dance

Miku Kawai (Iwate Prefectural University, Japan),
Jumpei Ono (Aomori University, Japan),
Takashi Ogata (Iwate Prefectural University, Japan)

We conducted a detailed analysis of the stage performance structure of Kyōganoko Musume Dōjōji and reproduced it on the animation tool KOSERUBE with the music, lyrics, and images aligned. In this paper, 11 scenes of Kyōganoko Musume Dōjōji are organized. The actual flow is compared with the movement on the system. The goal of this paper is to understand the entire stage structure of Kyōganoko Musume Dōjōji in detail. In addition, this research aims to use the stage performance structural system for an integrated narrative system that the authors are developing. We believe that this system will be the basis for the construction of the Kabuki stage. Further, we want to apply it not only to Kabuki but also to other genres, such as computer games and commercials.

OS15-2 Unchiku Generation with Moving Illustration Using Kabuki Knowledge

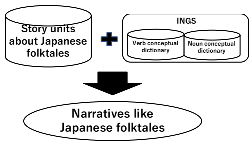
Jumpei Ono (Aomori University, Japan),
Miku Kawai (Iwate Prefectural University, Japan),
Takashi Ogata (Iwate Prefectural University, Japan)

In this study, we combined the unchiku generation system and an animation system, and added the animation system as a moving illustration. We have been studying the unchiku generation system and animation system for a long time. The first system is an unchiku generation system. Unchiku is the unchiku of deeply preserved knowledge. Unchiku in the story is one of the rhetorical techniques of the story. The second system is an animation system that expresses the stage performance structure. This system is based on the Kabuki-dance Kyōganoko Musume Dōjōji. In this study, we created a prototype and considered a new system that combines both the systems. The system provides an animated description for the generated story. Its purpose is to provide an explanation that deepens the understanding of the story.

OS15-3 Implementing Story Units of Japanese Folktales with Conceptual Dictionaries

Takuya Ito (Iwate Prefectural University, Japan),
Jumpei Ono (Aomori University, Tokyo)
Takashi Ogata (Iwate Prefectural University, Japan)

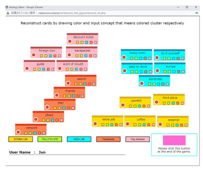
This paper reports the development of story units that are knowledge representing story structure for story generation. A story structure in our narrative generation study is formed by a tree that consists of events connected based on the relationships, such as the causal and temporal relations of events. We have been implemented story units based on the “Type of Folktale” in the *Nihon Mukashi-banashi Taisei* (*The Concluded Compilation of Japanese Folktales*). The current main theme is to use our conceptual dictionaries, including verb and noun conceptual dictionaries, to combine the current version of story units to our “integrated narrative generation system,” which is a general framework of our narrative generation study. In this paper, we explain the above mechanism.



OS15-4 The Story Generation Process and Cognitive Biases

Jun Nakamura (Chuo University, Japan)

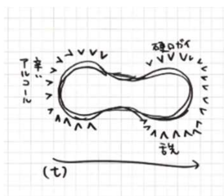
When it comes to designing a product, a service, or a new technology, whatever the case may be, author believes that in order to attract people, it is important to tell a story about the subject matter in a narrative style, that makes people impressed with the subject matter. In this paper, author asked several groups to create stories based on the same given data set. As expected, a variety of stories were created for each group. Cognitive biases were evident in the process of creating the stories. It is assumed that cognitive bias is a sort of driving force behind the creator's intentions, then it can be interpreted that the creator must have a certain strong intention to create a story.



OS15-5 Time in an Aesthetic Experience of a cup of Sake

Hiroki Fukushima (Kyushu Womens’ University, Japan)

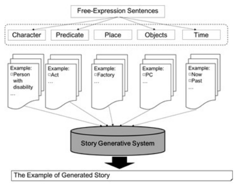
There are two broad genres of artworks depending on the nature of their appreciation: arts of space and arts of time. In the aesthetic appreciation of taste, which is the focus of the experience of a cup of Sake? Is there the dimension of time in the experience of taste, and if so, what is it? Because taste has never been considered a formal subject of aesthetics, there is little discussion of describing and appreciating the aesthetic experience. This study deals with the concept of time in the appreciation of Sake, with some reference to the concept in narrative studies. As with wine, the time associated with Sake is multifaceted, and each is intrinsic to the aesthetic experience. Consider some examples: Time from the Sake is brewed to the time it is opened on the table (so-called "vintage"; for some years). The time from opening the bottle to pour into the glass (for minutes to tens of minutes). The time between pouring into the glass and entering the mouth (for a few minutes). Time to swallow in the mouth (a few seconds). The flavor felt after swallowing ("after flavor"; tens of seconds). This study focuses on the period from the Sake is placed in the mouth to the time it is swallowed. The purpose of this study is to reveal how time is narrated (explicitly or metaphorically) in the tens of seconds of a sake tasting.



OS15-6 Extension of Clinical/Psychological Approach Using Post-Narratology: Possibility of application on Artificial Intelligence and Robot

Kai Seino (National Rehabilitation Center for Persons with Disabilities, Japan)

A purpose of this resarch is to examine the theories and methods of the psychological approach in treatment and research from a narrative viewpoint. In addition, we discuss the possibility of expanding collaboration with AI and Robot. This paper provides a literature review of the narrative psychological approaches in the fields of psychology and psychiatry. It addresses not only narrative therapy but also related therapeutic approaches that use the narrative framework. And Psychotherapy and Psychological, Practice and research of Self-help group, Open Dialogue, Tohjisha-Kenkyu (Self-directed studies), etc. are outlined, and the possibility of integration to A.I. Artificial Intelligence or application on Robots is proposed. In addition, the actual research which aimed at integration to A.I. Artificial Intelligence is reviewed. In discussions. First, literary narratology as basic theory is compared to psychological narratology as a practice or application theory. Second, the extension and possibility of a psychological narratology are considered.



OS15-7 A relationship between narratology and marketing

Akinori Abe (Chiba university, Japan)

We have been studying an expression of the taste of Japanese sake. Actually, the descriptions of the taste can be regarded as stories. Accoringly, we conducted an experiment, in which we asked participants to draw descriptions of the taste of the Japanese sakes and to design labels of them. From the results of the experiment, we will analyse the relationship between narratology and marketing of Japanese sake.

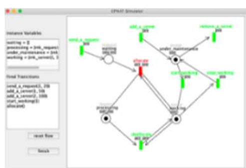
OS16 Software Development Support Method (5)

Chair: Tetsuro Katayama (University of Miyazaki, Japan)
Co-Chair: Tomohiko Takagi (Kagawa University, Japan)

OS16-1 Simulation and Regression Testing for Behavior of Software Models Based on Extended Place/Transition Net with Attributed Tokens

Tomohiko Takagi, Ryo Kurozumi (Kagawa University, Japan)

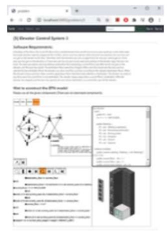
Extended Place/transition Net with Attributed Tokens (EPNAT) is a formal specification description language for modeling the expected behavior of state transition-based software that consists of multiple objects. Each attributed token corresponds to an object, and the values of its variables are updated while moving on the net. Engineers need to understand such complex aspects of EPNAT models when constructing, validating and refining them. In order to address this problem, we propose a simulation and regression testing technique for the behavior of software EPNAT models represent, and then show an early prototype tool to support it. In the tool, fireable transitions and values of variables are indicated for the simulation, and also the process of the simulation is recorded as test cases for the regression testing.



OS16-2 Development of an Early Prototype Tool for Learning Software Modeling Using Extended Place/Transition Net

Tomohiko Takagi, Akio Usuda (Kagawa University, Japan)

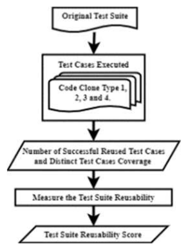
Extended Place/transition Net (EPN) is place/transition net that includes some additional elements written in VDM++ to enhance its representation power, and can be used to model the state transition-based behavior of software in development processes. The use of EPN is based on technical knowledge and skills, and therefore engineers will need to learn them. This paper shows an early prototype tool for learning software modeling using EPN, and then gives the discussion about its effectiveness and problems. A user of the tool, that is, an engineer tries to construct his/her EPN model based on given software requirements by selecting and putting the components of EPN. The behavior of the EPN model is partially visualized by using animated graphics, and finally its correctness is automatically checked.



OS16-3 Test Suite Reusability Measurement Based on Frequency and Coverage of Reused Test Cases

Mochamad Chandra Saputra*, Tetsuro Katayama*, Yoshihiro Kita†,
Hisaaki Yamaba*, Kentaro Aburada*, Naonobu Okazaki*
(*University of Miyazaki, Japan) (†University of Nagasaki, Japan)

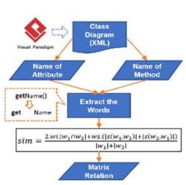
A test suite that consists of a set of test cases is an important element in software testing. The objectives of white-box testing process are to execute all the test cases in the test suite and also achieve 100% code coverage. One of the good criteria for test suite is possible to reuse it for testing another program. Reusability of a test suite is one of important factors to decrease the cost and time in testing the program. This research defines and measures the test suite reusability score expressed with the degree of reusability of the test suite. It is calculated by considering both frequency and code coverage of successful reused in the test suite. Test suite reusability measurement provides useful information to improve the efficiency of software testing, especially in regression testing and automated testing.



OS16-4 The Seven Information Features of Class for Blob and Feature Envy Smell Detection in a Class Diagram

Bayu Priyambadha*, Tetsuro Katayama*, Yoshihiro Kita†,
Hisaaki Yamaba*, Kentaro Aburada*, Naonobu Okazaki*
(*University of Miyazaki, Japan) (†University of Nagasaki, Japan)

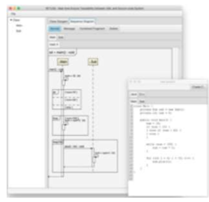
Measuring the quality of software design artifacts is difficult due to the limitation of information in the design phase. The class diagram is one of the design artifacts produced during the design phase. The syntactic and semantic information in the class is important to consider in the measurement process. The class information is used to detect the smell as an indicator of a lack of quality. All information related to the class is used by several classifiers to prove how informative it to be used to detect the smell. The smell types that are a concern in this research are Blob and Feature Envy. The experiment using three classifiers (j48, Multi-Layer Perceptron, and Naïve Bayes) confirms that the information can be used to detect the smell. The average of correctly detected by the classifiers is about 80.67%.



OS16-5 Improvement of RETUSS to Ensure Traceability between Sequence Diagram in UML and Java Source Code in Real Time

Kaoru Arima*, Tetsuro Katayama*, Yoshihiro Kita†,
Hisaaki Yamaba*, Kentaro Aburada*, Naonobu Okazaki*
(*University of Miyazaki, Japan) (†University of Nagasaki, Japan)

Ensuring traceability of software deliverables is one of the methods to ensure software quality. It has two problems: taking much labor and time, and causing mistakes by human handling. In order to solve them, our laboratory developed RETUSS (Real-time Ensure Traceability between UML and Source-code System). RETUSS ensures traceability between UML and source code by transforming them to each other in real time. However, RETUSS is not useful in ensuring traceability between sequence diagram and Java source code due to its limited scope of application. This paper improves the usefulness of RETUSS by extending the scope of application of sequence diagram and Java source code. The experimental results confirmed that the improved RETUSS also saves labor and time, and eliminates mistakes by human handling.



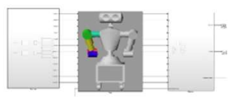
OS17 Robotics and machine vision (3)

Chair: Jiwu Wang (Beijing Jiaotong University, China)
Co-Chair: Junxiang Xu (Beijing Jiaotong University, China)

OS17-1 Kinematic modeling and Simulation of humanoid dual-arm robot

Jiwu Wang, Junxiang Xu (Beijing Jiaotong University, China)

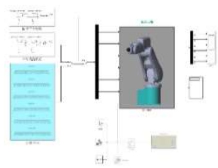
In recent years, dual-arm robots have attracted more and more attention due to their advantages such as strong cooperation ability and high flexibility. With the improvement of real-time requirement, the inverse kinematics solution of robot becomes a key problem to be solved urgently. In order to solve the time-consuming problem of inverse kinematics of robot arm, a closed inverse kinematics solution algorithm for humanoid dual-arm robot was proposed. The effectiveness of the algorithm was verified by simulation.



OS17-2 Kinematics analysis and simulation of 6R Robot Based on MATLAB/Simulink

Jiwu Wang, Shuo Han (Beijing Jiaotong University, China)

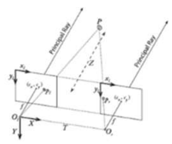
The intelligent robot has played an important role in improving working conditions and production efficiency with the development of intelligent robot technology. Here the robot kinematics was studied and simulated based on the MATLAB/Simulink. The forward and inverse kinematics equation were established based on the standard Denavit-Hartenberg method, and eight solutions regarding angle variables of robot joints were obtained. Then the motion trajectory of the robot was simulated in the joint space, and the smooth curves of angular displacement, angular velocity. Furthermore, the visualization of the robot kinematics model was realized based on the Simulink, and the kinematics simulation control system was established. The paper provides necessary theoretical basis and correct kinematics model for the research on kinetic analysis of 6R robots.



OS17-3 Distance measurement system based on binocular stereo vision

Jiwu Wang, Xin Pei (Beijing Jiaotong University, China)

In order to solve the problem of low precision and time consuming in traditional ranging, this paper studies key technologies such as binocular ranging principle, camera calibration, and stereo matching.On this basis, a planar ranging system based on parallel binocular vision is designed. Firstly, through the stereo camera calibration toolbox of MATLAB, the internal and external parameters are obtained. Then stereo correction is used to make the two images coplanar and aligned by line and performing stereo matching on the corrected image to obtain the parallax. Finally, it is able to select any two target points through the mouse to obtain their spatial coordinates and calculate the distance between the two points. Experimental results show that the method meets the precision requirements.



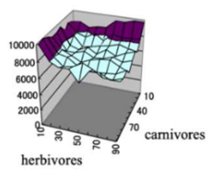
OS18 Natural Computing (3)

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

OS18-1 Contribution to the Theory of Periodic Reaction of Three Bodies Systems

Yasuhiro Suzuki (Nagoya University Japan)

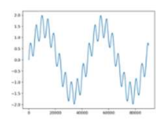
A chemical ecosystem is an ecosystem in which chemical substances mediate interactions. When herbivores feed, plants analyze herbivore saliva and produce volatile chemicals that attract herbivores' natural enemies. Natural enemies are attracted to volatile chemicals and eliminate herbivores. This system is a trytrophic system consisting of herbivores, carnivores and volatile chemicals. The basic equation of mathematical ecology is the Lotoka Volterra equation, which is a bitrophic system of herbivores and carnivores. This paper proposes the fundamental equations of chemical ecosystems and shows their mathematical features.



OS18-2 Deep Micro Vibrotactile, DMV and its Applications

Yasuhiro Suzuki (Nagoya University, Japan)

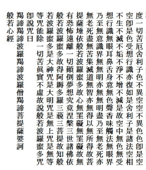
Deep Micro Vibrotactile, DMV is infrasound. DMVs produce sound effects when mixed with audio. Since the DMV has a transcendental low frequency, mixing it with sound such as music does not affect the frequency spectrum in the audible range. However, through demonstration experiments at concerts, it was confirmed that DMV changes the tactile qualities of the sound. This paper introduces the acoustic features and case studies of DMVs



OS18-3 Inter-Induce computation and its Philosophical Foundation

Yasuhiro Suzuki (Nagoya University, Japan)

Set theory is based on the distinguishability of elements. How to recognize and identify the world is the essence of set theory. If each element cannot be identified, all the elements are one set. So the set does not make sense. The Heart Sutra is highly rational and can be interpreted mathematically. The mathematical interpretation of the Heart Sutra shows the divergence of how to discriminate. Based on this world view of Heart Sutra, we propose Inter-Induce computation, IIC as a novel calculation paradigm that does not depend on set theory. This paper gives an overview and philosophical foundation of IIC.



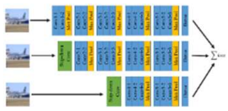
OS19 Artificial Intelligence for Embedded Systems and Robotics (4)

Chair: Hakaru Tamukoh (Kyushu Institute of Technology, Japan)
Co-Chair: Sansei Hori (Kyushu Institute of Technology, Japan)

OS19-1 Convolutional Network with Sub-Networks

Ninnart Fuengfusin, Hakaru Tamukoh (Kyushu Institute of Technology, Japan)

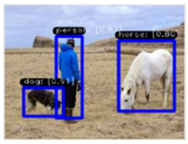
We propose a convolutional network with sub-networks (CNSN), a convolutional neural network (CNN) that can be detached in a depth-wise direction into sub-models. Due to a conventional design of CNN which feature map shapes are varied throughout the model, a hidden layer may not able to receive an input image without changing the last fully-connected layer. To deal with this problem, we attach a step-down convolutional layer as an additional layer to each sub-model. The step-down layer acts as an input layer of the sub-model which reduces input to the preferred representation to the sub-model. To train CNSN, we treat the base- and sub-models as separable models and by using multi-losses, a combination of losses from the base- and sub-model, we can update weights to able to be utilized to both base- and sub-models.



OS19-2 Influence of FPGA implementation methods in High-Level Synthesis

[1]Yusuke Watanabe, [2]Hakaru Tamukoh ([1]RAFT WORK Co., Ltd, Japan, [2]Kyushu Institute of Technology, Japan)

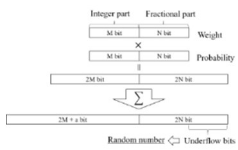
We explain about how the difference of implementation methods written in C++ in High-Level Synthsis(HLS) influences on time it takes for tiny You only look once(YOLO) v2, a real-time object detection system to infer on an FPGA. To utilize features of FPGA, we need to implement hardware-oriented algorithms such as the weight binalization. We primarily focus on convolution in tiny YOLO v2 network and we report execution results on the Xilinx SDSoC development environment to know whether methods are appropriate or not.



OS19-3 A Hardware-Oriented Random Number Generation Method and A Verification System for FPGA

Sansei Hori, Hakaru Tamukoh (Kyushu Institute of Technology, Japan)

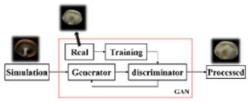
We propose a new random number generation method for digital hardware such as an FPGA. A random number generator is one of the important components to implement neural networks (NNs) such as a generative model on FPGA. However, conventional methods of random number generators (RNGs) require huge hardware resources. The proposed method employs underflow bits obtained during calculations of NN on the FPGA instead of the random number from the conventional RNGs to reduce the hardware resources. In this paper, we employed the proposed method to a restricted Boltzmann machine to evaluate the method and trained MNIST dataset on a software environment. Furthermore, this paper proposes an FPGA infrastructure that can easily connect NNs to a host computer.



OS19-4 Synthesis of realistic food dataset using generative adversarial network (GAN) based on RGB and depth images

Obada Al Aama, Hakaru Tamukoh (Kyushu Institute of Technology, Japan)

Constructing a large food dataset is time and effort consuming due to the need to cover the feature variations of food items. Hence, a huge data is needed for training neural networks for robot vision systems. Current study aims to advocate the Cycle-GAN to build up large food dataset based on large number of simulated images and relatively few real captured images thus obtaining more realistic images effortlessly compared with traditional capturing. Real RGB and depth images of variant food samples allocated over turntable were captured in three different angles using real-sense depth camera with different backgrounds. Furthermore, for simulated images, the Autodesk 3D_Maya software was employed using the same parameters of captured real images. Results showed that generally, realistic style transfer on simulated food objects was obtained as a result of employing Cycle-GAN. GAN proved to be an efficient tool that could minimize imaging efforts resulting in realistic images.



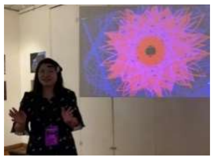
OS20 Virtual Reality and Intelligent Interactions (4)

Chair: R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)
Co-Chair: Yoshimasa Tokuyama (Tokyo Polytechnic University, Japan)

OS20-1 NeuroPhyllotaxis: An Interactive Application for Generative Art Based on EEG Data

Chien-Tung Lin, R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)
Yoshimasa Tokuyama (Tokyo Polytechnic University, Japan)

Generative art is produced by procedural techniques. It has obtained a lot of attention from the beginning of computer graphics. Many works of art are inspired by nature, among which phyllotaxis is more It is a combination of mathematics and the beauty of nature. Not only can it be seen everywhere in nature, but also often appear in man-made objects, becoming part of culture or religion. This paper presents the development of an interactive generative art application that is created from a phyllotaxis pattern by using the user's EEG data. When people are using it, it is easy to relax the mood and achieve the function of art therapy. We tried to use EEG data to make an interactive installation art which creates phyllotaxis patterns and project them on the wall. Everyone has a different state, the generated patterns are also different from person to person, which creates interesting interactive contents. In addition, sound can also be changed by EEG to become dynamic and real-time contents.



OS20-2 VRMAZU: VR Visualization of Mazu Temple for Passive Interaction with Generated Sound from the ML Technique(withdraw)

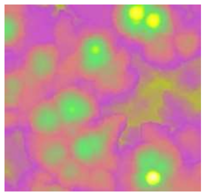
Yi-Li Liang. R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)
Jen-Tun Lee (Japan Advanced Institute of Science and Technology, Japan)
Yoshimasa Tokuyama (Tokyo Polytechnic University, Japan)

Passive interaction based on virtual reality applications has the potential to identify personalized user activities. Through personalized interaction methods to access the virtual scene, the user gets a better sense of immersion. This paper introduces a VR visualization of Taiwan traditional Mazu Temple for passive Interaction with generated sound from the machine learning (ML) technique. In this VR application, user can control their own situation through the feedback information of the generated sound based on the user's location in the VR scene, so that their physical condition can reach a good state. This development that personalized interaction in VR applications is very important to get different experiences in each time of the user's temple activities, so an ML approach to design sound ambiance is needed. This paper also concerns the different perspectives on understanding Taiwan traditional temple activity knowledge and user interaction.

OS20-3 ThoughtMix: Interactive Water Color Generation and Mixing Based on EEG Data

R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)
Yoshimasa Tokuyama (Tokyo Polytechnic University, Japan)

In modern art, new types of installations and artworks use sensor-based inputs as an interactive method from which to create a new form of media art. Most of those installation artworks have static and pre-captured contents; they tend to be played as dynamic, and by some concerned as interactive creations. However, most of the installations were not be able to capture the visitor's rich feedback which could trigger the contents. In order to identify the relation between the cybernetic-aesthetics and visitor's brain wave activities, we propose an electroencephalogram (EEG) data-driven approach to generated watercolor effects and mixing application which visualizes the changes in brain wave activities in real-time. Our proposed application transforms the brain waves into real-time watercolor effects that are wrapped around by generating the data about emotion, attention, meditation, and neural mechanisms with EEG data. "ThoughtMix" creates watercolor effects as an immersive experience to influence creativity and art therapy by actually visualizing the colors user's mind creates.



OS20-4 HaptWarp: Soft Printable and Motion Sensible Game Controller

Jen-Tun Lee, Kazunori Miyata (Japan Advanced Institute of Science and Technology, Japan)
R.P.C. Janaka Rajapakse (Tainan National University of the Arts, Taiwan)

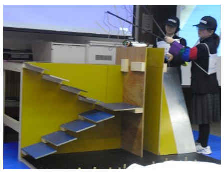
Many interaction techniques have been introduced for controlling virtual reality contents and navigation. Traditional controllers need to map between two-dimensional inputs and three-dimensional actions while interactive surfaces enable more natural approaches, they still lack tactile feedback. We present real-virtual bend and twist, a tactile feedback technology that delivers the twist and bends effect into a virtual environment. The method uses two inertial measurement unit (IMU) sensors fixed on both sides on the 3D printed soft material bar to acquire object deformation to simulate the virtual deformation based on the changes in the angle exerted on a physical device. People can customize the 3D printed connection shape to apply to different virtual scenarios.

OS21 Educational Application of Control Technology (4)
Chair: Kazuo Kawada (Hiroshima University, Japan)
Co-Chair: Yoshihiro Ohnishi (Ehime University, Japan)

OS21-1 Development of Shock Sensitive Tiny Dummy Robot for Junior High School Rescue Robot Challenge

Kazuo Kawada, Keisuke Iuchi, Keita Murai, Hiroyuki Y. Suzuki
(Hiroshima University, Japan)

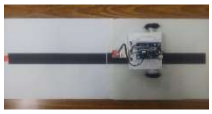
Junior High School Rescue Robot Challenge" is an annual tournament held by our group in Hiroshima Univ. by sponsorship of a construction machinery company. The tournament has 16 years of history and the rescue theme has been changed in each year. The theme in 2019 was to convey injured people "tenderly" from top of half collapsed building to ground, without giving any shocks. We developed shock sensitive tiny dummy robot, controlled by M5STACK microcomputer since it rigged with accelerometer. The dummy can remember how much shocks were given during transportation. The remote (wired) controlled robots developed by junior high school student teams conveyed the dummy and evaluated the "tenderness" during transportation.



OS21-2 Basic Research on Parameter Tuning Skills Evaluation Based on Sensor Car Behavior Data in Technology Education

Teruyuki Tamai (Hiroshima University / Ehime University), Yoshihiro Ohnishi (Ehime University), Kazuo Kawada (Hiroshima University)

In this research, the evaluation of programming included in the contents of junior high school technology education is explained. In that case, a new evaluation method using the sensor car as a teaching material is proposed. First, the guidance of programming in junior high school technology education is explained. In addition, the evaluation method of the programming which has been made so far is explained. Next, the new evaluation method is explained. Quantitative data obtained by operating the sensor car used as a teaching material will be described. And how to analyze the data will be described. Next, the method to evaluate based on the data is explained. Finally, the usefulness of the evaluation method is examined.



OS21-3 Mini Windmill Generator Kit for homework for Hiroshima Univ. Monozukuri Junior Doctor Special Educational Program

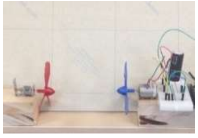
Hiroyuki Y. Suzuki, Kazuo Kawada, Masayasu Nagamatsu (Hiroshima University, Japan)

"Hiroshima Univ. Monozukuri Junior Doctor" is a special educational program for young (5-6 grades in elementary school and junior high school) volunteered students. We originally prepared around 10 seminars on various topics and completed 2/3 of them. We had to give up, however, all remained seminars because of sudden attack of COVID-19. Since then, we introduced a set of homework in which students could study by themselves. We modified commercially available "Mini Windmill Generator Kit" for the homework, to make easy to measure generated voltages with three color LEDs and a tester. Students tried to find best sizes and shapes of tail fin and windmill, using a manual of "Procedure of quantitative research" that we made. Majority of the students could find appropriate sizes and shapes of fins and windmills, through the data collection, drawing figures and by discussion.

OS21-4 Development of Teaching Materials to Learn the Efficient Use of Energy

Yoshihiro Ohnishi[1], Teruyuki Tamai[1], Shinnosuke Mori[1] Kawada Kazuo[2]
([1]Ehime University, [2]Hiroshima University, Japan)

This paper introduces the programming learning at junior high school. In this research, an experimental device using the fan which driven by EDLC was developed. Learning activities using this device is to maximum the driving time of fan by the charged energy on EDLC. The goal is to learn programming as a technique for efficiently using electrical energy.



OS22 Robot Competitions and Education (5)

Chair: Yasunori Takemura (Nishinippon Institute of Technology, Japan)
Co-Chair: Kazuo Ishii (Kyushu Institute of Technology, Japan)

OS22-1 Robustness Verification Against Noise of Self-localization Method Using Omni-directional Camera for Soccer Robot

Yuehang Ma, Kaori Watanabe, Hidekazu Suzuki (Tokyo Polytechnic University, Japan)

The focus of the RoboCup competitions is the game of soccer, where the research goals concern cooperative multi-robot and multi-agent systems in dynamic environments. In the field of RoboCup, self-localization technique is important to estimate own position including goal and other robot positions and to decide strategy. This paper presents a self-localization technique using an omni-directional camera for an autonomous soccer robot. We propose the self-localization method with white line information of soccer field, and recognizes the robot position by optimizing the fitness function using Genetic Algorithm. Moreover, we also verified the robustness of the proposed method against noise through experiments.

OS22-2 Tomato-Harvesting Robot Competition: Aims and Developed Robot of 6th Competitions

Takayuki Matsuo[1], Yasunori Takemura[2], Takashi Sonoda[2], Shinsuke Yasukawa[3], Yuya Nishida[3], Kazuo Ishii[3] ([1]National Institute of Technology, Kitakyushu College, Japan, [2]Nishinippon Institute of Technology, Japan, [3]Kyushu Institute of Technology, Japan)

Tomato is one of the important fruit vegetables and most tomatoes are produced in the greenhouses, or large-scale farms, where the high temperature and humidity, and long harvest age force the farmer heavy works. To develop the tomato harvesting robot, many research issues exist such as manipulator design, end-effector design, collaborative behavior, artificial intelligence, motor control, image processing, target recognition and so on. With an aim to promote the automation of tomato harvesting, we have organized the tomato harvesting robot competition since 2014. In this paper, we discuss the aim and the results of 6th tomato harvesting robot competition in 2019.

OS22-3 Smart Agriculture IoT Education Course in enPiT-everi (Education Network for Practical Information Technologies - Evolving and Empowering Regional Industries)

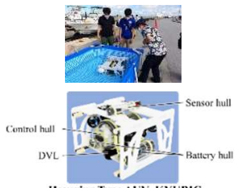
Yasunori Takemura[1, 2], Keiji Kamei[1, 2], Atsushi Sanada[1, 2], Kazuo Ishii[1]
([1]Kyushu Institute of Technology, Japan, [2]Nishinippon Institute of Technology, Japan)

Fostering human resources who can solve social issues by using Information Technology is an important issue in Japan. The Ministry of Education, Culture, Sports, Science and Technology (MEXT) has formed enPiT (Education Network for Practical Information Technologies) programs a practical education network for industry-academia collaboration and practice problem-based learning for the purpose. We have planned a smart agriculture IoT course in the program enPiT-everi (Evolving and Empowering Regional Industries) directed by Kitakyushu university. In the smart agriculture IoT course, the programming of IoT devices to measure the temperature of the green house and soil, soil moisture and CO_2, image processing using AI technology to detect maturity of fruits, green worms was carried out.

OS22-4 Development of a Handy Autonomous Underwater Vehicle “Kyubic”

Toshimune Matsumura, Yuuichiro Uemura, Kentaro Yanagise, Yoshiki Tanaka, Yuya Nishida, Kazuo Ishii
(Kyushu Institute of Technology, Japan)

Ocean is one of big challenging and extreme environments, and hard for human to access directly. As the tool for ocean survey, Autonomous Underwater Vehicles: AUVs are expected and developed from ‘80s. The recent rapid progress of computer and information technologies makes the development of AUVs easier and more practical. We had developed a handy AUV “Kyubic” for the observation of shallow water and artificial structures. In this paper, we describe the system architecture of Kyubic and the experimental results in Underwater Robotics Competition in Okinawa 2020.



Hovering Type AUV KYUBIC

OS22-5 A Greenhouse Project toward Smart Agriculture

Kazuo Ishii[1], Yuya Nishida[1], Shinsuke Yasukawa[1], Kanako Shirahashi[1], Yasunori Takemura[2], Takayuki Matsuo[3]
([1]Kyushu Institute of Technology, Japan, [2]Nishinippon Institute of Technology, Japan, [3]National Institute of Technology, Kitakyushu College, Japan,)

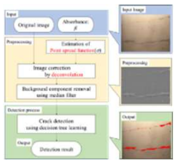
In the future estimation of 2050, the food demand increases 70% and the production decreases 15% caused by global warming, and farmer population be 1/5 of current workers. Until 2050, we need an agricultural system of twice production with the same farmland area and 5 times effective operations. To realize the sustainable society, smart agriculture including robot technology, AI, IoT is one of the solutions for food issues. We have been working for a greenhouse project for evaluation of robotic, AI and IoT technologies. In this paper, we introduce the project and robotic applications.

OS23 Advances in Field Robotics and Their Applications (7)
Chair: Keisuke Watanabe (Tokai University)
Co-Chair: Kazuo Ishii (Kyushu Institute of Technology)

OS23-1 Three-dimensional Measurement Using Laser Pattern And Its Application to Underwater Scanner

Yuya Nishida, Tomoya Shinnoki, Shinsuke Yasukawa, Kazuo Ishiii
(Kyushu Institute of Technology, Japan)

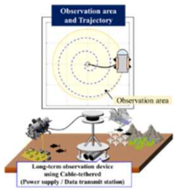
Underwater infrastructure and facilities are getting increasingly degraded and need inspections for cracks, scouring, however, the inspections are not carried out because of underwater environment and water turbidity. As a solution for the inspections, underwater vehicles attract attentions and expected to take precise photos of brides and their seabed. In this paper, to realize efficient underwater infrastructure inspection, automatic crack detection by image processing is proposed. In first process of our method generates enhanced image based on the absorbance from the turbidity meter and removes background component, and then detects crack from the enhanced image by using decision tree learning algorithm. This paper explains the algorithm of our method and shows evaluation experiment results.



OS23-2 Motion control of a cable-restricted underwater vehicle for long-term spot observation

Yoshiki Tanaka, Yuya Nishida, Kazuo Ishii (Kyushu Institute of Technology, Japan)

There are many marine resources in the Japanese EEZ, and recently, un-tethered underwater vehicles, AUVs (Autonomous Underwater Vehicles) are used as the practical tools for wide area seafloor survey. However, AUVs have the risks of their losing by hardware or software failures, AUVs have the risks to be trapped on the seafloor and not to be back on surface. The current trend of observation is 3D data in time series, 4 D data sampling. Underwater station can measure for long period in a certain point, not for wide area survey. This research aims to develop a long-term observation device with a low operation risk and wide area. The proposed system consists of a base station for supplying electric power and an underwater vehicle. The underwater vehicle is tethered to the base station with a cable. In this paper, we explain the developed observation device and report the result of motion control of the vehicle.



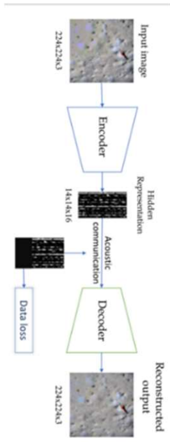
OS23-3 Development of current sensors for digitizing expert knowledge in fish feeding towards sustainable aquaculture

Dominic B. Solpico, Yuya Nishida, Kazuo Ishii (Kyushu Institute of Technology, Japan)

Improving the efficiency of fish feeding contributes to achieving sustainable expansion of the aquaculture industry. However, expert knowledge on efficient feeding remains a skill acquired through experience. This paper presents a new approach of digitizing such knowledge by measuring underwater currents induced by fishes as indicator of their behavior and appetite. Initial measurements to be made shall provide additional insight to the sensor system design. For this, a prototype current sensor suite has been constructed to measure the current around the fish cage, especially during feeding times. It consists of two modified flow sensors to measure the current flowing into or out of the cage at two target depths, and an IMU sensor to detect other induced water movements on the sensor mounting.

OS23-4 Underwater image reconstruction using convolutional auto-encoder

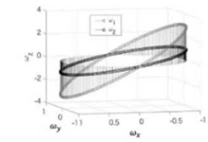
Shinsuke Yasukawa, Sreeraman Srinivasa Raghura, Yuya Nishida, Kazuo Ishii
(Kyushu Institute of Technology)



One of the main tasks of AUVs is to capture deep-sea images like fishes, crabs, other living organisms and resources for information leading to research on deep-sea ecosystems. Acoustic transmission are used to establish wireless underwater communications between the AUV and the ship. However, there are some limitations in the communication channels due to limited bandwidth, multi-path, temperature distribution and change in the direction of transmitting source and receiving sensor which results in losses in data being transmitted. Initially, the captured images are enhanced to reduce the effect of light attenuation and then compressed for transmission through acoustic modems. Only an important part of image is being transmitted through set of data packets. The received data packets in the ship will be reconstructed to predict the presence of living organisms. The loss in data during transmission creates a difficulty for the operators to predict the exact information. In this research, to compensate this transmission loss, an efficient compression and reconstruction technique using convolutional autoencoder with minimal distortion is proposed. Finally, for evaluation of the proposed image compression technique, the quality of reconstruction of images with and without data loss will be compared using the quality metrics signal to noise ratio(PSNR), structural similarity index(SSIM) and perceptual quality of image.

OS23-5 Spherical Robot Transfer Problem with Minimal Total Kinetic Energy

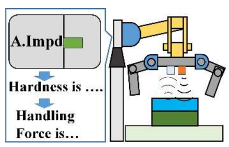
Kenji Kimura[1], Kazuo Ishii[2] ([1]Fukuoka Daiichi High School, Japan, [2]Kyushu Institute of Technology, Japan)



Previous spherical mobile robots were driven by two rollers with a fixed rotational axis, which restricts the angular velocity vector of the sphere to two dimensions. Three-dimensional freedom is expected to improve the rotational diversity of the sphere. This study proposes a spherical mobile robot with a variable roller-rotational axis that allows three-dimensional freedom of movement. Furthermore, the kinetic energy of transporting the sphere by the roller is minimized by an optimization procedure.

OS23-6 Acoustic Impedance Measurement through the Modelling of Ultrasonic Wave Transmission

Ryuugo Mochizuki, Yuya Nishida, Kazuo Ishii (Kyushu Institute of Technology, Japan)

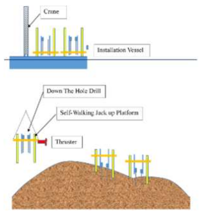


In food industry, shortage of workers is one of a serious problem. Automation of food handling is one of critical issues. To alleviate the damage caused by food picking operation by robotic hand, we propose non-contact acoustic impedance estimation with ultrasonic wave, which is preceded before the picking for optimization of picking force control. By assuming the correlation between hardness and acoustic impedance, the impedance is calculated by the product of sonic velocity and density of a medium. We derive an ultrasonic transmission model considering attenuation by ultrasonic reflection and absorption, then, made an experiment to estimate unevenness of the impedance. As the result, we succeeded in detecting the difference of acoustic impedance from the ultrasonic reflection.

OS23-7 Installation Method of In-situ Drilling Platform(withdraw)

Keisuke Watanabe (Tokai University, Japan)

Drastic cost reduction for subsea drilling is desired for scientific research and subsea mining field exploration. The traditional drilling method needs a special drilling platform which equips massive drilling facilities with many trained crews. Low cost in-situ drilling platform is one of the possible alternatives and we are studying about a self-walking jack up platform with a down the hole drill. In this paper, I focus on consideration on installation method to reduce its installation cost. The platform is suspended from a vessel and its motions in the horizontal plane are controlled by thrusters attached to the platform. The dynamics of a suspended platform with thrusters is formulated and a simulation program to estimate position accuracy is developed. Through simulations the effectiveness of the method is confirmed.



OS24 Mathematical Informatics (5)

Chair: Takao Ito (Hiroshima University, Japan)
Co-Chair: Makoto Sakamoto (University of Miyazaki, Japan)

OS24-1 Verification of CG character operation by brain wave discrimination

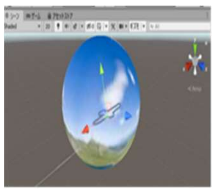
Kenji Sakoma[1], Kodai Miyamoto[1], Taketo Kamasaka[1], Makoto Sakamoto[1], Amane Takei[1],
Tsutomu Ito[2], Takao Ito[3] ([1]University of Miyazaki, Japan)
([2]National Institute of Technology, Ube College, Japan) ([3]Hiroshima University, Japan)

Since 2016, VR technology has been researched and developed all over the world and has made significant progress. In this research, I studied whether VR space could be played simply by thinking in my head, and conducted research. The equipment used at that time is an electroencephalograph, but since this equipment is generally very expensive, we used Neurosky's Mind Wave Mobile to determine if it could be done as cheaply as possible. This time, we used the fast Fourier transform and the support vector machine as the methods of EEG analysis. As a result, various problems were found. In the future, we will improve this problem and try to improve it further by changing the method of EEG analysis and other conditions.

OS24-2 Approach to tourism support by aerial video using CG

Kenji Sakoma[1], Noritake Seto[1], Kodai Miyamoto[1], Taketo Kamasaka[1], Makoto Sakamoto[1],
Amane Takei[1], Tsutomu Ito[2], Takao Ito[3] ([1]University of Miyazaki, Japan)
([2]National Institute of Technology, Ube College, Japan) ([3]Hiroshima University, Japan)

In recent years, rural areas have been declining due to the concentration of work in cities and the aging of the population. So, there are local governments that are focusing on tourism to regain the vitality of the past. As a first step in the tourism business, we will consider whether it is possible to increase the number of tourists by incorporating IT technologies such as CG in videos introducing PVs and commercials. This research is part of a joint development with Takachiho-cho.In this paper, we created a CG image of a bear flying in the sky combined with a real Takachiho video taken with a drone. However, various other problems have been identified, and I hope to improve them in future research to produce even better ones.



OS24-3 Method for detecting eye misalignment based on movement near the center of the pupil

Uchida Noriyuki[1,2], Kayoko Takatuka[2], Hisaaki Yamaba[2], Masayuki Mukunoki[2], Naonobu Okazaki[2]

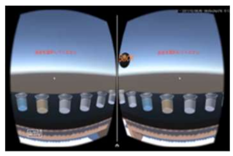
([1]Kyushu University of Health and Welfare, Japan), ([2]University of Miyazaki, Japan)

Only ophthalmologists and orthoptists are eligible to diagnose misalignment of the eyes amid persistent understaffing. We automated the traditional eye misalignment evaluation (Cover-test) in our earlier studies and developed a technique to distinguish the misaligned eyes from its behavior. This method, however, had a problem with the vertical eye movements being affected by the eyelids and eyelashes, which prevented thorough detection. Therefore, by observing only the movement near the center of the pupil, where it is less likely to be disturbed by the eyelids and eyelashes, we arranged an alternative method to recognize anomalies. Based on this process, we achieved an inspection system and carried out a performance experiment. We confirmed that we could observe not only horizontal motion, but also that of the vertical eye.

OS24-4 Basic research on video production for educational support by virtual technology

Kodai Miyamoto[1], Taketo Kamasaka[1], Kenji Sakoma[1], Makoto Sakamoto[1], Amane Takei[1], Tsutomu Ito[2], Takao Ito[3] ([1]University of Miyazaki, Japan) ([2]National Institute of Technology, Ube College, Japan) ([3]Hiroshima University, Japan)

As a result of conducting a questionnaire about science classes to high school students in 2016, the percentage of high school students who answered "I like science" and "Science is important" is lower than other subjects. However, more than 80% of elementary and junior high school students said they like experiments and observations. In addition, the 2019 smartphone penetration rate survey found that it is popular among about 90% of students. In addition, VR technology has recently made remarkable progress. From the above, I researched the idea that creating a simulation application using VR technology using smartphones would change the way high school students think about science classes. In this paper, we have developed a simulation application for science experiments. Subjects were asked to experience the newly created app and complete a questionnaire. As a result, the average score is 4 out of 5 and it is not bad. But at the same time, a problem was found. The problem was that this app was a simulation app, so there wasn't much user operability, so I wanted a little more operability. I want to make apps in other fields while improving the problem.



OS24-5 There is a movement towards the development of hula costume CAD

Taketo Kamasaka[1], Kenji Sakoma[1], Kodai Miyamoto[1], Makoto Sakamoto[1], Amane Takei[1], Tsutomu Ito[2], Takao Ito[3] ([1]University of Miyazaki, Japan), ([2]National Institute of Technology, Ube College, Japan) ([3]Hiroshima University, Japan)

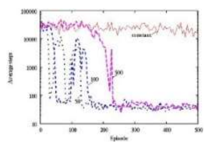
In this paper, the collision between the pow skirt and the body is determined. We verified each simple animation using the Euler method, the FB Euler method, and the Runge-Kutta method. As a result of verification, the Euler method and Runge-Kutta method animations showed contact with the body, movements away from the body of clothes, and violent movements, and the FB Euler method completed a simple and stable animation. However, I encountered problems such as heavy calculations that required manual parameter changes. In addition, issues such as research on texture mapping methods for pow skirts were also raised. Therefore, we will proceed with future research and aim for more realistic reproduction and practical research.

GS abstracts
GS1 Chaos & Application (5)
Chair: Masato Nagayoshi (Niigata College of Nursing, Japan)

GS1-1 A Method of Role Differentiation Using a State Space Filter with a Waveform Changing Parameter in Multi-agent Reinforcement Learning

Masato Nagayoshi, Simon Elderton (Niigata College of Nursing, Japan), Hisashi Tamaki (Kobe University)

Recently, there have been many studies on the multi-agent reinforcement learning (MARL) in which each autonomous agent obtains its own control rule by RL. Here, it is considered that different agents having individuality is more effective than uniform agents in terms of role differentiation in MARL. Therefore, we have proposed a promoting method of role differentiation using a wave-form changing parameter in MARL. In this paper, we confirm the effectiveness of role differentiation using a state space filter with a waveform changing parameter through computational experiments using "Pursuit Game" as one of multi-agent tasks.



GS1-2 The research about editing system of performance information for player piano. - Make inferences about both handed musical composition by using DP matching system -

Ryo Kinoshita, Eiji Hayashi (Kyushu Institute of Technology, Japan)

Playing the piano expressive by player piano, it is necessary to adjust the volume, length, and timing of music. In the case of piano music, there are often 1000 or more notes in the score of even a short piece of music. So, to edit music data manually requires not only knowledge but also a huge amount of time and effort. Therefore, we aimed to develop a system that, like a skilled pianist, can perform even the first musical score based on information related to previous skills and experience. So, we developed a system that automatically estimates the performance expression of unedited music using edited performance data and score data. In this paper, we described a phrase search using DP matching and a method for selecting an optimal phrase, how to infer parameters of notes, evaluation of an inferred entire both-handed song.

GS1-3 Development of a LiDAR based Navigation System for Tree Harvester

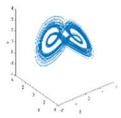
Ayumu Tominaga, Ryusuke Fujisawa, Eiji Hayashi (Kyushu Institute of Technology, Japan), Abbe Mowshowitz (The City College of New York, US)

This study deals with an autonomous moving system for the automation of felling process by high-performance machines in the forestry. Many fatal accidents are caused by logging operations. In this research, a navigation system was developed to realize autonomous movement between accumulation sites and trees to be harvested in order to improve productivity and safety. A three-dimensional map is generated by LiDAR observation, and harvester moves autonomously to the tree as specified by the operator. A test of the harvesting process was conducted in an experimental environment. The evaluation was focused on the required time of autonomous movement in the process. The result showed that the effectiveness of the system was confirmed in operations such as row-thinning.

GS1-4 Numerical study on a class of chaotic financial chaotic systems(withdraw)

Lei Gong, Minghan Song (Tianjin University of Science and Technology, China)

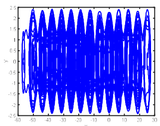
In this paper, a class of chaotic financial system is studied based on numerical simulation. Firstly, chaotic dynamics of the chaotic financial system is briefly introduced. Then, the stability of the equilibrium points is analyzed by using the Routh Hurwitz criterion. Finally, the influence of different parameters on the system is studied by means of phase diagram, Lyapunov exponent diagram and bifurcation diagram, which show a theoretical method to control chaotic dynamics of financial market.



GS1-5 Dynamic characteristics analysis of a multi-scroll conservative chaotic system with sinusoidal nonlinearity

Zhonggao Chen (Tianjin University of Science and Technology, China)

In this paper, a multi-scroll conservative chaotic system with sinusoidal nonlinearity is studied. Based on theoretical analysis and numerical analysis, such as equilibrium point and its stability analysis, Lyapunov exponent spectrum and bifurcation diagram, the system is found to show some complex dynamics. In addition, NIST test and FPGA implementation verify that the pseudo randomness and feasibility of the proposed system is satisfactory.



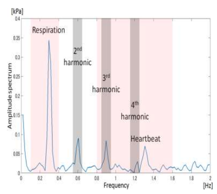
GS2 Image & Signal Processing 1 (4)

Chair: Tohru Kamiya (Kyushu Institute of Technology, Japan)

GS2-1 Unconstrained Measurement of Heart Rate Considering Harmonics of a Respiratory Signal by Tactile Sensor

Kazuya Matsuo (Kyushu Institute of Technology, Japan),
Toshiharu Mukai (Meijo University, Japan),
Shijie Guo (Hebei University of Technology, China)

Measurements of the sleeping state would be useful to monitor the health of a nursed person. The sleeping state can be estimated from biological information such as respiration rate, and heart rate. This study describes a heart rate measurement method by considering the harmonics of a respiratory signal. Respiration and heart rates are measured by applying frequency analysis to the time-series data of body pressure. The harmonics of a respiratory signal serve as noises in heart rate measurement. Therefore, heart rate measurement is improved by eliminating the effects of the harmonics. The average of the frequency errors of heart rate measurement by our proposed method is 0.144 Hz. The results of experiments demonstrated that our proposed method enhances the precision of heart rate measurement.



GS2-2 Image Registration Method for Chest MDCT Images Based on 2-D Finite Element Method

Takuji Ogimoto, Tohru Kamiya (Kyushu Institute of Technology, Japan)
Takatoshi Aoki (University of Occupational and Environmental Health, Japan)

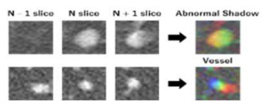
Multi detector-row computed tomography (MDCT) device has been used for early detection of lung cancer. However, there is concern that an increase in the burden on doctors will be caused by improvements in CT performance. Therefore, developing a computer aided diagnosis (CAD) reduces the burden on doctors. A temporal subtraction (TS) technique is one of the CAD systems and it is the subtraction operation of the current image and the previous one of the same patients that emphasizes temporal changes. In the TS image, when the position of the current image and the previous image are misaligned, subtraction artifacts are remained. In this paper, we propose a registration method based on the 2-D finite element method (FEM). In our proposed method, to improve the high computed cost that is the biggest problem of the FEM, we introduce the 2D FEM registration. We applied this method to 31 series MDCT image sets which was obtained previous and current from the same subjects and evaluated. As a result, full width at half maximum (FWHM) of 28.0, artifact to lung volume ratio of 5.77% and computational time of 140 sec were obtained.

(c)output image
Fig.4 Experimental result

GS2-3 Detection of Abnormal Shadows in Low-dose CT Images Using CNN

Hiromu Ikeda, Tohru Kamiya (Kyushu Institute of Technology, Japan)

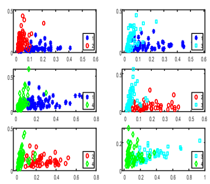
CT imaging is a very effective way to detect abnormal shadows in the lungs. Considering the physical burden on the patient, it should be taken at a low dose. Our approach consists of two main elements. First, candidate abnormal shadow areas are extracted from low-dose CT images of the chest by threshold processing and filtering. Second, we implement a CNN model to classify candidate region on images. To improve identification accuracy in CNN, we devised a method to add spatial information to the input image. In this paper, the proposed method is applied to 13 abnormal shadow images and 100 normal tissue images, and true positive ate of 84.6%, false negative rate of 15.4% were achieved respectively.



GS2-4 EEG signal extraction method based on HHT and CSP

Lei Gong (Tianjin University of Science and Technology, China)

In this paper, the Hilbert Huang transform (HHT) is used to extract the marginal spectrum (MS) of the four-task motor imagery of EEG signals. For the first time, a spatial filter constructed by common spatial pattern (CSP) is used to extract the feature of signal marginal spectrum, namely the marginal spectrum CSP feature (MSCSP). The MI EEG Data2a provided by The University of GRAZ is verified by the experiment, and the results show that compared with the traditional single CSP feature extraction method, the effect of MSCSP feature extraction was significantly improved. Which provide a theoretical and experimental foundation for the application of MI based on BCI system.

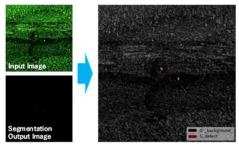


GS3 Image & Signal Processing 2 (3)
Chair: Keiko Sakurai (University of Miyazaki, Japan)

GS3-1 Deep-Learning Based Segmentation Algorithm for Defect Detection in Magnetic Particle Testing

Akira Ueda, Huimin Lu, Tohru Kamiya (Kyushu Institute of Technology, Japan)

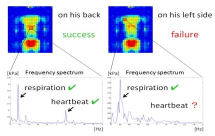
Magnetic Particle Testing (MPT), also referred to as magnetic particle inspection, is a nondestructive examination (NDE) technique used to detect surface and slightly subsurface flaws in most ferromagnetic materials such as iron, nickel, and cobalt, and some of their alloys. In a bad environment, the procedure is complicated, and automation of MPT is strongly desired. To find defects in the formed magnetic powder pattern, it is required to be highly skilled and automation has been considered difficult. In recent years, many defect detection methods based on deep learning have been proposed, and the effectiveness of deep learning has been shown in the task of automatically detecting various types of defects having different shapes and sizes. In this paper, we describe the development of deep learning based segmentation algorithm for defect detection in MPT images. We have achieved a F2 score of 84.04% by using U-Net as the segmentation model and by utilizing a strong backbone network and an optimal loss function.



GS3-2 A Method for Improving Recognition of Lying Postures Using a Measured Signal Intensity of Respiration and Heartbeat by Flexible Tactile Sensor Sheet

Kazuya Matsuo (Kyushu Institute of Technology, Japan), Toshiharu Mukai (Meijo University, Japan),
Shijie Guo (Hebei University of Technology, China)

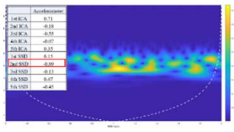
Measurements of the sleeping state would be useful to monitor the health of a nursed person. The sleeping state can be estimated from biological information such as respiration rate, heart rate, and lying postures. This study describes a method for improving recognition of lying postures using a measured signal intensity of respiration and heartbeat. We can obtain respiration and heartbeat by means of using the time series data of the body pressure measured at the suitable location determined by a lying posture. Therefore, a recognition rate of lying postures and a measured signal intensity of respiration and heartbeat have positive correlation. In the experiments, we show that recognition of lying postures is improved by means of using a measured signal intensity of respiration and heartbeat.



GS3-3 Basic research for the realization of online MEG using SSD

Kazuhiro Yagi, Yuta Shibahara, Lindsey Tate, Keiko Sakurai, Hiroki Tamura (University of Miyazaki, Japan)

Neurofeedback systems have been found to be effective in the clinical rehabilitation of paralysis. However, most systems exist only for use with EEG, which is cumbersome to apply to patients and has lower spatial resolution than MEG. Furthermore, the best practices for neural data feature extraction and feature selection are not well established. The inclusion of the best performing feature extraction algorithms is critical to the development of clinical neurofeedback systems. Using simultaneously collected MEG and accelerometer data before and during 10 spontaneous finger movements, we performed an in-depth comparison of independent components analysis (ICA) and spatio-spectral decomposition (SSD) algorithms for their individual abilities to isolate movement-relevant features in brain activity. Having restricted raw data to that from sensorimotor rhythm (SMR) frequencies in select MEG sensors over sensorimotor cortex, we compared ICA and SSD components using: (1) 2D topographies, (2) activations over time, (3) and correlations with accelerometer data at both 0ms and 60ms time delays. SSD performed more quickly and produced components that were more highly correlated with the behavioral data than ICA. We will discuss these results and suggestions for application to neurofeedback systems. In particular, we will present detailed visualizations of SSD results and discuss potential strategies and pitfalls for feature selection.



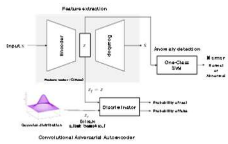
GS4 Image & Signal Processing 3 (3)

Chair: Taro Shibanoki (Ibaraki University, Japan)

GS4-1 Landslide Area Detection from Synthetic Aperture Radar Images Using Convolutional Adversarial Autoencoder and One-class SVM

Shingo Mabu, Soichiro Hirata, Takashi Kuremoto (Yamaguchi University, Japan)

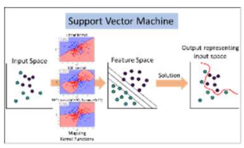
In this paper, an anomaly detection system using deep learning for detecting disaster-stricken (landslide) areas in satellite images is proposed. We used synthetic aperture radar (SAR) images because SAR can observe a wide area of the earth even at night or in bad weather. However, it is difficult to obtain a large number of training images with annotations. Especially it is much more difficult to obtain disaster area images than normal area images. To overcome this problem, we proposed an anomaly detection system that only uses normal area images for the training. The proposed system combines a convolutional adversarial autoencoder (CAAE) and one-class SVM, which realized good feature extraction and anomaly detection. In the experiments, the detection ability of normal and abnormal areas was evaluated.



GS4-2 Anomaly Detection in Time Series Data Using Support Vector Machines

Umaporn Yokkampon[1], Sakmongkon Chumkamon[1], Abbe Mowshowitz[2], Eiji Hayashi[1] and Ryusuke Fujisawa[1] ([1]Kyushu Institute of Technology, Japan) ([2]The City College of New York, USA)

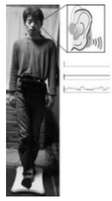
Analysis of large data sets is increasingly important in business and scientific research. One of the challenges in such analysis stems from uncertainty in data, which can produce anomalous results. In this paper, we propose a method of anomaly detection in time series data using a Support Vector Machine. Three different kernels of the Support Vector Machine are analyzed to predict anomalies in the UCR public data set. Comparison of the three kernels shows that the defined parameter values of the RBF kernel are critical for improving the validity and accuracy in anomaly detection. Our results show that the RBF kernel of the Support Vector Machine can be used to advantage in detecting anomalies.



GS4-3 Relationship Between Tactile Stimuli and Human Body Sway

Masaya Tadokoro, Taro Shibanoki (Ibaraki University, Japan)

This paper describes the relationship between tactile stimulation and human body sway. We previously proposed a body sway mitigation system based on tactile stimulation and revealed that simultaneous stimulation behind both auricles were significantly improve human balance function, however, the mechanism of it were not fully discussed. In this paper, some stimulation patterns were applied to participants and COP eccentricities before/after stimulation was extracted. The results showed that COP values after stimulation could be inclined to the opposite of the stimulation site. It indicates that tactile stimuli can control human balance function.

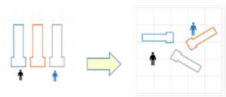


GS5 Genetic Algorithms & Robotics 1 (3)
Chair: Hideyuki Tanaka (Hiroshima University, Japan)

GS5-1 Layout decision system for multiple production lines using work-flow-line and GA

Masato Noda[1], Hidehiko Yamamoto[1] (Gifu University, Japan),
Hirohumi Tsuji[2], Yasuhisa Terawa[2], Yoshinori Nakamura[2], Masayuki Tsuchida[2] (Infofarm Corporation, Japan),
Katsuaki Yamada[3], Yukiyasu Kuriyama[3] (Kai Industries Corporation, Japan)

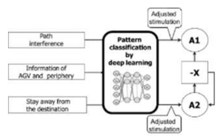
We develop the system to decide the efficient layout of assembly production line by using Genetic Algorithm (GA). The target production line has some operators. We call the system as System of Production-line-layout Decision by Chameleon-code and GA (PDCG). PDCG decides the efficient layout of production line by using GA, work flow-lines acquired by Chameleon code and the machine failure data. Specifically, PDCG evaluates the layout efficiency by calculating the operator's walking time in order to fix the machine failures which sometimes occur on the production line. By the evaluation, PDCG can find the good layout which has multiple production lines with one operator. We applied PDCG to the actual razors assembly line's layout in order to verify the performance of PDCG. We compare the layout obtained by PDCG with the parallel layout currently used in razors assembly factory. As a result, PDCG's layout reduced the walking time, compared with the parallel layout.



GS5-2 Autonomous decentralized FMS's AGVs moving control by mind change with deep learning

Ryunosuke Yamane, Hidehiko Yamamoto (Gifu University)

We have developed the system of autonomous decentralized flexible manufacturing system (AD-FMS) factory. Especially, automated guided vehicles (AGVs) in the AD-FMS factory avoided path interferences between AGVs by inserting the mind model into the AGVs. By using the mind model, the interferences can be avoided by repeating to change two types of mind, the arrogant mind and the modest mind. However, the problem is still left that it takes time to make the decision which way an AGV moves to avoid the interference. In order to solve the problem, we develop the new mind model. The new mind model includes the deep learning system. The new model is called Minimum Unit of Mind with Deep learning (MUMD). We applied the new mind model including the deep learning system to AD-FMS and ascertained MUMD is useful in the time of path interferences avoidances.



GS5-3 Secondary School Robotics Education in Camarin High School: Developments and Challenges for Improvement

Jeffrey Rivera Galino (Camarin High School, Philippines and Hiroshima University, Japan)
Hideyuki Tanaka (Hiroshima University, Japan)

Camarin High School is one of the schools in the Philippines that is making an initiative in improving STEM education through robotics education. In this paper, the authors report the development of a robotics program for students in junior high school. Through the Teacher Training Program of the Japanese Government for international teachers, localized challenges were identified and advances in technology education were examined for consideration. This report is hoped to provide useful insights to both practitioners of robotics education and initiators in a similar situation.

GS6 Robotics 1 (4)
Chair: Noritaka Sato (Nagoya Institute of Technology, Japan),

GS6-1 Digital Testing Device for Active Range of Motion of Finger Joints Utilizing Artificial Neural Network

Huu-Hieu Quang, Yoshifumi Morita (Nagoya Institute of Technology, Japan), Makoto Takekawa (everfine, Japan)

This paper proposed a digital goniometer device for measuring the active range of motion (A-ROM) of finger joints by artificial neural network (ANN). Although a therapist typically measures the A-ROM by a physical goniometer, it consumes a significant amount of time, and imposes a relatively heavy burden on patients. The proposed device comprises an Intel RealSense depth camera and a computer. The camera is set up toward the patient`s finger joints. Subsequently, the therapist captures to collect RGB-D data. The three-dimensional position coordinates of the specified joint are acquired from the RGB-D data. These data are used as inputs to an ANN model for predicting the joint`s angles. The effectiveness of the proposed device is confirmed experimentally on the third joint of the middle angle.

Digital Testing device for A-ROM of finger joints

GS6-2 Position and Force Teaching Method for 6 DoF Manipulator Using Contact Teaching Tool and Teaching Data Editor

Duy-Do Bui, Hiroki Tanaka, Quang-Trung Chu, Hideki Inuzuka, Yoshifumi Morita (Nagoya Institute of Technology, Japan) Masao Sakai (Aichi Prefecture, Japan)

This paper proposed a method for teaching tasks performed by a six degree-of-freedom robot manipulator while the robot’s tool makes contact with a workpiece. By using our previously developed direct teaching device, the time to teach pressing angle and force of the tool is greatly reduced compared to using the teaching pendant. In this study, a contact teaching tool and a teaching data editor were proposed on the assumption that the compliance characteristics of the robot hand are known. The tool allows us to accurately grasp the shape of the workpiece. The editor generates the target trajectory to realize the desired pressing angle and force on the basis of the teaching data and the known characteristics. The effectiveness was confirmed experimentally from the viewpoint of improving the teaching accuracy and reducing the teaching time.

PAWTED with a contact teaching tool

GS6-3 Haptic Device that Presents Sensation Corresponding to Palm on Back of Hand for Teleoperation of Robot Hand Report2: Consideration on Decided Specification

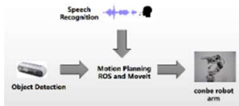
Kyosuke Ushimaru,Noritaka Sato,Yoshifumi Morita (Nagoya Institute of Technology, Japan)

Recently, teleoperated rescue robots are required. However, it is known that the teleoperation of a robot hand mounted on a rescue robot is difficult. Therefore, we propose a new haptic device that presents haptic sensation for teleoperation of a robot hand. The device stimulates the back of the hand instead of the palm of the operator. Determination of required specifications by an experiment with subjects is written in this paper. To design the device, the interval of the stimulation cue (i), the diameter of the stimulation cue (d), and the force of the stimulation (f) should be optimized. As a result of the experiment, we found that the accuracy rate is highest, when (i, d, f) = (30mm, 6mm, 0.9kgf). Moreover, we deeply considered on the decided specification with an additional experiment.

GS6-4 Robot Assisting Water Serving to Disabilities by Voice Control

Yang Chunxin, Sakmongkon Chumkamon, Eiji Hayashi (Kyushu Institute of Technology, Japan)

ROS is an open-source robot operating system. In this paper, we use ROS to control Conbe robot arm. By introducing the YOLACT real-time instance segmentation, we trained our own model for Object Detection. Secondly, the Speech-Recognition system is established through Deep speech and Mozilla Text-To-Speech with Tacotron2 DDC model. Deep speech is an end-to-end speech system, where deep learning supersedes these processing stages. Combined with a language model, this approach achieves higher performance than traditional methods on hard speech recognition tasks while also being much simpler. In this way, we create an artificial intelligence, which accomplished a simple conversation with people. And the voice control system is established based on Speech-Recognition system. In the experiment, we successfully control the robot arm move positions and do water serving for disabilities by voice command. With this research, voice control robot arm can be apply in the life support area, it will be more convenient for disabilities in daily life

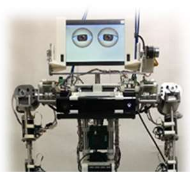


GS7 Robotics 2 (4)
Chair: Wisanu Jitviriya (Kyushu Institute of Technology, Japan)

GS7-1 Development of Interactive Robot – Emotion Estimation System Using Speech by 1dCNN -

Yugo Kawachi, Eiji Hayashi (Kyushu Institute of Technology, Japan)

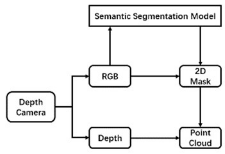
With the expansion of the market for the robotics industry, the development of service robots is becoming more popular. These robots are designed to be used in homes and in medical and welfare setting where they can communicate with people, and they need to act and talk in friendly. In this research, we are developing an interactive robot that pursues human-like movements by focusing on non-verbal interactions such as facial expressions and body language. We can guess the other person's emotions from the intonation of speech. In addition to the facial expressions and body language used in previous studies, we considered that a function for emotion estimation from speech is necessary. Therefore, we developed a machine learning system using 1dCNN for emotion estimation of speech.



GS7-2 Deep Learning Methods for Semantic Segmentation of Dense 3D SLAM Maps

Pei Yingjian, Sakmongkon Chumkamon, Eiji Hayashi (Kyushu Institute of Technology, Japan)

Most real-time SLAM systems can only achieve semi-dense mapping, and the robot lacks specific knowledge of the mapping results, so it can only achieve simple positioning and obstacle avoidance, which may be used as an obstacle in the face of the target object to be grasped, thus affecting the realization of motion planning. The use of semantic segmentation in dense SLAM maps allows the robot to better understand the map information, distinguish the meaning of different blocks in the map by semantic labels, and achieve fast feature matching and Loop Closure Detection based on the relationship between semantic labels in the scene. There are many semantic segmentation datasets based on street scenes and indoor scenes available for use, and these datasets have some common tags. Based on these training data, we can derive a semantic segmentation model based on RGB images by using the Pytorch platform for training.



GS7-3 Robot Motion Generation by Hand Demonstration

Sakmongkon Chumkamon, Umaporn Yokkampon, Eiji Hayashi, Ryusuke Fujisawa
(Kyushu Institute of Technology, Japan)

Since traditional robot teaching requires time and instruction to the robot motion, we present a system framework and experiment for generating robot motion trajectories from human hand demonstration, which the worker could teach robot easier rather than assigning the instruction to the robot controller manually. Therefore, the robot can imitate the action in a new situation instead of directly teaching the robot arm. Our contributions are including three-point 1) the extracting method of hand movement with marker-less using hand detection in 3D from human 2) the motion generalization of the hand trajectories from human 3) Robot path planning for grasping and place the object to the target. We also present the experiment conducted by the user movement for real data and evaluate the system using the manipulator's gripper. The experiment shows the pick-and-place task of the robot for food by hand demonstration.

GS7-4 Medical Telerobotics: IRAPs SHaRE-aGIVeR

Noppadol Pudchuen, Jiraphan Inthiam, Wisanu Jitviriya, Amornphun Phunopas, Chirdpong Deelertpaiboon
(King Mongkut's University of Technology North Bangkok, Thailand)
Aran Blattler (Kyushu Institute of Technology, Japan)

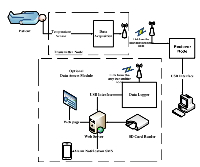
The Coronavirus disease 2019 (COVID-19) pandemic has affected the global population. In particular, the medical personnel in direct contact with patients have been exposed to high risk. To reduce the spread of COVID-19 and protect healthcare workers and patients, we would like to present the fully automated medical telerobot as the IRAPs SHaRE-aGIVeR robot. Our robot is capable of generating both 2D and 3D maps automatically, delivering medical supplies, food, or medical devices such as blood pressure monitors, pulse oximeters, and so on. In addition, the user interface system is also vital part. Users are able to connect and control the robot using a computer, a mobile device, or a tablet via the wireless network which is installed inside the robot. Currently, our robots are being operated on with medical personnel at hospitals, regarding the feedback from the formal caregivers that can prove our robot's efficiency in reducing the risk of COVID-19 spread.

GS8 Medical Data Processing (5)
Chair: Shingo Mabu (Yamaguchi University, Japan)

GS8-1 Development of in-Home Wireless Continuous Temperature Data Logging and Alarming System for Fever Monitoring in Pediatrics

Ali S. AlMejrad(University of Hail, Kingdom of Saudi Arabia)

Fever is one of the most common concern to be taken care by parents that sometimes force them to bring their child to the hospital emergency for continuous follow up especially at night when sleeping. Fever is defined when a rectal temperature reaches over 38.0 C (100.4 F). Temperatures measured from other sites of body are usually less. There is no defined threshold for fever due to variation of body temperatures between different individuals as much as 1 F. Fever itself is not life threatening unless the rectal temperature is persistent over than 41.6 C (107 F). Fever may be due to a serious illness and usually is caused by common infections that are not serious. However, a fever can be caused by other causes other than infections. Since continuous temperature monitoring is inapplicable at the rectal site, the paper aims at developing a prototype system using Wireless Sensor Network (WSN) for wireless continuous fever monitoring system from the head skin or wrist that are the most applicable area even so one must add 2 degrees for the read temperature correction. The system can record the temperature at the patient terminal and transmit the fever status and critical situations to an alarming system in the side of taking-care people that will help in in-home care for pediatrics suffering from fever.

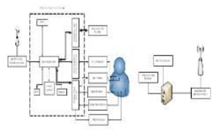


System Block Diagram

GS8-2 Design and Development of in-home Wireless Crucial Events Logging and Alarming System for Elderly and Disabled People Care

Ali S. AlMejrad(University of Hail, Kingdom of Saudi Arabia)

This paper aims at developing a system that will help in in-home care for elderly or disabled people. The system can record the patient vital parameters such as temperature, SpO2, ECG, heart rate, bed moisture and fall off sensor which could be installed at the bathroom or shower area at the patient terminal and transmit the critical situations such as high or low values to an alarming system. The system will be developed using the National Instrument Wireless Sensor Network (NI-WSN) to control the crucial events transmission and LabVIEW software to design the user interface.



System Block Diagram

GS8-3 Virtual bird's-eye view for remote operation of unmanned construction machinery

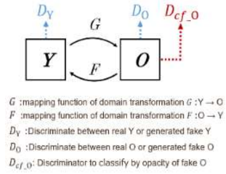
Noritaka Sato, Akihiro Fukuda (Nagoya Institute of Technology, Japan)

Recently, remote operation of construction machinery is required to release humans from dangerous places. However, the operator only watches a video from a camera mounted on a construction machine in general and lacks the sense of distance. Therefore, risk of collisions and operation time increase. In order to solve those problems, we propose a method to display virtual bird's-eye view which is created by superimposing a CG model of a construction machine on pre-captured pictures. The operator can recognize positional relationship between the machine and the environment by using the proposed method. Experiments are carried out to verify the effectiveness of the proposed method which can reduce the number of collisions and operation time comparing to the conventional method.

GS8-4 Domain Transformation of Chest CT Images Using Semi-Supervised Cycle GAN for Opacity Classification of Diffuse Lung Diseases

Masashi Miyake[1], Shingo Mabu[1] , Shoji Kido[2], Takashi Kuremoto[1]
([1]Yamaguchi University, Japan, [2]Osaka University, Japan)

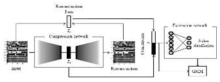
The aim of this research is to perform domain translation of chest CT images so that medical institutions can effectively use a computer-aided diagnosis (CAD) system trained at a different institution. We propose a semi-supervised Cycle GAN for domain transformation by combining the standard Cycle GAN and the trained CAD. In detail, after some images are domain-transformed by Cycle GAN, the trained CAD classifies them to examine whether they are appropriately transformed or not. Then, the results are fed back to the Cycle GAN to re-train. In the experiments, classification of normal opacity and five kinds of abnormal opacities of diffuse lung diseases in CT images are carried out and the effectiveness of the domain transformation of the proposed method is clarified.



GS8-5 Anomaly Detection of Lung Sounds Using DAGMM

Ryosuke Wakamoto[1], Shingo Mabu[1], Shoji Kido[2], Takashi Kuremoto[1]
([1]Yamaguchi University, Japan, [2]Osaka University, Japan)

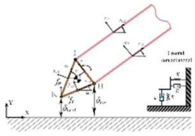
There are few studies using deep learning for auscultation, and also, there are only a few small-scale benchmark datasets of lung sounds that are annotated for the training of machine learning. Therefore, we aim to build an anomaly detection system that only uses normal data that can be obtained more than abnormal data. To realize the anomaly detection system, we propose an algorithm that improves the Deep Autoencoding Gaussian Mixture Model (DAGMM), that simultaneously implements feature extraction and clustering. In this study, various types of neural networks are applied to DAGMM as the compression networks for efficient feature extraction. From the experimental results, it is clarified that the proposed methods show effective classification performance for lung sounds.



GS8-6 A ground reaction force analysis in walking and running gaits in horse leg model on viscoelastic hoof-ground contact

Dondogjamts Batbaatar, Hiroaki Wagatsuma (Kyushu Institute of Technology, Japan)

In this paper, we focused on the hoof-ground interaction in the simplified horse leg model because walking and running gaits are known to be different in trajectories; however, the force analysis still remains as unsolved issues. The computational experiments in Matlab, elastic and inelastic impact with the ground was resolved by using the dissipative contact force model and the ground reaction force was clearly examined in four different conditions from the combination of walking/running and elastic/inelastic contact. The proposed method contributes to the establishment of the detail time course analysis of the ground reaction force for animal and human gaits in a consistent manner.



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)

[A]			
Abdullah	Azian	OS1-3	39/43
	Azamimi		
Abe	Akinori	OS15-7	26/75
Aburada	Kentaro	OS16-3	30/76
		OS16-4	30/76
		OS16-5	30/77
Adam	Asrul	OS1-1	39/42
Ahn	Seong-Ha	OS4-1	34/48
Aizawa	Shogo	OS5-6	29/50
Al aama	Obada	OS19-4	21/80
AlMejrad	Ali S.	GS8-1	33/98
		GS8-2	33/98
Alpizar	Ixchel	OS11-1	29/61
	Ramirez		
Aoki	Takatoshi	GS2-2	24/91
Arima	Kaoru	OS16-5	30/77
Asada	Taro	OS2-2	22/44
		OS2-3	22/44
		OS2-4	22/45
[B]			
Batbaatar	Dondogjamts	GS8-6	34/99
Blattler	Aran	GS7-4	40/97
Bui	Duy-Do	GS6-2	32/95
Bulatov	Sergey	OS9-1	22/57
[C]			
Chao	Shu-Tsung	OS7-3	32/53
Cho	Young Im	OS4-2	34/48
		OS4-3	34/48
Chen	Chen-Pang	OS8-1	23/55
Chen	Shuailin	OS12-13	38/65
Chen	Shuo-Tsung	OS9-5	22/58
Chen	Yue	OS12-1	37/62
Chen	Zhonggao	GS1-5	33/90
Cheng	Yuhui	OS12-6	37/64
		OS13-5	36/68
Cheng	Yung-Hsin	OS9-7	22/59
Chu	Quang-Trung	GS6-2	32/95
Chuang	Ching-Chun	OS9-4	22/58
		OS9-5	22/58
Chuang	Li-Ming	OS7-3	32/53
Chumkamon	Sakmongkon	GS4-2	35/93
		GS6-4	33/96
		GS7-2	39/97
		GS7-3	40/97
[D]			
Dahari	Mahidzal	OS1-4	39/43
Dai	Fengzhi	OS13-1	36/67
		OS13-2	36/68
		OS13-3	36/68
		OS13-4	36/68
		OS13-5	36/68
		OS13-6	36/69
		OS13-7	36/69
		OS13-8	36/69
		OS13-9	36/69

		OS13-10	36/70
		OS13-11	37/70
		OS13-12	37/70
		OS13-13	37/71
		OS13-14	37/71
Deelertpaiboon	Chirdpong	GS7-4	40/97
Deng	Jianhua	OS12-14	38/66
		OS12-15	38/66
Domae	Yukiyasu	OS11-1	29/61
Dudin	Vitaly	OS9-1	22/57
[E]			
Egawa	Haruka	OS14-1	28/71
Elderton	Simon	GS1-1	33/89
[F]			
Fuengfusin	Ninnart	OS19-1	21/79
Fujisawa	Ryusuke	GS1-3	33/89
		GS4-2	35/93
		GS7-3	40/97
Fukuda	Akihiro	GS8-3	33/98
Fukuda	Kazuki	OS14-4	28/72
Fukushima	Hiroki	OS15-5	26/74
[G]			
Galino	Jeffrey Rivera	GS5-3	26/95
Gao	Longyu	OS12-16	38/66
		OS12-17	38/67
		OS12-18	38/67
		OS13-14	37/71
Ge	Jing	OS6-3	30/51
Gong	Kai	OS6-2	30/51
		OS6-4	30/52
Gong	Lei	GS1-4	33/90
		GS2-4	25/91
Guan	Zhe	OS10-3	24/60
		OS10-4	24/60
Guo	Jr-Hung	OS8-5	24/57
		OS9-6	22/59
Guo	Shijie	GS2-1	24/90
		GS3-2	27/92
[H]			
Hamasaki	Yosiki	OS14-5	28/73
Han	Shuo	OS17-2	31/77
Han	Leixin	OS12-17	38/67
Hanajima	Naohiko	OS6-1	30/51
Hao	Hongbo	OS12-7	37/64
		OS12-8	37/64
		OS12-9	38/64
		OS12-12	38/65
		OS12-13	38/65
		OS13-7	36/69
		OS13-8	36/69
Hao	Junlong	OS12-18	38/67
Harada	Kensuke	OS11-1	29/61
		OS11-2	29/62
		OS11-3	29/62
Halin	Abdul Hafiz	OS1-3	39/43
Hayashi	Eiji	GS1-2	33/89
		GS1-3	33/89
		GS4-2	35/93
		GS6-4	33/96
		GS7-1	39/96
		GS7-2	39/97
		GS7-3	40/97
Hiraoka	Toru	OS5-1	28/49
		OS5-2	28/49
		OS5-5	29/50
Hirata	Soichiro	GS4-1	35/93
Hirota	Koichi	OS10-7	24/61
Hirota	Masaharu	OS5-3	29/49
		OS5-4	29/50
Hori	Sansei	OS19-3	21/79
Hsia	Kuo-Hsien	OS8-5	24/57
		OS9-6	22/59
Hsiao	Jia Ming	OS9-7	22/59

Hu	Manqian	OS12-18	38/67
Hua	Chih-Chiang	OS9-4	22/58
		OS9-5	22/58
Hung	Chung-Wen	OS8-1	23/55
		OS8-2	23/55
		OS9-4	22/58

[I]

Ibrahim	Zuwairie	OS1-1	39/42
Iiyama	Shota	OS5-4	29/50
Ikeda	Hiromu	GS2-3	25/91
Imai	Shinichi	OS10-1	24/59
Inthiam	Jiraphan	GS7-4	40/97
Inuzuka	Hideki	GS6-2	32/95
Ishii	Kazuo	PS1	21/41
		OS22-2	31/83
		OS22-3	31/84
		OS22-4	31/84
		OS22-5	31/84
		OS23-1	27/85
		OS23-2	27/85
		OS23-3	27/85
		OS23-4	27/86
		OS23-5	27/86
		OS23-6	27/86
Ishikawa	Seiji	OS14-1	28/71
		OS14-2	28/72
Ito	Takao	OS24-1	25/87
		OS24-2	25/87
		OS24-4	25/88
		OS24-5	25/88
Ito	Takuya	OS15-3	26/74
Ito	Tsutomu	OS24-1	25/87
		OS24-2	25/87
		OS24-4	25/88
		OS24-5	25/88
Iuchi	Keisuke	OS21-1	23/82

[J]

Jhang	Guang-Ming	OS8-3	23/56
Jhang	Yun-Syuan	OS9-2	22/57
Ji	Chengxu	OS13-5	36/68
Jia	Yuxin	OS6-2	30/51
		OS6-4	30/52
Jia	Yingmin	OS6-2	30/51
		OS6-4	30/52
Jian	Tianye	OS3-2	34/45
		OS3-8	35/47
Jitviriya	Wisanu	GS7-4	40/97

[K]

Kamasaka	Taketo	OS24-1	25/87
		OS24-2	25/87
		OS24-4	25/88
		OS24-5	25/88
Kamazlan	Muhammad Zubir	OS1-3	39/43
Kamei	Keiji	OS22-3	31/84
Kamiya	Tohru	GS2-2	24/91
		GS2-3	25/91
		GS3-1	26/92
Kang	Ho-Sun	OS4-1	34/48
Kang	Ruming	OS12-17	38/67
		OS13-6	36/69
Katayama	Tetsuro	OS16-3	30/76
		OS16-4	30/76
		OS16-5	30/77
Kawada	Kazuo	OS21-1	23/82
		OS21-2	23/82
		OS21-3	23/82
		OS21-4	23/83
Kawai	Miku	OS15-1	25/73
		OS15-2	26/73
Kawachi	Yugo	GS7-1	39/96
Kharisova	Enzhe	OS9-1	22/57
Khazetdinov	Artur	OS9-1	22/57
Kido	Shoji	GS8-4	34/99
		GS8-5	34/99

Kimura	Kenji	OS23-5	27/86
Kinoshita	Takuya	OS10-2	24/60
		OS10-5	24/60
Kinoshita	Ryo	GS1-2	33/89
Kita	Yoshihiro	OS16-3	30/76
		OS16-4	30/76
		OS16-5	30/77
Kitagawa	Kohei	OS14-2	28/72
Kobayashi	Hiroyuki	OS8-2	23/55
Koyama	Keisuke	OS11-2	29/62
Kumano	Minoru	OS5-5	29/50
Kumar	Anirban	OS1-2	39/42
Kuremoto	Takashi	GS4-1	35/93
		GS8-4	34/99
		GS8-5	34/99
Kuriyama	Yukiyasu	GS5-1	26/94
Kurozumi	Ryo	OS16-1	29/75
[L]			
Lai	Chia-Chun	OS7-2	32/53
Lavrenov	Roman	OS9-1	22/57
Lee	Jang-Myung	OS4-1	34/48
Lee	Jen-Tun	OS20-2	39/81
	Jen-Tun	OS20-4	39/81
Lee	Tung-Lin	OS7-4	32/54
Lee	Yu-Po	OS7-3	32/53
Li	Di	OS3-4	35/46
Li	Jung-Shian	OS7-1	32/53
		OS7-2	32/53
		OS7-4	32/54
Liang	Yi-Li	OS20-2	39/81
Lin	Chien-Tung	OS20-1	38/80
Lin	Tzu-Yuan	OS9-6	22/59
Liu	Chengzhi	OS3-3	34/46
Liu	Chuan-Gang	OS7-2	32/53
Liu	Hsiu-Hao	OS9-2	22/57
		OS9-3	22/58
Liu	I-Hsien	OS7-1	32/53
		OS7-2	32/53
		OS7-4	32/54
Liu	Jie	OS3-5	35/46
Liu	Mingfei	OS12-17	38/67
Liu	Qunpo	OS6-1	30/51
Liu	Zilong	OS12-16	38/66
		OS13-14	37/71
Liu	Zonghui	OS6-1	30/51
Lu	Huimin	GS3-1	26/92
Lu	Shengyang	OS12-1	37/62
Lu	Peng	OS12-5	37/63
		OS13-3	36/68
		OS13-4	36/68
		OS13-5	36/68
Lv	Yongzheng	OS12-16	38/66
[M]			
Ma	Yuehang	OS22-1	31/83
Mabu	Shingo	GS4-1	35/93
		GS8-4	34/99
		GS8-5	34/99
Maeda	Kohei	OS5-1	28/49
Magid	Evgeni	OS9-1	22/57
Matsui	Ayumi	OS2-3	22/44
Matsumura	Toshimune	OS22-4	31/84
Matsuo	Kazuya	GS2-1	24/90
		GS3-2	27/92
Matsuo	Takayuki	OS22-2	31/83
		OS22-5	31/84
Matsuura	Kenta	OS11-2	29/62
Miyake	Masashi	GS8-4	34/99
Miyamoto	Kodai	OS24-1	25/87
		OS24-2	25/87
		OS24-4	25/88
		OS24-5	25/88
Miyata	Kazunori	OS20-4	39/81
Miyauchi	Shoko	OS11-3	29/62
Mochamad Chandra	Saputra	OS16-3	30/76
Mochizuki	Ryuugo	OS23-6	27/86

M. Khairuddin	Anis Salwa	OS1-1	39/42
Mohamed Shah	Noraisyah	OS1-4	39/43
Mokhtar	Norrima	OS1-1	39/42
		OS1-3	39/43
		OS1-4	39/43
Mori	Shinnosuke	OS21-4	23/83
Morita	Yoshifumi	GS6-1	32/95
		GS6-2	32/95
		GS6-3	33/96
Morooka	Ken'ichi	OS11-3	29/62
Mowshowitz	Abbe	GS1-3	33/89
		GS4-2	35/93
Mukai	Toshiharu	GS2-1	24/90
		GS3-2	27/92
Mukunoki	Masayuki	OS24-3	25/88
Murai	Keita	OS21-1	23/82
[N]			
M. Nor	M. Rudzuan	OS1-3	39/43
Nagamatsu	Masayasu	OS21-3	23/82
Nagayoshi	Masato	GS1-1	33/89
Nakamoto	Masayoshi	OS10-6	24/61
Nakamura	Akira	OS11-1	29/61
Nakamura	Jun	OS15-4	26/74
Nakamura	Yoshinori	GS5-1	26/94
Nakatani	Yuki	OS10-2	24/60
Ning	Lei	OS12-14	38/66
		OS12-15	38/66
Nishida	Yuya	OS22-2	31/83
		OS22-4	31/84
		OS22-5	31/84
		OS23-1	27/85
		OS23-2	27/85
		OS23-3	27/85
		OS23-4	27/86
		OS23-6	27/86
Nishiya	Y.	OS10-5	24/60
Niu	Hong	OS12-2	37/63
Noda	Masato	GS5-1	26/94

[O]			
Oda	Tetsuya	OS5-3	29/49
		OS5-4	29/50
Ogata	Takashi	OS15-1	25/73
		OS15-2	26/73
		OS15-3	26/74
Ogawa	Hanako	OS10-6	24/61
Ogimoto	Takuji	GS2-2	24/91
Ohnishi	Yoshihiro	OS21-2	23/82
		OS21-4	23/83
Okazaki	Naonobu	OS16-3	30/76
		OS16-4	30/76
		OS16-5	30/77
		OS24-3	25/88
Ono	Jumpei	OS15-1	25/73
		OS15-2	26/73
		OS15-3	26/74
Ouchi	Tomohito	OS2-1	21/44
[P]			
Pei	Xin	OS17-3	31/78
Pei	Yingjian	GS7-2	39/97
Peng	Wanlong	OS3-2	34/45
		OS3-8	35/47
Peng	Yizhun	OS3-1	34/45
		OS3-2	34/45
		OS3-3	34/46
		OS3-4	35/46
		OS3-5	35/46
		OS3-6	35/46
		OS3-7	35/47
		OS3-8	35/47
		OS3-9	35/47
Phunopas	Amornphun	GS7-4	40/97
Priyambadha	Bayu	OS16-4	30/76
Pudchuen	Noppadol	GS7-4	40/97
[Q]			

Qin	Zhongxu	OS12-3	37/63
Quang	Huu-Hieu	GS6-1	32/95
[R]			
Raghura	Sreeraman	OS23-4	27/86
	Srinivasa		
Rajagopal	Heshalini	OS1-1	39/42
		OS1-2	39/42
Rajapakse	R.P.C. Janaka	OS20-1	38/80
		OS20-2	39/81
		OS20-3	39/81
		OS20-4	39/81
[S]			
Sakai	Masao	GS6-2	32/95
Sakamoto	Makoto	OS24-1	25/87
		OS24-2	25/87
		OS24-4	25/88
		OS24-5	25/88
Sakoma	Kenji	OS24-1	25/87
		OS24-2	25/87
		OS24-4	25/88
		OS24-5	25/88
Sakurai	Keiko	GS3-3	27/93
Sakurai	Miki	OS2-3	22/44
Sanada	Atsushi	OS22-3	31/84
Sano	Tomoya	OS14-3	28/72
Sato	Noritaka	GS6-3	33/96
		GS8-3	33/98
Seino	Kai	OS15-6	26/75
Seki	Hiroaki	OS11-3	29/62
Seto	Noritake	OS24-2	25/87
Shakhnoza	Muksimova	OS4-2	34/48
Shen	Luqi	OS12-16	38/66
		OS12-18	38/67
Sheng	Chang	OS13-2	36/68
Sheng	Yuhang	OS12-4	37/63
Shi	LiYing	OS10-3	24/60
Shi	Songyun	OS3-1	34/45

		OS12-16	38/66
Shibahara	Yuta	GS3-3	27/93
Shibanoki	Taro	GS4-3	35/94
Shirahashi	Kanako	OS22-5	31/84
Shou	Ho-Nien	OS7-5	32/54
Sohara	Naoto	OS23-1	27/85
Solpico	Dominic B.	OS23-3	27/85
Song	Chau-Chung	OS8-1	23/55
		OS8-2	23/55
Song	Minghan	GS1-4	33/90
Song	Yunzhong	OS6-6	30/52
		OS13-1	36/67
Sonoda	Takashi	OS22-2	31/83
Su	Bo	OS6-1	30/51
Su	Kuan-Ming	OS7-1	32/53
Suzuki	Hidekazu	OS22-1	31/83
Suzuki	Hiroyuki Y.	OS21-1	23/82
		OS21-3	23/82
Suzuki	Yasuhiro	OS18-1	32/78
		OS18-2	32/78
		OS18-3	32/78
Suzuki	Yosuke	OS11-3	29/62
[T]			
Tabuse	Masayoshi	OS2-1	21/44
		OS2-2	22/44
		OS2-3	22/44
		OS2-4	22/45
Tadokoro	Masaya	GS4-3	35/94
Tajima	Sho	OS11-3	29/62
Takagi	Tomohiko	OS16-1	29/75
		OS16-2	30/76
Takaki	Ryosuke	OS5-2	28/49
Takatuka	Kayoko	OS24-3	25/88
Takei	Amane	OS24-1	25/87
		OS24-2	25/87
		OS24-4	25/88
		OS24-5	25/88
Takekawa	Makoto	GS6-1	32/95

Takemura	Yasunori	OS22-2	31/83
		OS22-3	31/84
		OS22-5	31/84
Tamai	Teruyuki	OS21-2	23/82
		OS21-4	23/82
Tamaki	Hisashi	GS1-1	33/89
Tamukoh	Hakaru	OS19-1	21/79
		OS19-2	21/79
		OS19-3	21/79
		OS19-4	21/80
Tamura	Hiroki	GS3-3	27/93
Tan	Joo Kooi	OS14-1	28/71
		OS14-2	28/72
		OS14-3	28/72
		OS14-4	28/72
		OS14-5	28/73
Tanaka	Hideyuki	GS5-3	26/95
Tanaka	Hiroki	GS6-2	32/95
Tanaka	Yoshiki	OS22-4	31/84
		OS23-2	27/85
Tate	Lindsey	GS3-3	27/93
Terawa	Yasuhisa	GS5-1	26/94
Tian	Shuhao	OS12-1	37/62
Tokuyama	Yoshimasa	OS20-1	38/80
		OS20-2	39/81
		OS20-3	39/81
Tominaga	Ayumu	GS1-3	33/89
Tsuchida	Masayuki	GS5-1	26/94
Tsuji	Hirohumi	GS5-1	26/94
Tsuji	Tokuo	OS11-3	29/62

[U]

Uchida	Noriyuki	OS24-3	
Ueda	Akira	GS3-1	26/92
Uemura	Yuuichiro	OS22-4	31/84
Ul Husnain	Anees	OS1-4	39/43
Ushimaru	Kyosuke	GS6-3	33/96
Usuda	Akio	OS16-2	30/76

[W]

W. M. Mahiyidin	Wan Amirul	OS1-1	39/42
Wada	Mayuka	OS2-2	22/44
Wagatsuma	Hiroaki	GS8-6	34/99
Wakamoto	Ryosuke	GS8-5	34/99
Wakitani	Shin	OS10-7	24/61
Wan	Khairunizam	OS1-1	39/42
		OS1-3	39/43
Wan	Weiwei	OS11-2	29/62
Wang	Chun-Chi	OS8-1	23/55
Wang	Chun-Chieh	OS8-3	23/56
		OS8-4	23/56
Wang	Jiwu	OS17-1	31/77
		OS17-2	31/77
		OS17-3	31/78
Wang	Mengke	OS6-1	30/51
Wang	Min	OS12-10	38/65
Wang	Nana	OS3-7	35/47
Watanabe	Kaori	OS22-1	31/83
Watanabe	Keisuke	OS23-7	27/87
Watanabe	Tetsuyou	OS11-3	29/62
Watanabe	Yusuke	OS19-2	21/79
Wei	Qiang	OS12-2	37/63
Wen	Haokang	OS12-7	37/64
		OS12-9	38/64
		OS12-10	38/65
		OS12-11	38/65
		OS13-7	36/69
		OS13-9	36/69
Wu	Chi-Che	OS7-2	32/53
Wu	Jiangyu	OS12-17	38/67
		OS13-14	37/71
Wu	Jun-Rong	OS8-2	23/55
Wu	Xiaowei	OS3-5	35/46
		OS3-9	35/47

[X]

Xiao	Chang-Sheng	OS8-5	24/57
Xiao	Zhiqing	OS12-18	38/67
		OS13-14	37/71

Xu	Junxiang	OS17-1	31/77
[Y]			
Yagi	Kazuhiro	GS3-3	27/93
Yamaba	Hisaaki	OS16-3	30/76
		OS16-4	30/76
		OS16-5	30/77
		OS24-3	25/88
Yamada	Katsuaki	GS5-1	26/94
Yamada	Takeshi	OS10-6	24/61
Yamamoto	Hidehiko	GS5-1	26/94
		GS5-2	26/94
Yamamoto	Toru	OS10-2	24/60
		OS10-3	24/60
		OS10-4	24/60
		OS10-5	24/60
		OS10-7	24/61
Yamane	Ryunosuke	GS5-2	26/94
Yamanobe	Natsuki	OS11-1	29/61
Yamawaki	Junichiro	OS2-4	22/45
Yan	Yuqi	OS12-3	37/63
Yanagise	Kentaro	OS22-4	31/84
Yang	Bo-Jung	OS8-5	24/57
Yang	Chunxin	GS6-4	33/96
Yang	Dezhi	OS3-3	34/46
Yang	Jialin	OS12-3	37/63
Yang	Zhou	OS3-2	34/45
		OS3-7	35/47
Yao	Chun-Jen	OS9-4	22/58
Yasukawa	Shinsuke	OS22-2	31/83
		OS22-5	31/84
		OS23-1	27/85
		OS23-4	27/86
Ye	Zhan-Xian	OS8-4	23/56
Yin	Di	OS13-10	36/70
		OS13-11	37/70
Yin	Mengqi	OS13-10	36/70
Yokkampon	Umaporn	GS4-2	35/93
		GS7-3	40/97
Yoshimura	Motohide	OS5-6	29/50
Yoshitomi	Yasunari	PS2	28/41
		OS2-2	22/44
		OS2-3	22/44
		OS2-4	22/45
Yuan	Yasheng	OS13-9	36/69
		OS13-10	36/70
		OS13-11	37/70
[Z]			
Zhang	Qianqian	OS12-7	37/64
		OS13-7	36/69
Znang	Tianyi	OS12-5	37/63
		OS12-6	37/64
		OS13-3	36/68
		OS13-4	36/68
Zhang	Weicun	OS6-3	30/51
		OS6-5	30/52
Zhang	Yuheng	OS3-4	35/46
		OS3-6	35/46
		OS3-7	35/47
		OS3-8	35/47
Zhao	Jichao	OS12-7	37/64
		OS13-1	36/67
		OS13-7	36/69
		OS13-12	37/70
		OS13-13	37/71
Zhao	Jikai	OS3-3	34/46
		OS3-5	35/46
Zhao	Huailin	OS12-1	37/62
Zhao	Lianchen	OS3-4	35/46
		OS3-6	35/46
Zhao	Yuqi	OS3-5	35/46
		OS3-9	35/47
Zheng	Fulin	OS12-12	38/65
Zheng	Qiang	OS12-11	38/65
Zhou	Ye	OS14-5	28/73
Zhou	Yusong	OS3-9	35/47
Zhu	Meiyu	OS6-5	30/52

Zhu	Yuxuan	OS13-10	36/70
		OS13-11	37/70
Zhumanazarova	Assiya	OS4-3	34/48

Tomato-Harvesting-Robot Competition Towards Smart Agriculture

Kazuo Ishii[1], Takayuki Matsuo[1,2], Yasunori Takemura[1,3], Takashi Sonoda[1,3], Yuya Nishida[1], Shinsuke Yasukawa[1], Takuya Fujinaga[1]

[1]Center for Socio-Robotic Synthesis, Kyushu Institute of Technology, 2-4, Hibkinino, Wakamatsu Kitakyushu, Fukuoka 808-0196, Japan

[2]National Institute Technology, Kitakyushu College 2-20-1, Shii, Kokuraminami-ku, kitakyushu-shi, Fukuoka, Japan

[3]Department of Engineering, Nishinippon Institute of Technology, 1-11 Aratsu, Kanda, Miyakogun, Fukuoka 800-0394, Japan

E-mail: ishii@brain.kyutechg.ac.jp, matsuo@kct.ac.jp, takemura@nishitech.ac.jp, sonoda@nishitech.ac.jp, y-nishida@brain.kyutech.ac.jp, s-yasukawa@brain.kyutech.ac.jp, fujinaga.takuya835@mail.kyutech.jp[1]

Abstract

In agriculture, the aging and depopulation of farmers cause the shortages of farmers and manpower. Most of commercialized robots are industrial robots for factory automation, and most of robots for the first industry, agriculture, forestry and fisheries are still under developing. The reasons are cost-efficiency of the robotization, safety of the works using robots, difficulty of outdoor operations, and knowledge transfer problem from farmers to computer, etc. As one of solutions for the problems, robot technology into the agriculture is expected to contribute to the laborsaving, improvement of production, production line automation, and also the management toward smart-agriculture. We organize the Tomato-harvesting-robot competition to offer the research field and welcome researchers into agricultural robotics.

Keywords: tomato harvesting, agriculture robot, smart agriculture, competition.

1. Introduction

In agriculture, a decrease in the number of farmers, an aging population, and a shortage of successors are significant problems. Ministry of Agriculture, Forestry and Fisheries of Japan reported that Japanese self-sufficiency ratio for food is about 40 percent, which is lowest level among developed countries. For such urgent problems, smart agriculture is expected to be an innovative method of utilizing robot technology and information communication technology, and researches aiming to realize smart agriculture have been carried out such as the autonomous operation of agricultural machinery, automation of harvesting, and monitoring of the field environment and crops, etc. Most of commercialized robots are for factory automation of automobile, and most robots for agriculture, forestry and fisheries are under developing. The reasons for the late social implementation are cost-efficiency of the robotization, safety of the works using robots, difficulty of outdoor operations, and knowledge transfer problem from farmers to computer, etc. If we can overcome these difficulties and implement the robots into agricultural fields, robots can contribute to the laborsaving, improvement of production, production line automation. Also, the management of agricultural products such as quality, quantity, and condition of environment become possible and the smart-agriculture will be realized.

For the autonomous operation of agricultural machinery, Noguchi proposed a robot farming system [1] that consists of intelligent robot vehicles to automate farming activities from planting to supplying products to consumers and described the importance of simultaneous operations by developing a multiple robot system. For the automation of harvesting, Kawamura et. al. developed a mobile robot with manipulator and proposed the tomato harvesting method using image processing and visual feedback [2][3]. Kondo et. al. proposed the method to improve the success ratio of tomato harvesting and speed-up technique [4][5]. Ota et. al. proposed the path planning method to pick up tomato [6]. Yaguchi et al. developed an end-effector that grasped a fruit using grippers and plucked it from the separation layer in the peduncle [7]. They described the results of harvesting experiments conducted on an actual farm, demonstrating that the harvesting time was shortened by improving the harvesting motion [8]. Yoshida et al. developed a robot that har- vested a cluster of cherry tomatoes [9][10]. They focused on detecting the peduncles that the robot required when harvesting the clusters. For the monitoring of the field environment and crops, Fukatsu et al. described a remote monitoring method that responded flexibly and dynamically to changes in the field environment for long-term field monitoring [11] which consists of field servers, small monitoring sensor nodes. Fukui et al. developed a robot that estimated the volume of tomato fruits to create a database of fruit growth [12] by detecting fruits based on a human attention model, saliency [13].

Tomato is one of important fruit vegetables and most tomatoes are produced in the greenhouses, or large-scale farms, where the high temperature and humidity, and long harvest age force the farmers heavy works. With an aim to promote the automation of tomato harvesting, we have organized the tomato harvesting robot competition. In this paper, we introduce the Tomato-Harvesting-Robot Competition.

2. Tomato-Harvesting-Robot Competition

The Tomato-Harvesting-Robot competition started in 2014 and 7th competition was held in 2020, which consists of two leagues, the Senior League and the Junior League. The target competitors for Senior League are supposed to be the teams with automated robots, and the Junior League are for high school or junior high school students, whom are asked to build robots with LEGO Mindstorms.

2.1 Senior League

The Senior League supposes that teams composed of undergraduate, graduate students or developers join the league with tomato harvesting robots, and compete the accuracy and speed of harvesting. The two kinds of competition field are designed, the one is the rail-style area and another is the free-style area. The rail-style area is designed to have the similar environment with the large-scale tomato factory, where pairs of pipes are arranged on the ground to control the temperature of the greenhouse, and pipes are also used as the rails for the platform trucks on which the workers sit down, harvest and carry tomatoes. The diameter of pipes is 50 mm and the distance between pipes is 600 mm. The position of tomato is always adjusted between 800 to 1200 mm heights as the tomato plants are growing up. The free-style area is for the robots of general tomato fields in outdoor environment in future. In the first to 6th competitions, artificial grasses are put on the field with the sizes of 3600 x 3600 mm instead of soils. The setting of tomato position is the same with the rail-style area.

The required specifications for robots are shown in Table 1. The projected area of the robot on the ground is within the 800 x 800 mm square including a box for storing tomatoes, and no height limitation. The robots should have an emergency stop switch on the easy-to-find location of robots. As recommendations, the weight of the robot is less than 50 kg and the electric power of each motor is less than 70 W.

The robots are classified mainly two types, manual control and autonomous control, and the former robots are classified by whether the operator observes tomato directly or indirectly using cameras mounted on robots, and by robot locomotion whether the robot uses rail or not. Totally, the robots are categorized into 6 types depend on operation and locomotion method as shown in Table 2. The success points for a tomato change depending on robot categories. the point for one tomato of category (1) manual control, direct tomato observation, and rail-type is 2, and that of category (6) autonomous control, free-type is 16.

The senior league consists of 3 stages as shown in Fig.2. In first stage, the basic function of the robot is evaluated whether the robot can move, have a manipulator and an end-effector, and recognize tomato. Single tomatoes are hanged from a bar and the robot that can touch a tomato proceeds to the second stage.

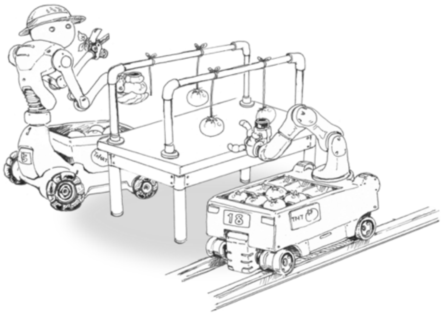
In the second stage, the robot harvests a tomato from a tomato cluster. If the robot gives damage to tomato, get -1 point for each damage tomato, and the damage to tomato plant is big minus points. And the target tomatoes should be redder than the reference tomato indicated by the committee. In the final stage, the robots harvest tomato from plant body as shown in Fig. 3, for the 1st competition.

Table 1 Required Specifications for Robots

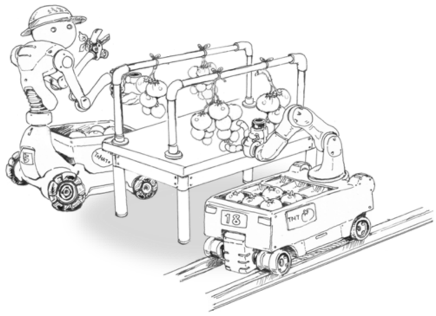
Requirement	Max. Size	W:800mm D:800mm No height limitation
	Safety	Emergency Stop Switch
Recommend	Max. Weight	Max. 50kg
	Max. Motor Power	Max. 70W each

Table 2 Robot Classification by Operation Method

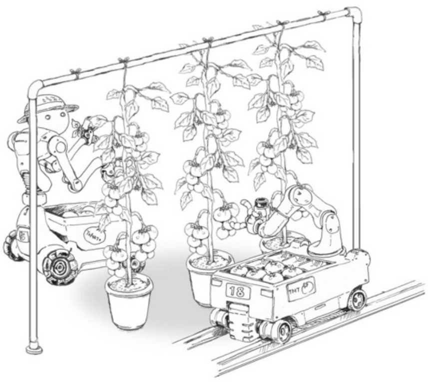
Robot Control	Manual				Autonomous	
Operator Manipulation	Direct Observation of Tomato		Indirect Observation of Tomato		Start and Stop commands	
Mode of Locomotion	Rail-type	Free-type	Rail-type	Free-type	Rail-type	Free-type
Category	(1)	(2)	(3)	(4)	(5)	(6)



(A) FIRST STAGE: PICK UP SINGLE TOMATO



(B) SECOND STAGE: PICK UP A TOMATO FROM TOMATO CLUSTER



(C) FINAL STAGE: PICK UP A TOMATO FROM TOMATO PLANT

Fig. 1 The three stages of the senior league.

Fig. 2 The overview of the final stages of the senior league in the 1st competition

2.1. *Ju 2.2 Junior League*

Figure 3 shows the concept of Junior League. We set the subjects to make the model of an autonomous transport system of tomatoes harvesting, such as the transportation and sorting of harvested tomatoes. Competition subjects are shown below.

(1) Line Trace

The robot moves along the white course line using the color sensor. In the harvest field, Tomato boxes are arranged. The robot can find three RGB colored tomato boxes if moves along the line.

(2) Color Identification

The robot explores the area and recognize color signs along course line, and find three RGB colored tomatoes in boxes (tomato box).

(3) Mechanism Design and Control

To pick up the tomato boxes, participants are expected to design and make a robot arm to get a tomato box. The robot is required to store, transport and relocate the boxes to the assigned storage area. After picking up the box, the robot is to return to the course line. Then, the robot carries the box to the specific storage location.

(4) Object detection

After putting the tomato boxes to the storage locations, the robot should return to the course line. Finally, the robot goes the charging station which is put at the position with a certain distance away from the course line. The robot must find the stop position by the distance measuring sensor.

Each team consists of 3 or 4 students and makes a robot using LEGO Mindstorms. The basic specification for robot is that the sizes of the robot are within 300 x 300 [mm] on the ground. The robot can have structural change after starting, so that the initial robot structure is regarded as the robot size.

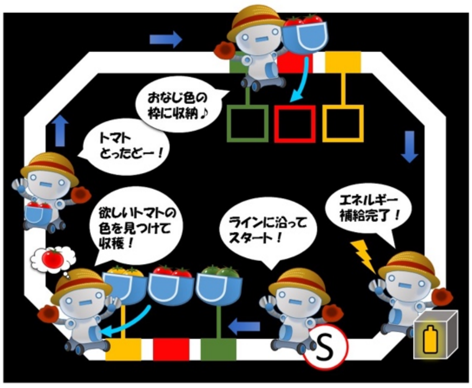


Fig. 3 The concept of Junior League

3. Regulation Changes toward Smart Agriculture

The regulation of the competition is changed by evaluating the achievement of teams toward actual greenhouse toward smart agriculture.

Table 3 shows the history of regulation changes. The first competition was held in 2014 with Senior League. From the second competition, Junior League started as the outreach activity of agricultural robotics. From the third competition, a bump was put on the free-style area to motivate the improvement of mobility for rough terrain. In the fifth competition, the black board shown in Fig.2, which is placed for making the image processing easy, was removed and the robot are asked to find tomatoes in various colors. From the 7th competition in 2020, the working field moved to an experimental greenhouse, so that the robot need the color consistency algorithm against sunlight environment. Junior league was held as in each school via internet.

The winner teams are shown in Table 4. The first winner robot from team JSK, the Univ. of Tokyo (Fig.4) joined the category (T-4) and the robot was operated by human in the remote place and succeeded to have 5 tomatoes in 10 min. 2 teams joined to autonomous categories could not harvest a tomato. In the 6th competition, the robot of Kyutech team whose category is T-5 autonomous robot for the rail-style field (Fig.5), succeeded to harvest 13 tomatoes in 10 min in the 2nd stage and 5 tomatoes in the final stage. The teams selected for the final stage of 7th competition, all teams entries for the category T-5, autonomous robot in the rail-style field. The winner is the team of National Institute Technology, Kitakyushu College (Fig.6) succeeded to take about 20 tomatoes in the greenhouse. The performance of each team grows up to harvest tomatoes autonomously in spite of the lighting condition changes. Moreover, some journal papers are published based on the results of the competitions such as [7][8][14]15].

Table 3 History of regulation changes

N-th	Year	Rule & Changes
1	2014	Senior League started with 6 categories from T1 (manual operation) to T6 (autonomous) and 2 kinds of working fields (rail and artificial grass)in indoor field (Gymnastium). The black boards are set behind tomato plants.
2	2015	Senior: No rule change. Junior: Junior League started. Lego mindstorm is used for basic platform.
3	2016	Senior: Bump is placed in the center of grass field. Junior: No rule chage.
4	2017	No rule change.
5	2018	Senior: Slope is placed instead of Bump in the grass field. The black boards are removed. In scoring, success rate is added. Junior: Original arm made of stationery is allowed. Tomato box removed.
6	2019	No rule change.
7	2020	Senior: Video evaluation instaed of 1st and 2nd rounds. The cometition field is changed to the ourdoor experimental green house. The grass field is changed to the soil field. Junior: Online cometition in each school.

Table 4 Results of competition and Winner teams

N-th	Year	Senior Teams	Junior Teams	Winner	2nd Place
1	2014	9	-	Inaba-Okada Lab., Univ. of Tokyo	National Institute of Technology, Kitakyushu College
2	2015	14	16	Hayashi Lab., Kyutech	Team SS
3	2016	12	23	Kyushu Polytech	NAIST
4	2017	12	18	Hayashi Lab., Kyutech	Ishii Lab., Kyutech
5	2018	9	23	Hayashi Lab., Kyutech	Kyushu Polytech
6	2019	9	25	Ishii Lab., Kyutech	NAIST
7	2020	8	13	National Institute of Technology, Kitakyushu College	Hayashi Lab., Kyutech

Fig.4 The winner robot of the 1st competition, JSK team from Univ. of Tokyo.

Fig. 5 The winner robot of the 6th competition, Kyutech team. The robot harvested 5 tomatoes in the final stage.

Fig.6 The winner robot of the 7th competition, National Institute of Technology, Kitakyushu College. The robot harvested 20 tomatoes in the experimental greenhouse.

4. Summary

We have organized the Tomato-Harvesting-Robot competition to promote agricultural robots and provide the test field for tomato harvesting. The performance of robots are growing up gradually and the regulation of the competition is revised to be closed to the actual working environment. Currently, the winner robot can harvest a tomato within 30 seconds in the sunlight condition. In the next, we expect to have robots working in the soil field and in the night time.

Acknowledgements

The Tomato-Harvesting-Robot competition is supported by FAIS, Kitakyushu-city, Hibikinada Green Farm, and YASKAWA Electric Corporation.

References

1. N. Noguchi, “Agricultural Vehicle Robot,” J. Robot. Mechatron., Vol.30, No.2, pp. 165-172, 2018.
2. Kawamura et. al., Study on agricultural robot (1st report), The journal of JASM , vol. 46 No.3,pp.353-358, 1984 (in Japanese).
3. Kawamura et. al., Study on agricultural robot (2nd report), The journal of JASM, vol.47 No. 2, pp.177-182, 1985 (in Japanese).
4. Kondo, et.al., A Machine Vision System for Tomato Cluster Harvesting Robot, Engineering in Agriculture, Environment and Food, Vol. 2, No. 2, pp.60-64, 2009.
5. Kondo, et.al, Development of an End-Effector for a Tomato Cluster Harvesting Robot, Engineering in Agriculture, Environment and Food, Vol. 3, No. 1, pp.20-24, 2010.
6. Ota et. al., Development of a Tomato Harvesting Robot with a Vision System Using Specular Reflection (Part 1), the journal of JASM, vol. 72 No.6, pp.595-603, 2010.
7. H. Yaguchi, K. Nagahama, T. Hasegawa, and M. Inaba, “Development of An Autonomous Tomato Harvesting Robot with Rotational Plucking Gripper, 2016 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 652-657, 2016.
8. H. Yaguchi, T. Hasegawa, K. Nagahama, M. Inaba, A Research of Construction Method for Autonomous Tomato Harvesting Robot focusing on Harvesting Device and Visual Recognition, J. Robot. Soc. Jpn., Vol. 36, No.10, pp. 693–702, (in Japanese with English summary), 2018.
9. T. Yoshida, T. Fukao, and T. Hasegawa, Fast Detection of Tomato Peduncle Using Point Cloud with a Harvesting Robot, J. Robot. Mechatron., Vol.30, No.2, pp. 180-186, 2018.
10. T. Yoshida, T. Fukao, and T. Hasegawa, Cutting Point Detection Using a Robot with Point Clouds for Tomato Harvesting, J. Robot. Mechatron., Vol.32, No.2, pp. 437-445, 2020.
11. T. Fukatsu and M. Hirafuji, Field Monitoring Using Sensor-Nodes with a Web Server, J. Robot. Mechatron., Vol.17, No.2, pp. 164- 172, 2005.
12. R. Fukui, K. Kawae, and S. Warisawa, Development of a Tomato Volume Estimating Robot that Autonomously Searches an Appropriate Measurement Position – Basic Feasibility Study Using a Tomato Bed Mock-Up –, J. Robot. Mechatron., Vol.30, No.2, pp. 173-179, 2018.
13. L. Itti et al., A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.20, No.11, pp. 1254-1259, 1998.
14. Takuya Fujinaga et al., Image Mosaicing Using Multi-Modal Images for Generation of Tomato Growth State Map, J. Robot. Mechatron., Vol.30, No.2, pp. 187-197, 2018.
15. Takuya Fujinaga, et al., Tomato Growth State Map for the Automation of Monitoring and Harvesting, J. Robot. Mechatron., Vol.32, No.6, pp. 1279-1291, 2020.

Convolutional Network with Sub-Networks

Ninnart Fuengfusin
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0196, Japan
Hakaru Tamukoh
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0196, Japan
E-mail: fuengfusin.ninnart553@mail.kyutech.jp, tamukoh@brain.kyutech.jp
https://www.brain.kyutech.ac.jp/~tamukoh/

Abstract

We propose a convolutional network with sub-networks (CNSN), a convolutional neural network (CNN) that can be detached into sub-models on fly. Due to a conventional design of CNN, shapes of feature map are varied throughout the model. Therefore, the hidden layer within CNN may not directly process the input image without any modifications. To address this problem, we propose a *step-down convolutional layer,* a convolutional layer which acts as an input layer for the sub-model. The *step-down convolutional layer* reshapes the input image to a preferred representation to the sub-model. To train CNSN, we treat the base-model and sub-models as different models. We separately forward- and back-propagate each model. By using *multi-model loss*, a linear combination of losses from base- and sub-models, we can update weights that can be utilized in both base- and sub-models.

Keywords: Convolutional Neural Networks, Supervised Learning, Model Compression.

1. Introduction

In recent years, a convolutional neural network (CNN) has been achieved state-of-art results in varied tasks from an image recognition[1], sematic segmentation[2], and object detection[3] tasks. However, when comparing the CNN with a multilayer perceptron (MLP) with same amount of weight parameters, the down-side of CNN is it significantly utilizes more multiply–accumulate operations (MAC) than MLP. This may lead into a bottleneck when deploying a CNN with a mobile device that has lower computational capacity. To address this problem, there are a several research directions attempt to address this problem from a structure pruning[4], network sliming[5], and early-exit network[6].

In this research, we focus on the network sliming direction. The network sliming focuses on create a neural network model that can be detached into smaller models during inference. In our previous work in this direction, we purposed Network with Sub-Networks[7] (NSN) that introduces a layer-wise detach ability to MLP. By detaching weight layers, this allows NSN to operate with less amount of MAC, therefore it decreases an overall latency of model with the trade-off in the model performance. However, to apply the methods in NSN to CNN, there are three compatibility issues. The first issue is NSN requires a targeted model to contain a same shape of feature map throughout the model. This creates a major constraint to the CNN model. The CNN cannot contain any spatial reduction layers, for example pooling layers or convolutional layers without padding. Without any spatial reduction layers, this may lead the CNN to operate with higher MAC than the layer detaching can reduce. The second issue is NSN requires to train *N+1* models in parallel, where *N* is a number of hidden layers. This indicates that NSN method is not feasible with the recent CNN models that contains more than hundred hidden layers. The last issue is NSN require to manually control the gradients across base- and sub-models. This causes the high complexity in programming.

In this research, we propose CNSN to address all mentioned issues from our previous work. CNSN allows the usage of different shapes of feature map by attaching a *step-down convolutional layer* as the input layer of sub-models. This layer pre-processes the input image to a preferred representation to the sub-model. To allow CNSN to operate more depth CNN models, CNSN reduces a number of sub-models by allowing only few selected sub-models to be detached with. Instead of manually control the gradients, we introduce a *multi-model loss*, the combination of losses from a base-model and sub-models. This loss allows the base model to be able to detached into sub-models during inference and reduces overall complexity to the user comparing with NSN.

By detach convolutional layers on demand, CNSN promises the greater reduction in MAC than NSN which can detach only fully connected layers. Our main contributions in this research are listed as follows:

- We propose CNSN, a CNN with ability to detach a base-model into sub models.
- We introduce *multi-model loss*, a combination of losses from base-model and sub-models.
- We introduce a *step-down convolutional layer* that acts as the input layer to sub-models.

2. Convolutional Network with Sub-Networks

CNSN consists of a base-model and several sub-models. The sub-model is a subset of the base-model with an exception, *step-down convolutional layer* that is not included in the base-model. Therefore, all parameters from sub-models except the *step-down convolutional layer* is identical to the base-model. In term of memory usage, we only required to store the base-model and several *step-down convolutional layers.*

Overview of CNSN is illustrated in Fig. 1. Fig. 1 illustrates a base-model at the top row and two sub-models at the two bottom rows. Each base- and sub-models can directly receive the input image and independently product the prediction to other models. With less amount of parameters in the model, the sub-model has less capacity comparing to the base-model. However, with the same reason, it promises to perform the inference faster than the base-model.

With both base- and sub-models, CNSN consists of the three components to make CNSN operate-able with CNN. These three components are as follows: *step-down convolutional layer*, *sub-models*, and *multi-losses*.

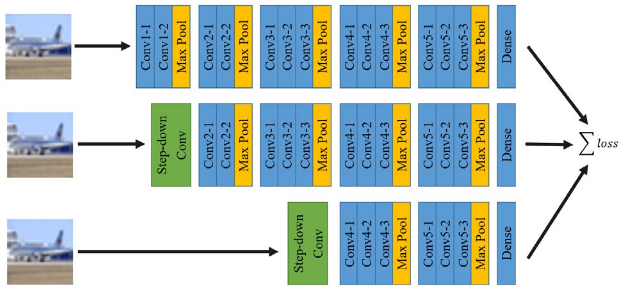


Fig. 1. Overview of CNSN. The top row represents a base model of CNSN while two bottom rows represent the sub-models. Each model can receive the image and produce the prediction. During training, losses from all model are collected as *multi-model loss.*

2.1. *Step-down Convolutional Layer*

The *step-down convolutional layer* is a convolutional layer that is designed to solve the spatial difference between the input and hidden layer in the CNSN. By solving this problem, this allows CNSN to reuse a hidden convolutional layer of base model as an input layer of sub model. However, to achieve this, there are some concerns as follows.

The first problem is the number of input filters of hidden layer may not be a same comparing to the input layer. With this issue, the hidden layer cannot process the input image. To solve this problem, we can place the *step-down convolutional layer* with the same amount of input filters to the input layer. Another issue is the expected size of the input feature is difference between base- to sub-models. By adjusting the number of stride and size of kernel of *step-down convolutional layer*, we can adjust the shape of output feature maps to be process further in the sub-model.

2.2. *Sub-models*

Instead of allowing every layer to be detach-able as in NSN, CNSN only allows few sub-models to be detached. In each batch trainings, both NSN and CNSN requires base- and sub-models to be propagated in the same iteration. This causes the problem that is parameters from base- and sub-models may not be fitted in a single GPU if the base-model is large and the base-model contains many sub-models. Therefore, to able to train on the bigger CNN with CNSN, we reduce the number of detach ability to few sub-models instead.

2.3. *Multi-model loss*

To allow both base-model and sub-models to operate as an individual model, we purpose a method called *multi-*

model loss. The *multi-model loss* is designed to balance between the loss from base-model and sub-models. By assume there are *N* sub-models in the base-model, the *multi-model loss* or l_{all} is formulated as shown in Eq. (1). Where l_N is the loss from a *N* sub-model, and l_{base} is a loss from the base-model. To utilize this loss, all models must be forward propagated in the same batch training first to find l_N and l_{base}. Then, we can use l_N and l_{base} to find l_{all}.

$$l_{all} = \sum_{1}^{N} l_N + l_{base} \tag{1}$$

3. Experimental results and Discussion

We implemented our CNSN using VGG-16[8]-like model as the base-model. Since VGG-16 is designed for ImageNet Large Scale Visual Recognition Challenge[9] (ILSVRC) 2014, some layers in VGG-16 are required to be modified to able to operate with images from CIFAR10[10]. We removed first two fully-connected layers and modified the last fully-connected layer to contain 512 input neurons and ten output neurons. To stabilize the training process, we also attach a batch normalization[11] (BN) layers after each weight layers except on the last fully-connected layer. We utilized ReLU as an activation function except the last fully-connected layer that we assigned with the log-softmax instead.

In this experiment, the overall setting is illustrated in Fig. 1. We assigned with two sub-models within the base-model. Without including any of activation, BN, and pooling layer, the first sub-model or *sub-model0* is the base-model after removing first two convolutional layers. The second sub-model, *sub-model1* is the base-model after removing first seven convolutional layers. The *step-down convolutional layer* for *sub-model0* is the convolutional layer with stride two and kernel size two. For *sub-model1*, it is the convolutional layer with stride eight and kernel size eight. We defined the *baseline model* as a VGG-16-like model that was trained without any modifications.

We conducted a benchmark with CIFAR10 dataset. CIFAR10 consists of 10 different classes. Each class consists of 5,000 training and 1,000 test images. We performed the data augmentations by padding 4 pixels into the training image and randomly crop back to original size. The training images were further augmented by horizontal flipping and normalizing with a channel-wise mean and standard deviation of CIFAR10 dataset. We trained all models with stochastic gradient descent with momentum of 0.9. We set an initial learning rate as 10^{-2} and step-down to one-tenth after we trained for 50, 100 and 150 epoch. We warmed-up the learning rate for an epoch and trained for totally 300 epochs using the training batch size as 32. We reported the best test accuracy that occurred during the training.

The experimental results are as shown as Table 1. In Table 1, we compared the CNSN base model with the *baseline model* with the same setting as CNSN except for the weight decay. We found out the optimized weight decays are differed between the *baseline model.*
and CNSN. Therefore, we applied the different weight decay to each model.

Table 1. Results of CNSN models comparing a base-line model on CIFAR10 dataset.

	Test accuracy	MAC	Weight decay
sub-model0	0.7738	**0.124 G**	6×10^{-4}
sub-model1	0.9115	0.275 G	
base-model	0.9267	0.314 G	
baseline model	**0.9316**	0.314 G	6×10^{-3}

Our *base-model* achieved the loss in test accuracy for 0.0049 in exchanging to the ability to detach into sub-models. The *sub-models1* able to reduces more than half of MAC comparing with the *baseline model*, however this came with the significantly drop in term of test accuracy.

4. Conclusion

We propose CNSN, a CNN that can be detached into smaller CNNs on fly. To gain the detach ability to CNN, we propose the *multi-model loss* and *step-down convolutional layer.* The base-model of CNSN can deliver the performance that is compare-able with the regular trained models, while sub-models significantly reduces the amount of MAC, however with the trade-off in loss in test accuracy.

Acknowledgements
This research was supported by JSPS KAKENHI Grant Numbers 17K20010.

References

1. Tan, Mingxing, and Q. V. Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." *arXiv preprint arXiv:1905.11946* (2019).
2. Chen, Liang-Chieh, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. "Encoder-decoder with atrous separable convolution for semantic image segmentation." In *Proceedings of the European conference on computer vision (ECCV)*, pp. 801-818. 2018.

3. Tan, Mingxing, Ruoming Pang, and Quoc V. Le. "Efficientdet: Scalable and efficient object detection." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10781-10790. 2020.
4. Li, Hao, Asim Kadav, Igor Durdanovic, Hanan Samet, and Hans Peter Graf. "Pruning filters for efficient convnets." *arXiv preprint arXiv:1608.08710*, 2016.
5. Yu, Jiahui, Linjie Yang, Ning Xu, Jianchao Yang, and Thomas Huang. "Slimmable neural networks." *arXiv preprint arXiv:1812.08928*, 2018.
6. Teerapittayanon, Surat, Bradley McDanel, and Hsiang-Tsung Kung. "Branchynet: Fast inference via early exiting from deep neural networks." In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 2464-2469. IEEE, 2016.
7. Fuengfusin, Ninnart, and Hakaru Tamukoh. “Network with Sub-Networks.” Proceedings of International Conference on Artificial Life and Robotics, vol. 25, 2020, pp. 191–194.
8. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556*, 2014.
9. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M. and Berg, A.C., 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision*, *115*(3), pp.211-252.
10. Krizhevsky, Alex, Vinod Nair, and Geoffrey Hinton. "The cifar-10 dataset." *online: http://www.cs.toronto. edu/kriz/cifar. html* 55 (2014).
11. Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *arXiv preprint arXiv:1502.03167*, 2015.

Influence of FPGA implementation methods in High-Level Synthesis

Yusuke Watanabe
CRAFT WORK Co.,Ltd,
5F OS Bldg 3-5-15 Shibasaki-cho, Tachikawa, Tokyo, 190-0023, Japan

Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0196, Japan

Hakaru Tamukoh
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0196, Japan
E-mail: watanabe.yusuke898@mail.kyutech.jp, tamukoh@brain.kyutech.jp
http://www.lsse.kyutech.ac.jp/english/

Abstract

We explain about how the difference of implementation methods written in C++ in High-Level Synthesis (HLS) influences on latency for tiny You only look once (YOLO) v2, a real-time object detection system to infer on an FPGA. To utilize features of FPGA, we need to implement hardware-oriented algorithms such as the weight binalization. We primarily focus on convolution in tiny YOLO v2 network and we report execution results on the Xilinx SDSoC development environment to know whether methods are appropriate or not.

Keywords: Convolution, FPGA, High-Level Synthesis, Hardware-Oriented Algorithm

1. Introduction

When a robot which power supply is generally limited detects objects, a deep neural network is often used. FPGAs are good choices to implement neural networks while reducing energy consumption. To reduce implementation time in FPGAs, High-Level Synthesis (HLS) has been used recently. It automatically creates digital hardware from C++ source codes and we don't need to use time-consuming hardware description languages. But typical C++ implementation is not hardware-oriented and therefore created hardware don't bring results as we would think.

In this paper, we experiment with some convolution operations in an object detection algorithm to utilize FPGAs better in HLS.

In conclusion, when we use HLS, hardware-oriented implementation is preferable and brings dramatically improved hardware.

2. Method

We explain about our implementation methods of convolution operations which occupy much of processing latency in object detection. As an object detection system, we use binarized tiny YOLO v2 which binarizes original tiny YOLO v2 input and weights data according to Ref. [1]. Convolution operations are executed in three dimensions of height, width and input channels. Due to input and weights data binarization, we can use bitwise operations which reduce FPGA resources utilization and latency and therefore can be regarded as hardware-oriented.

Eventually we use them as convolution operations and experiment with the following four methods.

(1) no bitwise operation
(2) width dimension
(3) channel dimension
(4) mix of width and channel dimension

The first method doesn't use bitwise operations. The second one uses bitwise operations along the width dimension and the third one along the channel. The last one is a mix of the second and the third. We apply bitwise operations to the dimension which has the largest value among height, width and input channel. We don't try along a height dimension because height values are always smaller than width ones.

3. Experiment

To experiment, we prepare for a specific input RGB image file which width and height are 384 and 288 pixels respectively. HLS was executed on 2018.3 release of Xilinx SDSoC Development Environment. Our targetFPGA board is a Zynq UltraScale+ MPSoC ZCU102 evaluation kit. We used automatically created synthesis reports to evaluate the performance of our implementation.

Table 1 shows the clock cycles to produce output. Differences between min and max come from the presence of conditional branches in source codes. Table 2 shows resources used to implement the binarized tiny YOLO v2 design in percentage.

Table 1. Latency comparison.

method		min	max
no bitwise op.		2,546,563,540	3,072,073,924
bitwise op.	width dimension	471,558,114	4,811,180,162
	channel dimension	159,132,162	3,695,079,650
	mix of width and channel dimension	115,836,258	4,634,795,522

Table 2. Resources usage comparison

method		BRAM	DSP	FF	LUT
no bitwise op.		1,364	67	31,355	54,777
bitwise op.	width dimension	328	22	22,388	42,965
	channel dimension	328	10	274,116	75,125
	mix of width and channel dimension	328	15	181,935	79,569
Available		1,824	2,520	548,160	274,080

4. Results

As for latency, if we use bitwise operations, the min clock cycles become lower and the actual execution time on the FPGA board also becomes shorter although max clock cycles become higher. This fact means that most processing doesn't take the max clock cycles paths. Further if we apply bitwise operations to the mix of width and channel dimension as shown in Table 1, clock cycles reduction becomes higher.

About resources, if we use bitwise operations, we can substantially reduce BRAM usage. If a network becomes deeper, we use more BRAM and BRAM shortage is likely to happen. It is important for BRAM not to get wasted.

5. Conclusion

We conclude that hardware-oriented algorithms like bitwise operations in this paper are essential even when we design hardware from C++ using HLS. This is because HLS environments doesn't automatically create the best hardware yet. There are challenges to exactly grasp which implementation brings better hardware. So we need to become used to how to write hardware-oriented codes in C++. Actually our current implementation is far from the best and we need to make it more hardware-oriented.

References

1. Courbariaux, Matthieu & Hubara, Itay & Soudry, Daniel & El-Yaniv, Ran & Bengio, Y.. (2016). Binarized Neural Networks: Training Deep Neural Networks with Weights and Activations Constrained to +1 or -1.
2. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.
3. Nakahara, Hiroki & Yonekawa, Haruyoshi & Iwamoto, Hisashi & Motomura, Masato. (2017). A Batch Normalization Free Binarized Convolutional Deep Neural Network on an FPGA (Abstract Only). 290-290. 10.1145/3020078.3021782.

A Hardware-Oriented Random Number Generation Method and A Verification System for FPGA

Sansei Hori, Hakaru Tamukoh
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka, 808-0196, Japan
E-mail: hori-sansei@edu.brain.kyutech.ac.jp, tamukoh@brain.kyutech.ac.jp
https://www.kyutech.ac.jp

Abstract

Deep learning technology has made remarkable progress in recent years and has been applied to a variety of applications such as smartphones and cloud servers. These systems employ dedicated processors to save power consumptions and process massive data. In this paper, we introduce a hardware-oriented restricted Boltzmann machine and propose a field-programmable gate array (FPGA) infrastructure for easy verification of user circuits. The infrastructure makes it easy to communicate and control between the host PC and the user circuit.

Keywords: FPGA, Hardware Accelerator, Xillybus, Deep Learning, Random Number Generator, RBM

1. Introduction

Recent days, neural network technologies such as deep learning[1] have been utilized many applications such as image processing and natural language processing. Especially, these technologies have been actively applied to various embedded fields, including smartphone applications.

However, most training of deep neural networks (DNNs) require a massive amount of calculation resources and are often performed on high-performance computers with GPUs[2]. On the other hand, some embedded systems have restrictions such as power consumption and physical size to implement high-performance computers. Therefore, application-specific integrated circuits (ASIC) and system-on-a-chip (SoC) dedicated DNNs[3,4] is actively developed to accelerate the processing and reduce power consumptions. These some of the technologies have already yield practical applications.

Field-programmable gate arrays (FPGAs)[5] can also be used to create dedicated circuits and can be rewritten, making it possible to build more general-purpose systems. We have proposed resource-saving hardware implementation of a restricted Boltzmann machine (RBM)[6,7], which is a building block of Deep Belief Networks. However, the hardware implementation of the user circuit on an FPGA is costly work. In this paper, we introduce the hardware-oriented RBM, and we propose an FPGA infrastructure for verifying the modules implemented as IP in FPGA. Also, to verify the operation of the proposed FPGA infrastructure, we implemented a hardware circuit for learning an RBM[8,9] and confirmed that the circuit could be controlled from a host computer.

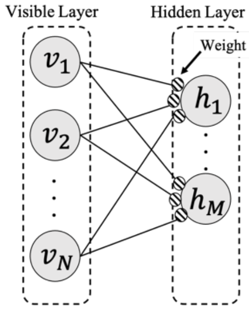


Fig. 1. Restricted Boltzmann machine.

2. Restricted Boltzmann machine

Restricted Boltzmann machine (RBM) is one of the generative model and a part of the element structures DNNs. An RBM has two layers called visible layer and hidden layer, as shown in figure 1. The probability distribution of an RBM calculated by Eq. (1), where $\boldsymbol{v}$ and $\boldsymbol{h}$ represent the states of visible and hidden units, and $\boldsymbol{\theta}$ is a set of the network parameters. Z($\boldsymbol{\theta}$) shown in Eq. (2) is a normalized term, and Φ shown in Eq. (3) is an energy function, where a_i and b_j represent the biases of visible and hidden units, and w_{ij} is the weight.

$$p(\boldsymbol{v},\boldsymbol{h}|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} e^{-\Phi(\boldsymbol{v},\boldsymbol{h},\boldsymbol{\theta})}. \quad (1)$$

$$Z(\boldsymbol{\theta}) = \sum_{\boldsymbol{v},\boldsymbol{h}} e^{-\Phi(\boldsymbol{v},\boldsymbol{h},\boldsymbol{\theta})}. \quad (2)$$

$$\Phi(\boldsymbol{v},\boldsymbol{h},\boldsymbol{\theta}) = -\sum_j a_i v_i - \sum_j b_j h_j - \sum_i \sum_j w_{ij} v_i h_j . (3)$$

The firing probabilities of the visible and hidden units of the RBM are calculated by the following equations, where σ(x) is a sigmoid function. This firing probability determines the state of each unit in the RBM.

$$p\left(h_j = 1 \middle| \boldsymbol{v},\boldsymbol{\theta}\right) = \sigma\left(b_j + \sum_i w_{ij} v_i\right). \quad (4)$$

$$p(v_i = 1|\boldsymbol{h},\boldsymbol{\theta}) = \sigma\left(a_i + \sum_j w_{ij} h_j\right). \quad (5)$$

3. Resource-saving random number generator

When training an RBM, it is necessary to sample each state from the firing probability of the visible and hidden

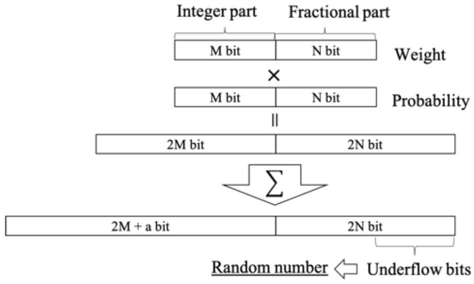


Fig. 2. Resource-saving random number generation method.

units. In this case, a large number of random number generators are required, but it is difficult to implement them in the hardware such as an FPGA. Therefore, the authors have proposed a method using truncated bits generated during fixed-point arithmetic operations as a substitute for random numbers[6,7].

In general, when some operations are implemented in digital hardware such as FPGAs, the various operations are implemented as fixed-point operations. In the operations, if there are variables with M bits and N bits in the integer and fractional parts, respectively, the result of multiplication of these values is 2M bits in the integer part and 2N bits in the fractional part, as shown in figure 2. Furthermore, the bit width of the integer part is increased by the addition process. When the result of this operation is stored in the register, the incremented bits are truncated. In the method, this truncated bit is used as a substitute for a random number.

4. Verification Infrastructure on FPGA

Figure 3 shows the configuration of a proposed user logic verification infrastructure. In this system, an FPGA is connected to the host PC via PCI Express bus to communicate and control a user logic. When configuring the FPGA, this system uses a joint test action group (JTAG). The user can connect the user logic by two AXI interfaces to verify the user logic. In this chapter, we describe the main components of this system.

4.1. *Interface between the host PC and the FPGA*

We use PCI Express bus to connect the host PC to the FPGA to control the user circuit and to communicate

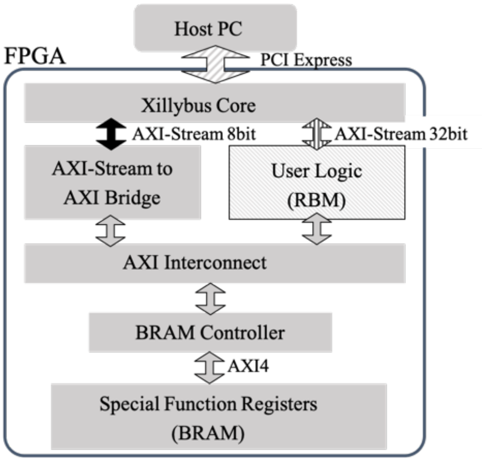


Fig. 3. User logic verification infrastructure.

data with the software on the host PC. This system applies Xillybus[10] to realize the PCI Express connection, which can convert the PCI Express data communication to first in first out (FIFO) or AXI-Stream. The host PC can access to the FPGA by reading or writing device files on its operating system.

4.2. *Internal bus and user logic interface*

This system applies AXI bus as an internal bus to connect all modules except the interfaces for Xillybus core. The user logic has AXI-Stream and AXI interface. The AXI-Stream interface connects to the Xillybus core directly, and this data path is used for data transfer. The AXI interface connects to AXI interconnect is used for the internal bus to control the user logic from the host PC through the special function registers (SFRs).

4.3. *Control registers*

The software running on the host PC can control and monitor the user circuits on the FPGA by accessing the control and status registers called SFRs. The user circuit on the FPGA can also access the SFRs and return its status, such as in-process or completed, to the software.

4.4. *AXI-Stream to AXI bridge*

The AXI-Stream to AXI bridge module extracts addresses and data from the 8-bit stream data sent from the Xillybus core to access the SFR. It is a necessary module to connect the core to the SFRs implemented in block RAM (BRAM) because the Xillybus core provides AXI-Stream interfaces.

5. Verification of the infrastructure

In order to verify the operation of the infrastructure, we synthesized an RBM using a conventional random number generators by Xilinx Vivado HLS, which is a high-level synthesis tool and implemented it as a user logic.

In this verification, the RBM trained the MNIST[11] dataset. The experimental conditions are listed below.

- Visible neuron: 784
- Hidden neuron: 150
- Without HLS optimization options
- Integer bit width: 14 bits
- Fraction bit width: 18 bits

Table 1 shows the resource utilization report of the RBM under the conditions. The target device is Xilinx Kintex 7 evaluation board KC705 (XC7K325T).

Table 1. Resource utilization report.

Resource	Utilization	Available	Utilization %
LUT	13202	203800	6.48
LUTRAM	580	64000	0.91
FF	16279	407600	3.99
BRAM	276.5	445	62.13
DSP	108	840	12.86
IO	5	500	1.00
GT	8	16	50.00
MMCM	2	10	20.00

We trained the RBM logic by the MNIST dataset on 25 epochs, and after the training, we obtained the weights of the RBM to verify the infrastructure. The weights are shown in Figure 4, and we can see the infrastructure is working in this application.

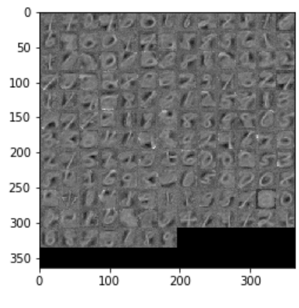


Fig. 4. Visualized weight parameter calculated on the FPGA infrastructure.

6. Conclusions and future works

In this study, we proposed and constructed a system that connects a host PC and an FPGA via PCI Express, and controls a user logic from software on the host PC. By using this system, the user does not need to construct the communication and control part by ownself except the module which user wants to verify, and the verification of the hardware-oriented system becomes possible more quickly.

Our future work is as follows. Firstly, to implement an RBM with the resource-saving random number generators in hardware, and to verify its operation in the verification infrastructure. Secondly, to connect external memory, such as an SDRAM to the verification infrastructure to be able to handle larger network parameters on the FPGA. In addition, since the user circuit is connected to the peripheral circuits only by AXI bus and AXI-Stream bus, we aim to create an environment in which the user circuit can be reconfigured while the peripheral circuits are running by utilizing the partial configuration technology. If this technology becomes available, the circuit can be verified and tested more easily and quickly. The goal of this project is to create an environment that enables easier and faster circuit verification and experimentation.

References

1. G.E. Hinton, S. Osindero, and Y.-W. Teh, *A fast learning algorithm for deep belief nets*, Neural computation, vol.18, no.7, pp.1527–1554, 2006.
2. NVIDIA, "NVIDIA Tesla V100 GPU Architecture", https://images.nvidia.com/content/volta-architecture/pdf/volta-architecture-whitepaper.pdf, 2017.
3. N.P. Jouppi et al. *In-datacenter performance analysis of a tensor processing unit*, Proceedings of the 44th Annual International Symposium on Computer Architecture. 2017. p. 1-12.
4. S. Wang, P. Kanwar, *BFloat16: The secret to high performance on Cloud TPUs*, https://cloud.google.com/blog/products/ai-machine-learning/bfloat16-the-secret-to-high-performance-on-cloud-tpus, 2019.
5. S. Trimberger, *A reprogrammable gate array and applications*, Proceedings of the IEEE, 1993, 81.7: 1030-1041.
6. S. Hori, T. Morie, and H. Tamukoh, *Restricted boltzmann machines without random number generators for efficient digital hardware implementation*, International Conference on Artificial Neural Networks, Springer, pp.391–398 2016.
7. S. Hori, and H. Tamukoh, *A random number generation method for hardware implemented neural networks*, IEICE Tech. Rep., vol. 119, no. 78, SIS2019-1, pp. 1-4, June 2019.
8. A. Fischer and C. Igel, *An introduction to restricted boltzmann machines*, Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, pp.14–36, Springer, 2012.
9. G.E. Hinton, *A practical guide to training restricted boltzmann machines*, Technical Report UTML TR 2010-003, Department of Computer Science, University of Tronto, 2010.
10. Xillybus PCIe IP, http://xillybus.com, access: 14th December, 2020.
11. Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, and LD. Jackel, *Backpropagation applied to handwritten zip code recognition*, Neural computation, vol.1, no.4, pp.541–551, 1989.

Synthesis of realistic food dataset using generative adversarial network based on RGB and depth images

Obada Al aama
Department of Life Science and Systems Engineering, Graduate School of Life Science and Systems Engineering,
Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan

Hakaru Tamukoh
Department of Human Intelligence System, Graduate School of Life Science and Systems Engineering,
Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan
E-mail:aama.obada-walidal425@mail.kyutech.jp, tamukoh@brain.kyutech.ac.jp
www.lsse.kyutech.ac.jp

Abstract

Constructing a large food dataset is time and effort consuming due to the need to cover the feature variations of food items. Hence, a huge data is needed for training neural networks. This paper aims to advocate the Cycle-GAN to build up large food dataset based on large number of simulated images and relatively few real captured images thus obtaining more realistic images effortlessly compared with traditional capturing. Real RGB and depth images of variant food samples allocated over turntable were captured in three different angles using real-sense depth camera with different backgrounds. Furthermore, for simulated images, the Autodesk 3D_Maya software was employed using the same parameters of captured real images. Results showed that generally, realistic style transfer on simulated food objects was obtained as a result of employing Cycle-GAN. GAN proved to be an efficient tool that could minimize imaging efforts resulting in realistic images.

Keywords: Cycle-GAN, Food dataset, RGB-D images.

1. Introduction

Recently, generative adversarial networks (GANs) have attracted particular attention. A GAN can generate an image that resembles a real image. A human face image dataset such as CelebA and a numeric character image dataset such as MNIST have been used for training the GAN as the target domains [1]. GANs have produced promising results in many generative tasks, such as photo-realistic image generation. One of the important obstacles of generating a large dataset is the need of a huge number of real images which is time and effort consuming. Meanwhile, few have reported food image generation or transformation using the GAN thus far. This paper aims to employ the automated Cycle-GAN technique to synthesize a large food dataset by the means of converting the simulated images to more relevant realistic images instead of only using large number of real images, hence, the Cycle-GAN produces the nearest real images based on processing a few numbers of real images along with the huge number of simulated images compared with the traditional imaging techniques. Results obtained from this study represent preliminary results that could be further used to train neural network for object detection which could be applied in various robotics applications.

2. Research Concept

The main concept focuses over using the depth camera to capture both RGB and depth images simultaneously, on the other hand, employing Autodesk 3D-Maya software to obtain the simulated images from 3D food models. Thereafter, training and testing the Cycle-GAN using those above mentioned images to investigate the capability of Cycle-GAN to produce the realistic images based training and testing few real images and huge simulated images.

3. Methodology

3.1. Real image capturing

This experiment was carried out through adjusting the Intel® RealSense™ Depth Camera D435i to be in front of turntable where the food objects including Apple, Kiwi, Mini-tomato and Mushroom were allocated. Both RGB and depth images were captured using this camera. The settled parameters of image capturing were as follows; distance from the object was 50 Cm, different capturing angles (0°, 45°, and 90°), turntable rotation rate was 1.22 rpm, capturing rate was 30 fps and the image resolution was 640×360 pixels. Images obtained were further extracted and processed as shown in Figure 1. All images for each food class were categorized in specific folders. Each class has 6570 images for RGB and the same number for depth in Jet-color mapping through obtaining 2190 images from each selected angle. In image capturing step the Rosbag_record was employed and accordingly the Region of Interest ROI cropping at 150×150 pixels and normalization were applied simultaneously resulting in Bag file format [2]. Extraction of images was carried out using the CV_bridge code [3, 4] which role is to re-save the images from RVIZ simulator to the proposed directory. Moreover, the saved images were processed through removing the background after the manual detection of the background color range for RGB via Chroma-key masking [5], and applying the same mask on the depth images.

3.2. Simulated image generation

The 3D simulated model for each candidate was downloaded from open source as an obj format. Thereafter these models were imported to the Autodesk

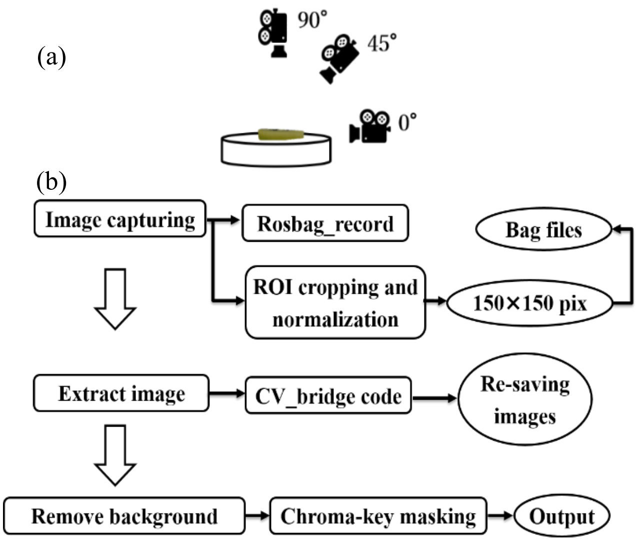


Fig. 1. Schematic diagram illustrating the different procedures of real image capturing beginning from the (a) image capturing process, and (b) different steps of extracting and processing of the obtained images.

3D_Maya software [6] to obtain the simulated images using the same parameters applied for capturing the real images (capturing angles, object capturing distance, turntable rotation rate, capturing rate, and resolution), resulting in the same number of captured images as those obtained from Real capturing. Finally, the background was automatically removed for both RGB and depth during the render step. However, the simulated depth images were obtained in grey-scale color model. The whole setup of the experiment was illustrated in Figure 2.

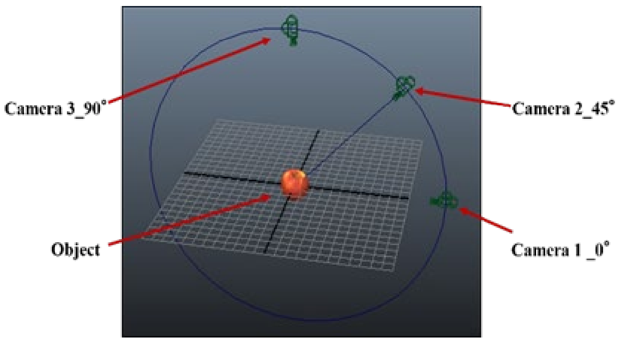


Fig. 2. Simulator experiment setup.

3.3. Cycle-GAN image generation

For each food class 500 real images and 500 simulated images were used to train the Cycle-GAN neural network.

The epoch was setup to be 200 epochs, and the training time was about 11 hours. Thereafter 6070 images were tested to validate Cycle-GAN neural network. Through Cycle-GAN training the equation used were obtained from 1. The main concept and various processes applied through Cycle-GAN was illustrated in Figure 3.

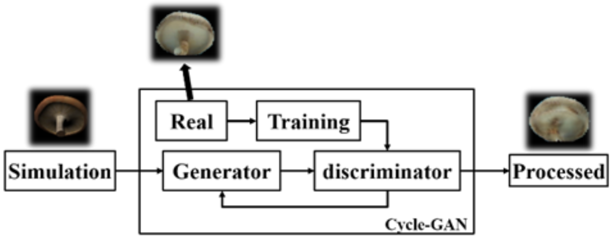


Fig. 3. Cycle-GAN main concept.

4. Results and discussion

4.1. Real imaging

Four classes were selected as food objects which were mushroom, mini-tomato, apples, and kiwi. For each object both RGB and depth were captured in three different capturing angles 0°, 45°, and 90°, resulting in 26280 total images as 2190 images for each category at each capturing angle. In Figures 4, 5, 6, and 7 the real images of different food objects were shown, along with the applied mask, and the obtained output. Moreover, at 0° the turn table surface appeared in all captured images and this require an extra cropping step to remove the interfered part before applying the Chroma-key masking. However, at 45° and 90° this process was not required and the Chroma-key masking was applied directly.

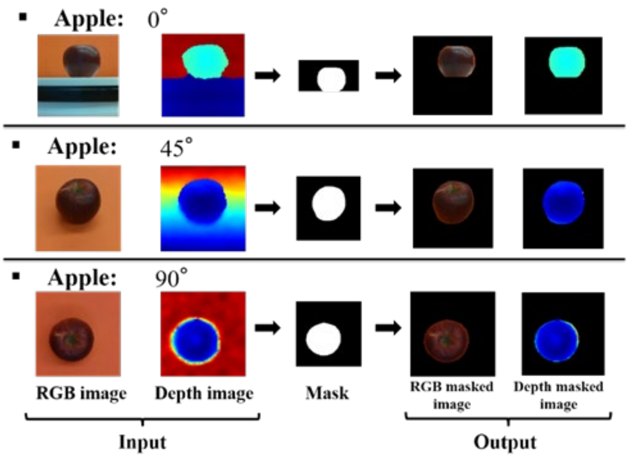


Fig. 4. Real captured RGB and depth images for apple, and the output after masking.

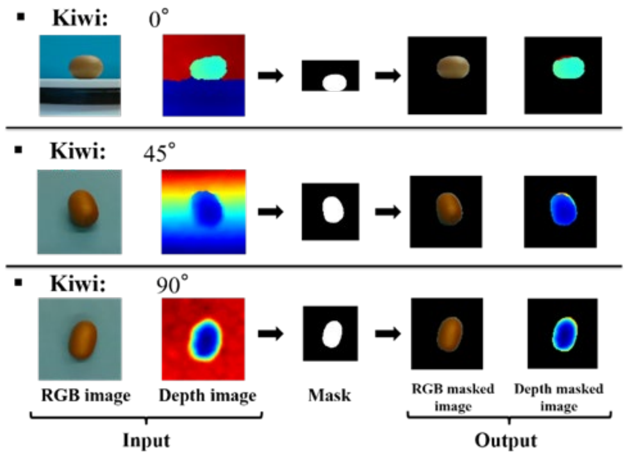


Fig. 5. Real captured RGB and depth images for kiwi, and the output after masking.

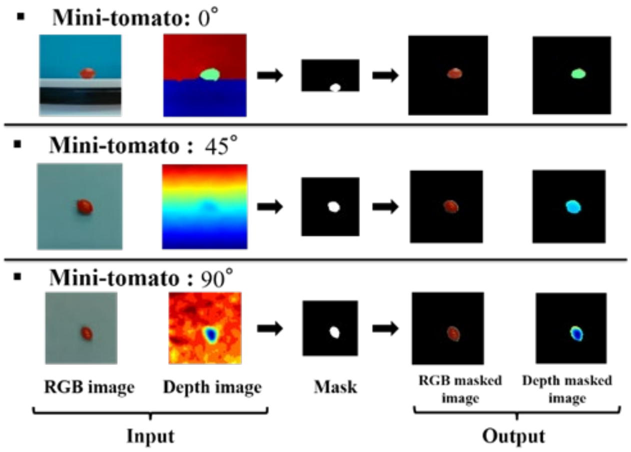


Fig. 6. Real captured RGB and depth images for mini-tomato, and the output after masking.

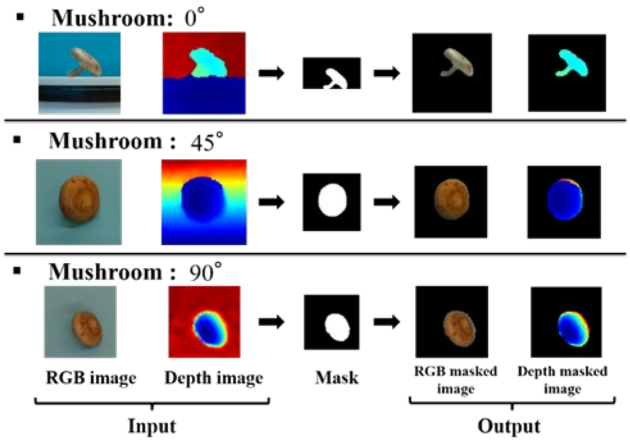


Fig. 7. Real captured RGB and depth images for mushroom, and the output after masking.

4.2. Simulated imaging

The same number of simulated images were obtained at each angle for both RGB and depth as those obtained in real imaging. However, the obtained depth images were grey-scale. Figure 8 shows the obtained RGB and depth simulated images for each investigated food class.

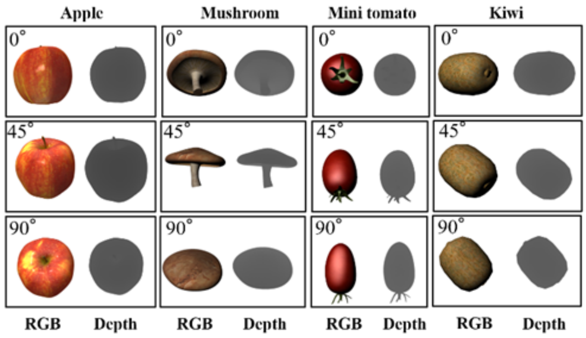


Fig. 8. Simulated RGB and depth images for apple, mushroom, mini-tomato, and kiwi.

4.3. Cycle-GAN

500 RGB Real images in addition to 500 RGB simulated for each object class were used to train the Cycle-GAN. Thereafter rest of the total obtained simulated images (6070) were used in Cycle-GAN testing step rather than those used in the training step as shown in Figure 9. It

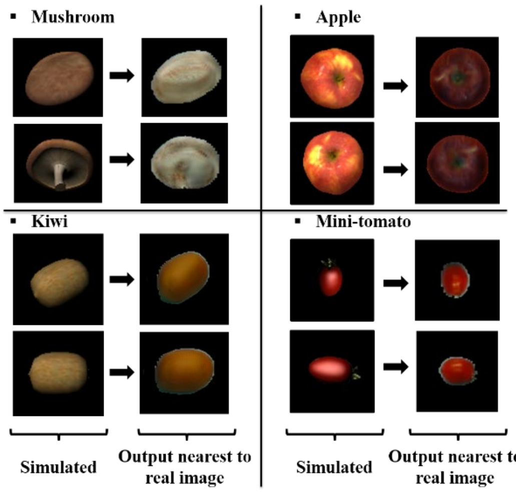


Fig. 9. Samples for Cycle-GAN output for different food classes RGB images at variant angles.

was noticed from the obtained RGB Cycle-GAN output that Cycle-GAN represent an efficient tool in modelling of near to real images in various angles. On the other hand, for the depth images 500 images were selected also for each object class to train the Cycle-GAN, and in the same manner the rest of the total obtained images were used to test the Cycle-GAN as illustrated in Figure 10 showing the obtained depth Cycle-GAN output for mushroom food class as example.

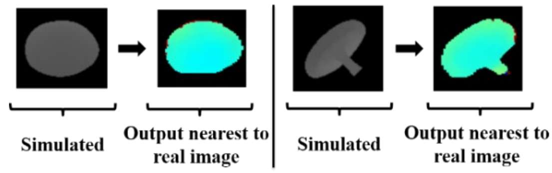


Fig. 10. Samples for Cycle-GAN output for mushroom class depth images.

5. Conclusion

This paper proposed to synthesize a large food dataset based on the use of few number of real images along with a huge number of simulated images utilizing the Cycle-GAN to produce realistic images. We used only 500 Real and another 500 simulated for both RGB and depth for training Cycle-GAN for each food class and this number represent a few number compared with that used for testing which was 6070 simulated images. Accordingly, our proposed method proved to be efficient in synthesizing of large food dataset. However, our data represent a preliminary result and a future study focusing on the effect of number of training images on the realistic quality of the output of Cycle-GAN, should be conducted, in addition to focusing on widening of the investigated food dataset including Japanese food.

References

1. HORITA, Daichi, et al. Food category transfer with conditional cyclegan and a large-scale food image dataset. In: Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management. 2018. p. 67-70
2. SHI, Wuzhen, et al. Reference image based method of region of interest enhancement for haze image. In: 2015 IEEE International Conference on Image Processing (ICIP). IEEE, 2015. p. 735-739.
3. Website:"http://wiki.ros.org/cv_bridge/Tutorials/ConvertingBetweenROSImagesAndOpenCVImagesPython", 15/12/2020.
4. MARTINEZ, Aaron; FERNÁNDEZ, Enrique. Learning ROS for robotics programming. Packt Publishing Ltd, 2013.
5. C. Teng, Y. Liao, Y. Chou and S. Lin, "Removing blue screen background under non-uniform illumination," 2017 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW), Taipei, 2017, pp. 307-308, doi: 10.1109/ICCE-China.2017.7991118.
6. Autodesk, INC. (2019). Maya. Retrieved from https:/ autodesk.com/maya.

Comparison of Data Augmentation Methods in Pointer-Generator Model Using Various Sentence Ranking Methods

Tomohito Ouchi, Masayoshi Tabuse

Graduate School of Life and Environmental Sciences, Kyoto Prefectural University,
1-5 Shimogamohangi-cho, Sakyo-ku, Kyoto 606-8522, Japan
E-mail: t_ouchi@mei.kpu.ac.jp, tabuse@kpu.ac.jp

Abstract

In the existing research, we proposed a data augmentation method using topic model for Pointer-Generator model. In this study, we add to the sentence ranking method in the data augmentation method. Specifically, we add two ranking methods using LexRank and Luhn. LexRank is based on Google's search method and Luhn defines sentence features and ranks sentences. We compare three data augmentation methods. We considered which method is suitable for data augmentation. We confirm that most accurate model is the model using data augmentation method by topic model.

Keywords: automatic summarization, data augmentation, Pointer-Generator Model, Extractive summarization.

1. Introduction

Currently, the amount of information on the Internet is increasing exponentially, and it is said that it will reach 59ZB in the research in May 2020[1]. It is also said that the total amount of data created in the next three years exceeds the total amount of data in the past 30 years, and the total amount of data created in the next five years is more than three times the total amount of data created in the past five years. Under such circumstances, the issue of selecting information is an urgent problem. Automatic summarization struggles that issue. However, it can be said that extractive summarization that only made up with sentences is not sufficient. Since the sentence-to-sentence connection is not taken into consideration, readability is lacking. Therefore, it is needed generative summarization as a technology that looks ahead. A generative summarization basically uses the Encoder-Decoder model, which learns the relationship between input and output and generates one word at a time in the output when a new input comes in during the test. Various models have been proposed [2,3]. In this study, the Pointer-Generator model [2] uses as the baseline model. One of the issues with the generative summarization model is that data maintenance is costly. We have to attach a manual summary to each article in order to make the generative summarization model. Therefore, we focused on data augmentation as a method that can be applied to any model. This is to create extended data from existing data. As a result, it was confirmed that the accuracy of the evaluation metric ROUGE [4] of Pointer-Generator model applied by the data augmentation method is improved by about 1% compared to baseline model.

Next, we explain the method of data augmentation simply. We decide the importance of each sentence in each article. And the sentence with the lowest importance is removed to obtain extended data. In the existing research [5], the topic model was used to measure the importance of sentences. In this study, in addition to that, the method called Luhn[6] and LexRank[7] was used. These three techniques are described in Section 2. Experiments and results are described in Section 3. And discussions are given in Section 4.

2. Data Augmentation Method

This section describes the three models used in the data augmentation method. In each method, each sentence is scored in an article, and the sentence with the lowest score is removed to obtain extended data.

2.1. *Luhn*

We measure position of the top 100 in most frequent words removed stop-words. We define words with a distance of 5 or less as one cluster. The score of a cluster is the square of the number of important words in a cluster divided by the distance between the first and last words of the cluster. Finally, the maximum score of each cluster becomes the score of the sentence.

2.2. *LexRank*

First, we explain PageRank, which is the basis of LexRank. The basic idea of PageRank is that linked pages are good pages, and links from more linked pages are evaluated highly. This rating is equivalent to the user inflow to the page. This is because if links are provided from many pages, it is easy to flow in, and the inflow from popular pages is larger than the inflow from normal pages. Links between pages can be represented by a matrix, which is the probability that the user will transition from that page to another linked page. The matrix is made with dividing by the total number of links on each page. The purpose of PageRank is to use this matrix to determine the probability that a user will stay on each page, that is, the rating of the page. PageRank is based on the premise that the page stay probability will eventually stabilize if the page transition is repeated many times, so that the transition matrix multiplied by the stay probability vector becomes closer to the transition matrix.

In LexRank, the transition matrix is the matrix of the cosine similarity of the Tf-Idf score between sentences. The basic idea of LexRank is that sentences similar to many sentences and sentences similar to important sentences are considered to be important sentences.

2.3. *Topic Model*

This method was used in the existing method. For how to determine the importance of sentences using the topic model, we referred to existing research [8]. The topic model is one of the language models that assumes that one document consists of multiple topics. In addition, each topic has an appearance word distribution. The method of determining the importance of a sentence is as follows.

1. Calculate the frequency of occurrence in a topic with words that make up a sentence
2. Sum of all the words that make up the sentence
3. Divide by the square root of the sentence length
4. Sum on all topics

3. Experiment and Results

In this section, the experimental conditions and the results of additional experiments using the Luhn method and the LexRank method are included.

3.1. *Parameter Setting*

The CNN / DailyMail dataset is used as the dataset for training, evaluating, and testing. The training data, evaluating data, and test data are 287,226 articles, 13,768 articles, and 11,490 articles, respectively. The model used for the experiment is the Pointer-Generator model, which is divided into a copy mechanism and a coverage mechanism when learning. The Copy mechanism calculates the error of the evaluating data each time the epoch ends and we uses the model of the epoch with the lowest error in Early Stopping. Early Stopping what we mean here, uses a model that waits twice as many epochs as the error seems to be the minimum, unless the minimum value is updated. Next, in the coverage mechanism, the same processing is performed in the coverage loss. We use ROUGE as using for evaluation on existing research.

The program used in this research uses PyTorch. It has been confirmed that this program can achieve the same result as [2]. The hidden layer vector size was set to 256 and the embedded vector size was set to 128. The batch size was set to 8. In the original paper, the batch size is 16, so double learning is required to learn the same number of articles. The beam size was set to 4. The beam search will be described later. The number of vocabulary was set to 50,000. The learning rate was set to 0.15.

In this program, the number of words used to encode an input article is limited to 400. This setting has no effect on learning an extended data. Specifically, an extended data is the same as an original data. This is because the extracted sentence may not be within 400 words from the beginning. We must confirm that the extracted sentence is present in the input article. Therefore, I found the article with the most number of words among the articles used in the training data. The number of words with the most words was 2,380. And the upper limit of the number of words used in encoding the input article was set to 2,380. Table 1 shows the values of ROUGE when the maximum number of words is 400 and 2,380. In the Table 1, f, r, and p represent the F value, recall, and precision, respectively.

Table1 the values of ROUGE when the maximum number of words is 400 and 2,380

	ROUGE-1-f	ROUGE-1-r	ROUGE-1-p	ROUGE-2-f	ROUGE-2-r	ROUGE-2-p
400	0.3935	0.4372	0.3800	0.1709	0.1891	0.1662
2380	0.3958	0.4181	0.3994	0.1741	0.1832	0.1770

	ROUGE-L-f	ROUGE-L-r	ROUGE-L-p
400	0.3616	0.4014	0.3493
2380	0.3644	0.3846	0.3679

Table 1 shows when the upper limit of the number of words is increased from 400 to 2,380, the value of ROUGE increases slightly. In the following, the experiment is performed with the upper limit of the number of words set to 2,380.

3.2. *Beam search*

Greedy method contrasts with beam search. This is because, when generating a word, one word with the highest generation probability is selected, while in beam search, processing is performed while holding the top K words. Then, we make the final summarizations by multiplying the probabilities of each word generation, and make the highest one the final summarization. In this experiment, this K value is set to 4.

3.3. *Results*

The results are as below.

Table 2 Results of learning 287226 articles using 6 methods

	normal	extended	ri	sr
rouge-1	0.3820	0.3948	0.3856	0.3869
rouge-2	0.1640	0.1724	0.1649	0.1683
rouge-L	0.3514	0.3624	0.3527	0.3551

	rs	rd	Luhn	LexRank
rouge-1	0.3911	0.3817	0.3739	0.3777
rouge-2	0.1696	0.1657	0.1640	0.1616
rouge-L	0.3596	0.3499	0.3449	0.3489

In order of good results, the existing method[5], baseline, ri, sr, rd, LexRank, rs, and Luhn.

A summary example generated in each model is shown.

reference
marseille prosecutor says `` so far no videos were used in the crash investigation " despite media reports .
journalists at bild and paris match are `` very confident " the video clip is real , an editor says .
andreas lubitz had informed his lufthansa training school of an episode of severe depression , airline says .

normal
new : `` a person who has such a video needs to immediately give it to the investigators " robin 's comments follow claims by two magazines , german daily bild and french paris match , of a cell phone video showing the harrowing final seconds from on board germanwings flight 9525 .
paris match and bild reported that the video was recovered from a phone at the wreckage site .

extended
marseille prosecutor brice robin says he was not aware of any video footage from the plane .
robin 's comments follow claims by two magazines , german daily bild and french paris match .
`` one can hear cries of ` my god ' in several languages , " paris match reported .

ri
marseille prosecutor brice robin told cnn that `` so far no videos were used in the crash investigation " robin 's comments follow claims by two magazines , german daily bild and french paris match .
all 150 on board were killed .

sr
new : `` it is a very disturbing scene , " official says .
new : `` one can hear cries of ` my god ' in several languages , " prosecutor says .
`` one can hear cries of ` my god ' in several languages , " official says .

rs
marseille prosecutor brice robin says he was not aware of any video footage from on board .
robin 's comments follow claims by two magazines , german bild and french paris match .
the video was found by a source close to the investigation .

rd
french prosecutor leading an investigation into the crash of germanwings flight 9525 .
robin 's comments follow claims by two magazines , german daily bild and french paris match , of a cell phone video showing the harrowing final seconds from on board the plane .
german airline lufthansa confirmed tuesday that co-pilot andreas lubitz had a `` previous episode of severe depression "

Luhn
new : `` so far no videos were used in the crash investigation , " official says .

new : `` a person who has such a video needs to immediately give it to the investigators , " says a forensic psychologist .

LexRank

marseille prosecutor brice robin told cnn that `` so far no videos were used in the crash investigation " french president francois hollande says it should be possible to identify all the victims using dna analysis by the end of the week .

4. Conclusion

In this study, in addition to the existing studies, we experimented with a data augmentation method using the Luhn and LexRank methods. The results confirmed that the best data augmentation method is to use the topic model of the existing research. In the future, we would like to confirm the effectiveness of data augmentation for state-of-the-art models. When extracting a sentence from an article, I would like to try a method for extracting multiple sentences instead of one sentence. We also expect that the number of sentences will change depending on the length of the article.

References

1. International Data Corporation (IDC) https://www.idc.com/
2. Abigail. See, Peter J. Liu, et al. “Get To The Point: Summarization with Pointer-Generator Networks” arXiv:1704.04368(2017)
3. Yang Liu, Mirella Lapata “Text Summarization with Pretrained Encoders” arXiv:1908.08345v2(2019)
4. Chin-Yew Lin, “ROUGE: A Package for Automatic Evaluation of Summaries” Workshop on Text Summarization Branches Out, Post-Conference Workshop of ACL 2004, Barcelona, Spain (2004)
5. T. Ouchi, M. Tabuse, “Effectiveness of Data Augmentation in Pointer-Generator Model” ICAROB(2020)
6. Luhn, H.P., “The automatic creation of literature of abstracts” IBM journal of research and development, vol. 2, No. 2, pp.159-165 (1958)
7. Gunes Erkan, Dragomir R Radev, “LexRank : graph-based lexical centrality as salience as text summarization” journal of Artificial Intelligence Research, vol. 22, pp. 457-479 (2004)
8. H.Sigematsu, I.Kobayashi “Generation of abstracts considering importance of potential topics” The Association for Natural Language Processing (2012) (In Japanese).

Music Recommendation System Driven by Variations in Fingertip Skin Temperature

Mayuka Wada[1], Taro Asada[2], Yasunari Yoshitomi[2], Masayoshi Tabuse[2]
1: Shimadzu System Development Corp., 1 Tokudaiji-cho, Nishinokyo, Nakagyo-ku,Kyoto, Japan
2: Graduate School of Life and Environmental Sciences, Kyoto Prefectural University,
1-5 Nakaragi-cho, Shimogamo, Sakyo-ku, Kyoto 606-8522, Japan
E-mail: t_asada@mei.kpu.ac.jp, {yoshitomi, tabuse}@kpu.ac.jp
http://www2.kpu.ac.jp/ningen/infsys/English_index.html

Abstract

In recent years, music therapy has been performed to recover the cognitive function of elderly people. In the previously reported system, it was necessary to input the subjective evaluation of the user to determine the next recommended song. In this study, variations in fingertip skin temperature were used as input instead of subjective evaluation, focusing on evaluation of emotion expressed as physiological response through variations in fingertip skin temperature. Our system can be used with no support for users, such as patients who have severe dementia or are bedridden and cannot respond readily.

Keywords: Music recommendation system, Music therapy, Peripheral skin temperature.

1. Introduction

In recent years, against the backdrop of increasingly aging societies in developed countries such as Japan, activities using music for the elderly ("music therapy") have been used with the aim of improving recognition ability. We previously developed music recommendation systems for music therapy for older people.[1]

However, in the previously reported music recommendation system,[1] it was necessary to input the subjective evaluation of the user ("I want to listen to it again" or "I don't want to listen to it again") into the recommended music in order to determine the next recommended song. Therefore, in the case of a user who has difficulty in responding by himself/herself, such as a dementia patient, the support of a caregiver may be required, and the problem of securing human resources hinders the use of our system.

In this study, we focused on the evaluation of emotions by measuring peripheral skin temperature,[2] developed a method for determining recommended songs based on the physiological evaluation of the user, and evaluated our system.

2. Related Studies

We now briefly explain the systems used in our previous study[1] and another study,[2] which we build upon to evaluate emotions in this study.

2.1. *Previously reported system for music recommendation*

The method for music recommendation used in this study is based on a previously reported methods.[1] Similar to the previously reported music recommendation system,[1] our system is a hybrid system for improving recognition ability using collaborative filtering and impression words.[3] See the paper[1] for detail on the music recommendation module in our system.

2.2. *Evaluation of emotions by measuring peripheral skin temperature*

In the study by Watanabe et al.,[2] the subject watched a horror video and an environmental video alternately, and the peripheral skin temperatures of the subjects viewing the videos were measured. Three peripheral parts, the front of auricle, the nose, and the back of the right index finger were measured, and temperature changes were compared and examined. No significant change was observed in the skin temperature on the front of the auricle and on the nose. However, the skin temperature on the back of the fingers of the right hand decreased when viewing the horror video and increased when viewing the environmental video. From this result, the influence of autonomic nerve activity appears to be reflected strongly in the peripheral skin temperature of the fingertip as a change in blood flow due to vasoconstriction.

3. Construction of a physiological evaluation method for music

A physiological evaluation method for music, which measures the peripheral skin temperatures of subjects listening to music, will be described. In this study, we selected the pad of the left index finger as the peripheral part for measuring temperature.

3.1. *Changes in peripheral skin temperature while listening to music*

In order to evaluate emotions by measuring peripheral skin temperature in a music recommendation system, we investigated the changes in peripheral skin temperature of subjects listening to music.

When subject A (a female in her twenties) listened to her favorite song ("Canon," composed by Johann Pachelbel) and least favorite song ("Mystery III - Heterophony for Swirl Butoh -," composed by Akira Nishimura), the peripheral skin temperature tended to rise while listening to her favorite song and decreased while listening to her least favorite song (Figures 1 and 2).[3]

3.2. *Determination of threshold for changes in peripheral skin temperature*

Similar to the previously reported music recommendation system,[1] the system used in this study determines the next recommended song by inputting a evaluation value (typing in "1" if you want to listen to the song again, or "0" if you don't want to listen to the song again). In this study, we determined the threshold value for dividing "1" when "physiologically preferable" and "0" when "physiologically unfavorable" based on the change in peripheral skin temperature.

Based on the results of the preliminary experiment, 0.4°C (the temperature difference between the maximum temperature and the subsequent minimum temperature) and 180 seconds (the elapsed time since the subject had begun listening to music) were set as the threshold values.

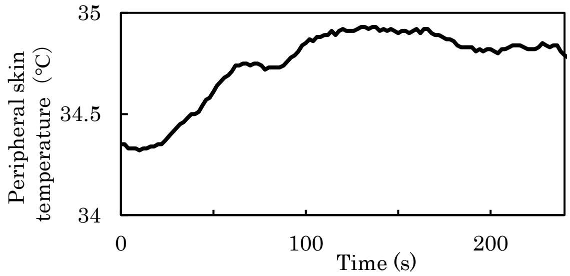


Fig. 1. Changes in peripheral skin temperature when listening to music (Favorite song).

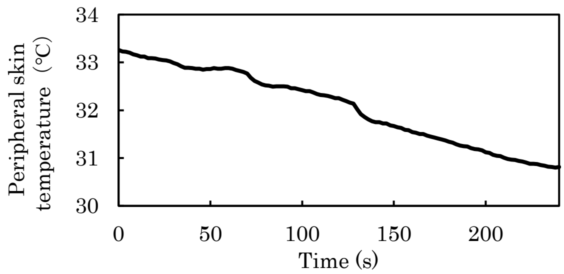


Fig. 2 Changes in peripheral skin temperature when listening to music (Least favorite song).

3.3. *Binarization of physiological evaluation*

Using the two thresholds described in subsection 3.2, the music being listened to was judged to be "Physiologically favorable (1)" or "Physiologically unfavorable (0)" based on the temperature change of the peripheral skin when listening to the music. Figure 3 shows the judgment processing flow for the physiological evaluation of music. In Figure 3, the elapsed time from the start of measurement is indicated by "time," and the measured temperature at that time is indicated by "temp." The maximum temperature from the start of measurement to that point is indicated as "max," and the difference between "max" and "temp" is indicated as "dif."

4. Experimental evaluation

4.1. *Conditions*

The music recommendation system[3] used in this study is an improved version of the previously reported system.[1] The improved system can support multiple music databases, and is mainly aimed at the elderly. It was reported that elderly people tend to choose songs that

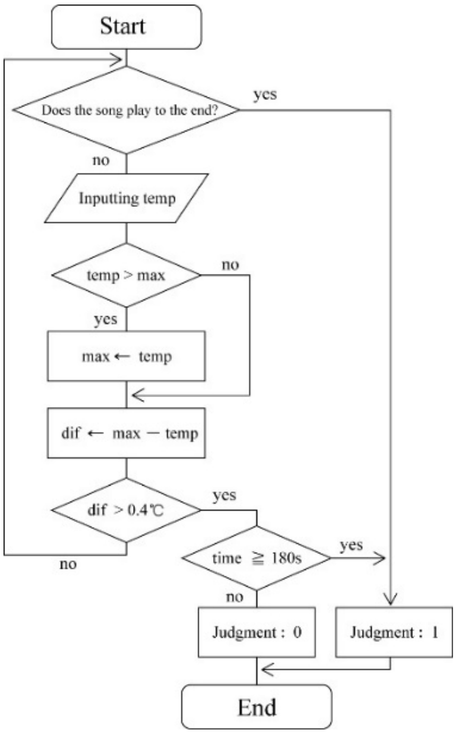


Fig. 3 Flowchart for binarizing physiological evaluation.[3]

are the subject of elementary school music textbooks as familiar songs (favorite songs).[5] Therefore, we selected a CD described as an anthology of older songs enjoyed by older people with dementia, and selected 52 songs on the CD that were also included in a music textbook database for elementary schools as well as 58 popular songs.[6] There were a total of 10 subjects: 4 females in their 20s (Subjects A to D), 5 males in their 20s (Subjects E to I), and 1 male in his 60s (Subject J). For programming, we used Visual C++ 6.0 (Microsoft) on a PC (Dell Optiplex790, CPU: Intel Corei7-2600 3.40GHz, main memory: 4.00 GB, and OS: Windows 7 Professional, Microsoft) for the experiment. We used a digital temperature and relative humidity sensor SHT11[7] to measure temperature.

4.2. *Results and discussion*

4.2.1. *Concordance rate*

Table 1 shows the results of the concordance rate between the physiological evaluation and the subjective evaluation. The concordance rate of all experimental results was relatively high, with an average of 70.9%. However, since there is a large difference in the concordance rate between subjects (Table 1), improvements such as reviewing the threshold are considered necessary. In addition, depending on the subject, the concordance rate of either the children's song or the popular song is high, and the concordance rate of the other is low. Therefore, it would appear that there are some problems that cannot be solved only by adjusting the threshold value.

4.2.2. *Correct answer rate of music recommendation*

Table 2 shows a comparison of the percentage of correct answers for the song recommendation results. The percentage of correct answers for the music recommendations was relatively low. The probable reason for this was that many of the subjects were relatively young, not elderly people who were originally targeted by our system. As an example, a male in his 60s (Subject J) had a relatively high percentage of correct answers (73.3-83.3%) in both the physiological and subjective evaluations. Figure 4 shows the change in peripheral skin temperature for the song that Subject J judged to be his "favorite song" in both the physiological and subjective evaluations. The peripheral skin temperature gradually increased over time (Figure 4).

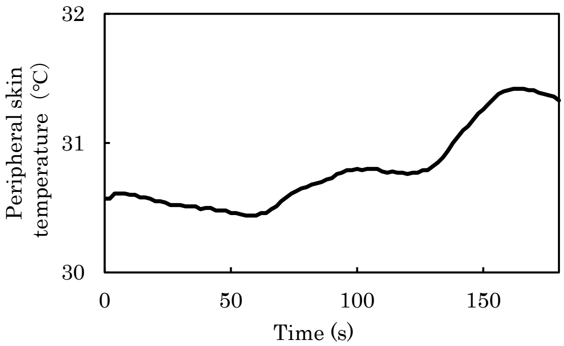


Fig. 4. Changes of subject J in peripheral skin temperature when listening to music (Favorite song).

Subject J (a male in his 60s) had a high rate of accuracy for both children's songs and popular songs. From this, the selection of the music database is considered to be one of the factors that contributed to the low recommendation accuracy for one of the subjects who was in their 20s. In future research, we plan to conduct experiments that focus on the elderly.

Table 1 Experimental results.[3]

	Subject	A		B		C		D		E		F		G		H		I		J		Total	
Children's songs	**Number of recommended songs**	**15**		**15**		**15**		**15**		**15**		**15**		**15**		**15**		**15**		**15**		**150**	
	Evaluation value	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**
	Physiological evaluation	8	7	5	10	6	9	6	9	7	8	9	6	8	7	10	5	8	7	11	4	78	72
	Subjective evaluation	9	6	8	7	14	1	9	6	8	7	5	10	7	8	11	4	5	10	13	2	89	61
	Number of concordances: Individual	8	6	3	5	6	1	5	5	4	4	5	6	7	7	8	2	2	4	11	2	59	42
	Number of concordances: All	14		8		7		10		8		11		14		10		6		13		101	
	Concordance rate (%)	93.3		53.3		46.7		66.7		53.3		73.3		93.3		66.7		40		86.7		**67.3**	
Popular songs	**Number of recommended songs**	**8**		**11**		**10**		**10**		**9**		**8**		**8**		**8**		**9**		**12**		**93**	
	Evaluation value	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**	**1**	**0**
	Physiological evaluation	3	5	3	8	9	1	8	2	6	3	1	7	3	5	6	2	6	3	10	2	55	38
	Subjective evaluation	3	5	6	5	10	0	8	2	2	7	2	6	5	3	7	1	5	4	9	3	57	36
	Number of concordances: Individual	3	5	1	3	9	0	7	1	2	3	1	6	2	2	6	1	4	2	9	2	44	25
	Number of concordances: All	8		4		9		8		5		7		4		7		6		11		69	
	Concordance rate (%)	100		36.4		90		80		55.6		87.5		50		87.5		66.7		91.7		**74.5**	

Table 2 Recommendation accuracy.[3]

	Subject	A	B	C	D	E	F	G	H	I	J	Total
Children's songs	**Physiological evaluation (%)**	53.3	33.3	40	40	46.7	60	53.3	66.7	53.3	73.3	**52**
	Subjective evaluation (%)	60	53.3	93.3	60	53.3	33.3	46.7	73.3	33.3	86.7	**59.3**
Popular songs	**Physiological evaluation (%)**	37.5	27.3	90	80	66.7	12.5	37.5	75	66.7	83.3	**57.6**
	Subjective evaluation (%)	37.5	54.5	100	80	22.2	25	62.5	87.5	55.6	75	**60**

5. Conclusion

By focusing on the evaluation of emotions by measuring the peripheral skin temperature, we have developed a music recommendation system that inputs subjective evaluations using peripheral skin temperature. This system can be used even when users who cannot respond actively, such as patients with severe dementia or bedridden patients, are not supported by a caregiver. The concordance rate between the physiological evaluation and the subjective evaluation varied depending on the subject and the music genre. One of the factors is that the threshold is set for one subject. In addition to music, there are many factors that affect peripheral skin temperature, such as room temperature and the physical condition of the subject.

Acknowledgements

We would also like to thank the subjects of our experiments for their cooperation.

References

1. S. Yoshizaki, Y. Yoshitomi, C. Koro, and T. Asada, Music recommendation hybrid system for improving recognition ability using collaborative filtering and impression words, *J. Artif. Life and Robotics* **18**(1-2), 2013, pp. 109–116.
2. T. Watanabe, M. Okubo, T. Kuroda, Evaluation of emotion using peripheral skin temperature-effectiveness of the back of finger measurement, *JJME* **36,** 1998, pp. 637.
3. M. Wada, T. Asada, Y. Yoshitomi, and M. Tabuse, Music recommendation system using peripheral skin temperature as input data (in Japanese), in *Proc. of Human Interface Symposium 2019,* Kyoto, Japan, Sep., 2019, pp. 736–739.
4. T. Kumamoto and K. Ohta, Design of scales to represent user's impressions of a music piece for use in a music-retrieval system (in Japanese), *IPSJ SIG Notes* **2001-NL-147**(6), (2002), 35–40.
5. T. Takahashi, Research report on songs familiar to people advanced in years (in Japanese), *J. Japanese music therapy associate* **15**(1) , 1997, pp. 68–75.
6. T. Akahoshi, *Good old anthology enjoyable for people advanced in years and troubled with dementia* (in Japanese) , Kirara shobo, Tokyo, 2009.
7. Sensirion Digital Humidity Sensor, https://www.sensirion.com/en/environmental-sensors/humidity-sensors/digital-humidity-sensors-for-accurate-measurements , Accessed 4 December 2020.

Music Recommendation System Driven by Interaction between User and Personified Agent Using Speech Recognition, Synthesized Voice and Facial Expression

Ayumi Matsui[1], Miki Sakurai[2], Taro Asada[3], Yasunari Yoshitomi[3], Masayoshi Tabuse[3]
1: Sumitomo Mitsui Card Co., Ltd., 4-5-15 Imabashi, Chuo-ku, Osaka 541-0042, Japan
2: TIS Inc.,8-17-1 Nishi-Shinjuku, Shinjuku-ku, Tokyo 160-0023, Japan,
3: Graduate School of Life and Environmental Sciences, Kyoto Prefectural University,
1-5 Nakaragi-cho, Shimogamo, Sakyo-ku, Kyoto 606-8522, Japan
E-mail: t_asada@mei.kpu.ac.jp, {yoshitomi, tabuse}@kpu.ac.jp
http://www2.kpu.ac.jp/ningen/infsys/English_index.html

Abstract

We propose a music recommendation system that is characterized by an interaction between a user and a personified agent. Speech is recognized using a speech recognition system called Julius, and the facial expression of the agent is then synthesized using preset parameters that depend on each vowel. We add a new function that changes the facial expression of the agent according to the response of the user to music recommended by our previously proposed system. The effectiveness of the proposed system is verified.

Keywords: Music recommendation, MMDAgent, Facial expression synthesis, Speech recognition, Speech synthesis.

1. Introduction

In Japan, the average age of the population has been increasing, and this trend is expected to continue. This trend is more pronounced in rural areas. Recently, music therapy has been used to improve the recognition ability of people, particularly older people. Music therapy may be more effective if music that is liked by an individual is adopted. We have previously developed a music recommendation system that aims to improve recognition ability, including a system through the Internet using a videophone system.[1-3] In our previously proposed system,[1-3] it is necessary to input the subjective evaluation of the user into a computer to determine the next music that is to be recommended. User input is inconvenient for elderly people, especially those with dementia.

In the present study, to overcome this inconvenience, we propose a system for music recommendation that is characterized by an interaction between a user and a personified agent, which uses speech recognition, synthesized voice and facial expression generation.

2. Proposed System and Method

2.1. *System overview*

Figure 1 shows the processing flow of the proposed system, the music recommendation module of which is based on the previously proposed system[2] that used collaborative filtering and impression words (see the paper[2] for detail on the music recommendation module). The proposed system is characterized by an interaction between a user and a personified agent, and uses speech recognition, synthesized voice and facial expression generation for integrative music recommendation.

To make the agent on a personal computer (PC) screen appear more human-like, we developed a system for agent facial expression generation that uses vowel recognition when generating synthesized speech.[4] We added a new function to change the facial expression of the agent according to the response of the user to music recommended by our system.

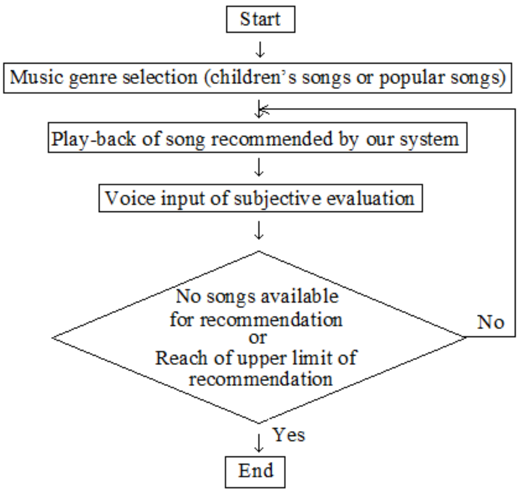


Fig. 1. Processing flow of the proposed system.

2.2. *Personified agent*

Figure 2 shows the process for agent generation in our system, which consists of six steps: creating facial expression data, recording vocal utterances, automatic WAVE file division, speech recognition by the Julius[5], insertion of expressionless data, and the creation of facial expression motion.[4] The facial expression data are created in advance.

Expressive motions are generated by combining the expression data of each vowel for each utterance motion. Then, utterance contents are input as text and used by the MikuMikuDanceAgent (MMDAgent),[6] which is a freeware animation program that allows users to create and animate movies with agents, to output synthesized voice that is then recorded by a stereo mixer inside a PC and saved as a WAVE file. Speech is recognized using a speech recognition system called Julius,[5] followed by facial expression synthesis of the agent using preset parameters depending on each vowel. Facial expression data were created with MikuMikuDance.[7] In this study, in order to generate more human-like agent facial expressions, facial expression data were created for the vowels / a /, / i /, / u /, / e /, and / o / (Fig. 3).[4] In order to create more natural agent facial expressions, processing is then performed to insert a neutral facial expression when the same vowel, for example / a /, is continuous.[4]

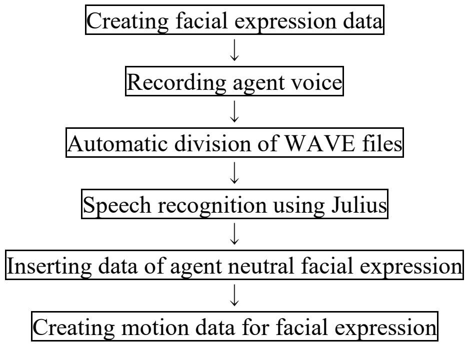


Fig. 2. Processing flow of agent generation in our system.[4]

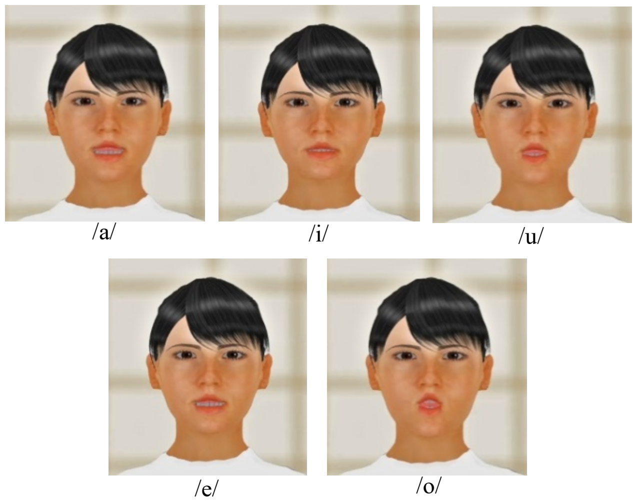


Fig. 3. Facial expression of the agent in uttering each vowel.[4]

Figure 4 shows the flow of creating a facial expression motion (see the paper[4] to understand the processing described in Fig. 4 in more detail).

2.3. *Navigations by improved agent*

In the music-recommendation process, all user navigations are performed by the synthetic voice of the agent, appearing on the PC screen facing the user. All dialogue spoken by the agent is situationally selected by the proposed system. The user's answers to the questions generated by the agent are recognized using the voice recognition function of the system, and the agent motions, including facial expressions, are then generated.

Reading audio data from Julius out file
↓
Calculating the total number of bones and writing it to the file header.
↓
Reading facial expression data according to vowel uttering time
↓
Change the vowel uttering time to the number of frames and writing corresponding data to the vmd file

Fig. 4. Flow of facial expression motion creation.[4]

(a) (b)

Fig. 5. Snapshots in the respective reactions of the agent after recognizing: (a) a positive answer, and (b) a negative answer.

Figure 5 shows two snapshots of the reaction of the agent after recognizing: (a) a positive answer, i.e., the user wishes to listen to the recommended song again in the future, and (b) a negative answer, i.e., the user does not wish to listen to the recommended song in the future. In the case of (a), the agent nods twice and raises the corners of the mouth slightly, while in the case of (b), the agent also nods twice, but instead lowers the corners of the mouth slightly. Figure 6 shows a snapshot of music recommendation being performed by the proposed system.

Fig. 6. Snapshot in performing song-recommendation by the proposed system.

3. Experiment

3.1. *Conditions*

Since older people tend to prefer children's songs,[8] we selected a CD[9] described as an anthology of older songs that are enjoyed by older people with dementia, and selected (1) 52 songs on the CD that were also included in a musical textbook database[10] for elementary schools, and (2) 58 other popular songs on the CD.

The experiment was performed on a Dell Inspiron 15 PC, equipped with an Intel Core i7-6700HQ 2.60 GHz central processing unit (CPU) and 8.0 GB of random access memory (RAM). The Microsoft Visual C++ 2010 Express and Visual C++ 6.0 were used as the development languages.

To evaluate the proposed method described in Section 2, 10 males (Subjects A to G in their 20s, I in his 40s, J in his 50s, and L in his 80s) and 2 females (Subject H in her 20s and K in her 50s) participated in the experiments. Each of the subjects navigated the music-recommendation process of the agent. The upper limit for the number of songs recommended by the proposed system was set to be 15. The experiment was performed separately using the database consisting of the 52 children's songs, and that consisting of the 58 popular songs, with two experimental conditions for each (Condition 1: with the input to the system being performed by a supporter instead of the agent navigating the system, and Condition 2: with the agent navigating the proposed system). Following the experiment, all subjects were requested to select one of five [evaluation value] answers ([5] absolutely yes, [4] yes, [3] I can't say either, [2] no, [1] absolutely no) to seven questions (Table 1) in order to evaluate the proposed system. The subjects were also asked to provide comments on the system.

Table 1. Questionnaire to evaluate the proposed system.

No.	Question
1	Was music-recommendation on condition 2 smoother than that on condition 1?
2	Were explanations by the agent easy to understand?
3	Were dialogues with the agent natural?
4	Were movements of agent mouth natural?
5	Were agent's reactions natural after recognizing user's positive answer for listening to the just recommended music again in the future?
6	Were agent's reactions natural after recognizing user's negative answer of no more he just recommended music from now on?
7	Did you feel enjoyable in using the proposed system?

3.2. *Results and discussion*

Table 2 shows the average-values for each question listed in Table 1. The mean value of the averages listed in Table 2 was 4.1, suggesting a positive overall evaluation of the proposed system. For questions 5 and 7, the evaluation averages were very large, while those of questions 1 and 4 were lower.

Table 2. Evaluation of the proposed system.

Question no.	1	2	3	4	5	6	7
Average	3.6	4.4	4.0	3.6	4.5	4.3	4.5

Table 3 shows the average values for each gender (M=male, F=female) and age group for each of the questions listed in Table 1. For Male Subject L, who was in his 80s, the evaluation values for questions 1 and 2 were very low compared with those of other subjects. Subject L described in his comments that the agent spoke too rapidly for him to understand its statements, and that it was difficult to input his voice using the microphone. Three male subjects (D and E in their 20s and Subject J in his 50s) also described in their comments that it was difficult to input their voices using the microphone. Consequently, speech recognition for Subjects D, J, and L was not performed smoothly by the proposed system, resulting in inputting voice again after training the system. Nearly all of the comments by the subjects expressed positive evaluation of the proposed system, except in relation to voice input to the system.

Table 3. Evaluation of the proposed system by each age group.

	Question no.	1	2	3	4	5	6	7
Average	M-20s (n=7)	3.7	4.7	4.4	3.6	4.6	4.3	4.1
	F-20s (n=1)	4	5	3	4	5	5	4
	M-40s (n=1)	3	4	4	3	4	4	4
	M-50s (n=1)	4	4	4	4	4	4	5
	F-50s (n=1)	5	5	5	5	5	5	5
	M-80s (n=1)	2	2	4	4	4	4	5

4. Conclusions

We proposed a system for music recommendation, characterized by an interaction between a user and a personified agent that uses speech recognition, synthesized voice and facial expression generation. Speech is recognized using a speech recognition system called Julius, and the facial expression of the agent is then synthesized using preset parameters that depend on each vowel. We used MMDAgent to create the agent. To produce the voice of the agent, we used the built-in speech synthesis function setting in MMDAgent. We added a new function that changes the facial expression of the agent according to the responses of the users to music recommended by our previously proposed system. The effectiveness of the proposed system was verified.

Acknowledgements

The authors would like to thank Professor J. Narumoto, of the Kyoto Prefectural University of Medicine, for his valuable support and helpful advice during the course of this research. We would also like to thank the subjects of our experiments for their cooperation. This research was supported by COI STREAM of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

1. C. Koro, Y. Yoshitomi, T. Asada, and S. Yoshizaki, Music recommendation aimed at improving recognition ability using collaborative filtering and impression words, *in Proc. 17th Int. Symp. on Artificial Life and Robotics*, ed. M. Sugisaka (Japan, Beppu, 2012), pp. 222–225.
2. S. Yoshizaki, Y. Yoshitomi, C. Koro, and T. Asada, Music recommendation hybrid system for improving recognition ability using collaborative filtering and impression words, *J. Artif. Life and Robotics* **18**(1-2) (2013) 109–116.
3. Y.Yoshitomi, T.Asada, R.Kato, Y.Yoshimitsu, M.Tabuse, N.Kuwahara, and J.Narumoto, Music recommendation system through Internet for improving recognition ability using collaborative filtering and impression words, *J. Robotics, Networking and Artif. Life*, **2**(1), 2015, 54–59.
4. T. Asada, R. Adachi, S. Takada, Y. Yoshitomi, and M. Tabuse, Facial expression synthesis using vowel recognition for synthesized speech, *in Proc. 2020 Int. Conf. on Artificial Life and Robotics*, ed. M. Sugisaka (Japan, Beppu, 2020), pp.398-402.
5. Julius, http://Julius.osdn.jp/, Accessed 24 November 2020.
6. MMDAgent, http://www.mmdagent.jp/ Accessed 24 November 2020.
7. MikuMikuDance, https://sites.google.com/view/vpvp/, Accessed 29 November 2020.
8. T. Takahashi, Research report on songs familiar to people advanced in years (in Japanese), *J. Japanese music therapy associate* **15**(1) (1997) 68–75.
9. T. Akahoshi, *Good old anthology enjoyable for people advanced in years and troubled with dementia* (in Japanese) (Kirara shobo, Tokyo, 2009).
10. Music textbook database for elementary school by Kanagawa prefectural education center (in Japanese), http://kjd.edu-ctr.pref.kanagawa.jp/daizai_music/, Accessed 27 November 2020.

Wallet Operation Evaluation System Using Deep Learning

Junichiro Yamawaki[1], Yasunari Yoshitomi[2], Masayoshi Tabuse[2], Taro Asada[2]
1: SKY Co., Ltd.,3-4-30 Miyahara,Yodogawa-ku, Osaka 532-0003, Japan
2: Graduate School of Life and Environmental Sciences, Kyoto Prefectural University,
1-5 Nakaragi-cho, Shimogamo, Sakyo-ku, Kyoto 606-8522, Japan
E-mail: {yoshitomi, tabuse}@kpu.ac.jp, t_asada@mei.kpu.ac.jp
http://www2.kpu.ac.jp/ningen/infsys/English_index.html

Abstract

As the average age of Japan's population increases, as a method for investigating mild cognitive impairment (MCI), wallet operation evaluations have been receiving considerable attention recently. Herein, we propose a system for conducting wallet operation evaluations based on deep learning. In our system, the bills and coins extracted from a wallet are automatically scanned and recognized, which makes it possible to evaluate a person's ability to correctly select and extract the correct bills and coins from the wallet within a reasonable period.

Keywords: Mild cognitive impairment, Dementia, Wallet operation evaluations, Deep learning

1. Introduction

The average age of Japan's population is increasing, and this trend is expected to continue. As a result, the number of elderly persons suffering from dementia is growing rapidly, and it has become necessary to find ways to identify persons suffering from mild cognitive impairment (MCI)[1] in order to ensure they have proper care while suppressing the progression of the disease.

As a method for investigating MCI, which is one of the pre-stages of dementia, wallet operation evaluations have been receiving considerable attention recently.[2] However, in the normal course of such testing, the test subject must be observed by an examiner who provides instructions on how the evaluation is conducted, observes the testing, and record the results.[2]

In contrast, this paper reports on an automatic evaluation system in which deep learning is used to conduct wallet operation evaluations automatically.

2. Wallet Operation Evaluations

During a conventional wallet operation evaluation,[2] a subject is tested by an examiner who instructs him or her on how to proceed, observes the process, and then records the results. This process can be broken down into the following six operational units:

(1) The examiner hands the test subject a wallet containing five (¥1, ¥5, ¥10, ¥50, and ¥100) Japanese coins each, five ¥1,000 Japanese bills, and one (¥2,000, ¥5,000, and ¥10,000) Japanese bills each , as partially shown in Fig. 1. (In the wallet, the total number of coins is 25, while that of bills is eight.)
(2) Next, the test subject is shown a sheet of paper describing two trials for 10 seconds. In the first trial, he or she is instructed to extract ¥1,165 from the wallet and place the money on the desk. The second trial is the same but the amount is ¥123.
(3) The test subject is instructed to begin the trials described in (2) when the examiner presses the button on his or her stopwatch.
(4) The test subject performs the two tests and reports his/her completion as instructed by the examiner.
(5) After visually confirming that both operations have been completed and receiving the test subject's report, the examiner presses the stopwatch button and informs the test subject that the evaluation is over.
(6) The time required by the test subject to complete both the first operation and second operation is then recorded by the examiner.

Fig. 1. Some of objects used in wallet operation evaluation.

3. Proposed System and Method

Our proposed system consists of two parts: (1) data preparation and deep learning and (2) the wallet operation evaluation module. Figure 2 shows the processing flow of the first part of this system, which consists of four processing units:

(1) Obtaining object image data
(2) Data augmentation
(3) Annotation
(4) Deep learning

Obtaining object image data
↓
Data augmentation
↓
Annotation
↓
Deep learning

Fig. 2. Processing flow of data preparation and deep learning.

As objects to be recognized, we use ¥1, ¥5, ¥10, ¥50, ¥100, and ¥500 Japanese coins and ¥1,000, ¥2,000, ¥5,000, and ¥10,000 Japanese bills. For dataset assembly and recording the bills and coins during actual testing, we use a web camera set 40 cm above a desk plate to obtain object images, as shown in Fig. 3.

Fig. 3. Example of object image.

In this study, we use a region-based convolutional neural network (R-CNN) known as Faster R-CNN[3] for deep learning and object recognition.

At the beginning of the evaluation, the same wallet containing the same coins and bills described in Section 2 is set in front of the test subject. The processing flow of the second part of the proposed system consists of the following 10 processing units:

(1)Japanese explanations of the evaluation process (three pages) are displayed sequentially on a personal computer (PC) screen, as shown in Fig. 4. The test subject indicates when he/she has finished reading each page by pressing the button connected to the PC.
(2)The PC then displays a 10-second message showing the two monetary amounts (first: ¥1,165, second: ¥123) the test subject will be requested to extract from the wallet.
(3)The evaluation begins when the test subject presses the button. At that point, the web camera and the system begin recording the elapsed time.
(4)The test subject extracts the appropriate bills and coins from the wallet and then presses the button again when he/she desires to report that an amount of money equal to ¥1,165 has been placed on the desk plate as instructed.
(5)An image showing the present state of the desk plate is then uploaded into the system.
(6)Next, the test subject proceeds to extract ¥123 from the wallet and place it on the desk plate. Note that a condition of this test portion is that the test subject is instructed not to use the coins extracted for the ¥1,165 test portion.
(7)The test subject then signifies the completion of that portion of the text by pressing the button again.
(8)An image showing the present state of the desk plate is uploaded into the system.
(9)Using object recognition, the total amount of money extracted from the wallet by the test subject, and the time required for both test segments are calculated by the proposed system.
(10)Finally, the time test and object recognition results are displayed on the PC screen, as shown in Fig. 5.

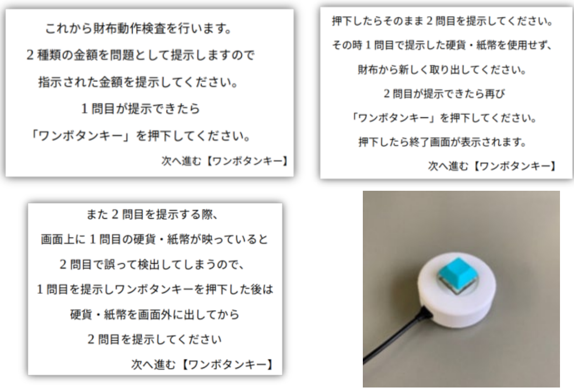


Fig. 4. Japanese language explanations of the evaluation process as displayed on the PC screen in order of upper left, upper right, and lower left. The bottom right photos show the button used for changing the display and recording elapsed time.

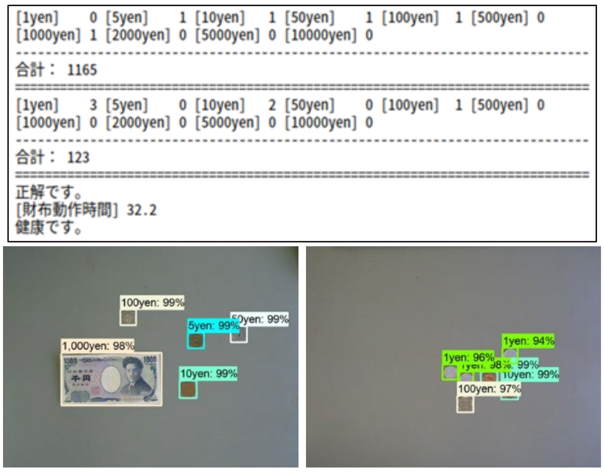


Fig. 5. Calculated evaluation results (upper) and object recognition results (lower left: ¥1,165, lower right ¥123).

4. Experiment

4.1. *Conditions*

The experiment was performed on a PC equipped with an Intel Core i7-6700K central processing unit (CPU), a GeForce GTX 1080 graphics processing unit (GPU), and 32.3 GB of random access memory (RAM). The Ubuntu LTS18.04 64-bit operating system (OS) was installed on the PC and Python 3.7.3 was used as the development language. Jupyter Notebook,[4] TensorFrow,[5] ImageDataGenerator,[6] labelImg,[7] CUDA10.0,[8] cuDNN,[9] and OpenCV[10] were used as toolkits and libraries.

For moving image input, a stream web camera was used. ImageDataGenerator was used for data augmentation, while rotation, vertical shift, horizontal shift, expansion, shrink, and channel shift were used as processing. Next, labelImg was used for annotations.

For deep learning, the TFRcode[11] for the faster_rcnn_resnet_101_coco model in TensorFlow[11] was generated using a program obtained from Ref. 12. The mean average precision (MAP) obtained by the learning model used in the present study was 0.755.

The resulting dataset contained 1,200 images showing the wallet and/or the Japanese bills and coins used in our experiment. Using ImageDataGenerator for data augmentation, 12,000 images were generated from the above 1,200 images. 10,000 of those images were used as learning data and the remaining 2,000 images were used for the evaluations. Figure 6 shows an example of data generated by augmentation.

To evaluate our proposed system, nine males (Subjects A to I) and one female (Subject J), all of whom are in their 20s, participated in our experiments.

Fig. 6. Original image (left) and augmented image (right).

The experiments were performed under two conditions listed below:

Condition 1: with the conventional method[2] described in Section 2.
Condition 2: with the proposed system described in Section 3.

After the experiments, all of the test subjects were requested to select one of five answer-choices ([5] definitely yes, [4] yes, [3] no difference, [2] no, [1] definitely no) for each question in Table 1 in order to evaluate the proposed system.

Table 1. Questionnaire to evaluate the proposed system.

No.	Question
1	Was the wallet operation test on condition 2 easier than that on condition 1?
2	Did you feel less stress during the wallet operation test on condition 2 than on condition 1?

4.2. *Results and discussions*

Table 2 shows the average time measured for the 10 subjects during these experiments. Here, it can be seen that the time required for the wallet operation evaluation using our proposed system was slightly longer than that of the conventional method. This might be caused by the additional button-pushing operations required during the process of our proposed method.

However, in both the conventional and proposed methods, none of the test subjects exceeded the limit time (120 s) that was assigned for the wallet operation evaluation in this experiment. This was thought to be because all subjects were in their 20s and in good mental and physical condition.

Table 2. Comparison of the average time of wallet operation evaluation between the conventional and proposed methods.

	Average time (s)
Conventional method	32.2
Proposed method	34.8

Table 3 shows the average questionnaire score for the 10 subjects. Here, it can be seen that the proposed method was better than the conventional method from the viewpoints of ease of use and stress reduction.

For all of the objects on the desk plate, the average correct recognition rate by the proposed system was 80% for ¥1,165 and 90% for ¥123. Additionally, ¥100 coins were sometimes misrecognized as ¥500 coins (Fig. 7) and ¥1 coins were sometimes misrecognized as ¥100 coins, possibly due to light reflection. However, mistakes in both cases were rare.

As our future targets for improving our proposed system, we intend to perform the wallet operation evaluation with the participation of elderly persons who may have MCI or other forms of dementia. Moreover, we will reduce potential misrecognitions of the ¥1, ¥100, and ¥500 coins, primarily by making improvements to the lighting conditions and/or the learning dataset.

Table 3. Average questionnaire scores for 10 subjects.

Question No.	Average score
1	4.2
2	4.1

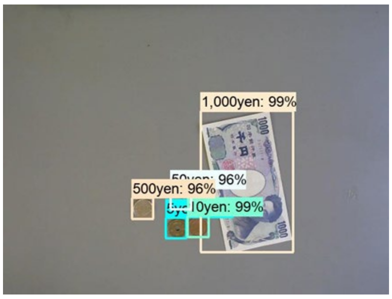


Fig. 7. Misrecognition of a ¥100 coin that was identified as a ¥500 coin.

5. Conclusion

As the number of elderly persons in Japan's super-aging society continues to increase, it is becoming increasingly important to find ways to identify persons suffering from MCI, which is one of the pre-stages of dementia, in order to ensure they receive proper care and make efforts to suppress the progress of the disease. As a method for investigating MCI, wallet operation evaluations have been receiving considerable attention recently.

Herein, we proposed a wallet operation evaluation system based on deep learning. In our system, a wallet containing bills and coins is supplied to the test subject, and his or her ability to extract bills and coins in the correct amounts is automatically recognized. The proposed system also determines whether the extractions from the wallet are made within a reasonable period.

Acknowledgments

The authors would like to thank Professor J. Narumoto of the Kyoto Prefectural University of Medicine for his valuable support and helpful advice during this research. We would also like to thank the test subjects of our experiments for their cooperation. This research was supported by COI STREAM of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

References

1. B. Winblad, K. Palmer, M. Kivipelto, V. Jelic, L. Fratiglioni, et al., Mild cognitive impairment – beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment, *J Int. Med.,* **256**(3), 2004, pp. 240–246.
2. R. Senba, K. Kamijo, A. Kubo, and S. Murata, Comparison of wallet operations of community-dwelling elderly people by cognitive function category (in Japanese), *Japanese J. Clinical Occupational Therapy,* **3**(1), 2016, pp. 5–9.
3. S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, in *Advances in Neural Information Processing Systems 28 (NIPS 2015),* eds C. Cortes and N. Lawrence and D. Lee and M. Sugiyama and R. Garnett, 2015.
4. Jupyter Notebook, https://jupyter.org/, Accessed 6 December 2020.
5. TensorFrow, https://github.com/tensorflow/, Accessed 6 December 2020.
6. ImageDataGenerator, http://keras.io/preprocessing/image/, Accessed 6 December 2020.
7. labelImg, https://github.com/tzutalin/labelImg, Accessed 6 December 2020.
8. CUDA, https://developer.nvidia.com/cuda-downloads, Accessed 6 December 2020.
9. cuDNN, https://developer.nvidia.com/cuDNN, Accessed 6 December 2020.
10. OpenCV, https://opencv.org/, Accessed 6 December 2020.
11. TensorFlow model zoo, https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/detection_model_zoo.md, Accessed 27 February, 2020.
12. https://github.com/EdjeElectronics/TensorFlow-Object-Detection-API-Tutorial-Train-Multiple-Objects-Windows-10, Accessed 27 February, 2020.

Architecture of a student training computer program for preparing professional outpatient consulting skills within an electronic medical records system during COVID-19 alertness situation

Sergey Bulatov[1], Evgeni Magid[2], Enzhe Kharisova[1], Roman Lavrenov[2], Vitaly Dudin[1], Artur Khazetdinov[2]

[1]Center of a practical skills, Kazan State Medical University, 49 Butlerov street, Kazan, Russian Federation
[2]Laboratory of Intelligent Robotic Systems (LIRS), Intelligent Robotics Department, Institute for Information Technology and Intelligent Systems, Kazan Federal University, 35 Kremlyovskaya street, Kazan, Russian Federation

Abstract

One of the important goals of a medical education in the context of the COVID-19 pandemic is a broad introduction of simulated-based teaching methods, which might include elements of robotics and artificial intelligence. We analyzed computer programs that are currently used for maintaining medical records of patients by various polyclinics of Kazan city. Based on these results, we summarized requirements for a training computer program that could provide students with medical records maintaining skills and developed program's preliminary architecture. Our main attention focused on modeling situations that are associated with pre-hospital stage processing of medical records for patients with suspected or confirmed COVID-19.

Keywords: medical education, computer program, COVID-19 infection

1. Introduction

The current epidemiological situation led to a transition of a higher medical education to distance learning[1]. While a main pedagogical emphasis moved to information technologies applications for educational process, including computer programs and artificial intelligence[2,3], specifics of medical information systems (MIS) do not allow all students to master them independently. A preliminary training of students to work with computer MIS used in practical health care has become acute.

The COVID-19 pandemic, which already had infected almost 60 million people all around the world and took over 1.4 million lives just within less than a year, has been a contributing factor in this process as it changed a traditional doctor-patient relationship and replaced face-to-face communication with a remote monitoring of patients[4]. In local polyclinics, computer programs create electronic medical records (EMR) of patients, draw up sick lists, monitor laboratory tests, and ensure preservation of information about patients' treatment. In literature sources, the authors had found only isolated reports about special training of students to work EMRs and MIS[5].

2. Related work

In medical universities of Russia, a special training of graduate students to work with MIS of a regional polyclinic is still experimental. Programs that are used in multi-specialty hospitals for inpatient patients might be employed as educational materials in universities, e.g., MIS "Aurora" for multi-specialty hospitals[6] provides a training interface "UMS University" for individual tasks, with a separate entrance for students and teachers, and an access to a server via the Internet. The authors[6] reported a positive impact of the training course on preparing students to work with a MIS.

In regional polyclinics of Kazan city various commercial MISs are used, including BARS.MED, ERP-system, 1C: Medicine (Registry), and VITACORE. These systems have centralized databases with a secure access for each user through standard operating environments (Microsoft, Linux) and a Web browser. Yet, application of training versions of these programs for student training was not reported.

In an international literature, an equivalent of a MIS is an EMR, which combines all types of MISs. A positive effect of using EMR in training of higher medical schools' students was established in[7,8,9] as an important part of an education, while students, in turn, reported that mastering EMR had a positive role in preparing them for a future profession[10,11].

Today, a unified pedagogical approach of teaching students to work with EMR/MIS does not exist[9]. We believe, the reasons are related to a commercial nature of such products, which prevents their free distribution, as well as an absence of a single EMR/MIS that is required for an obligatory use by a local government. Therefore, in order to facilitate medical education, local teams of educators create unique training computer programs that simulate real MIS behavior to support the necessary training. Main characteristics of such training programs (in order to provide an acceptable level of simulation) are their similarity to existing commercial products, a capability to evaluate student's actions and to help him/her during a task execution. This paper presents a preliminary architecture and interface for the MIS training program for students of higher medical schools. As an example, we discuss a consultation of an outpatient patient in a polyclinic with signs of COVID-19 infection.

3. Program architecture and interfaces

3.1 Situation modeling

Collecting information from a patient and conducting an objective examination of the patient comprise a preliminary stage of work. The stage begins from creating am instance for a patient database and filling out a corresponding EMR. A student manually enters his/her name, date and time of a consultation, and a patient's name to open active fields, which consist of complaints at the time of examination, medical history, patient's life history, and objective examination data (Fig. 1).

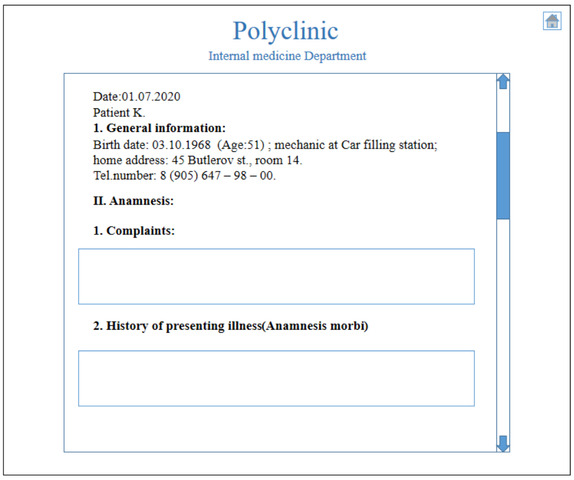


Fig. 1. Collecting information from a patient.

The stage ends with establishing a preliminary diagnosis and entering it in the corresponding field. The patient information is stored in the database and can be reused later by the student at the next stages.

3.2 Analysis

After making a preliminary diagnosis, in accordance with a treatment process algorithm, the patient should be assigned the necessary laboratory and instrumental assessment methods. It is necessary to select an optimal (strictly regulated by current standards) set of methods among proposed ones, which is performed by clicking interactive buttons on the right side of the screen. Next, the student can see all the results in a form of an actual data from a laboratory: a blood test, a CT scan of lungs, an electrocardiogram (Fig. 2).

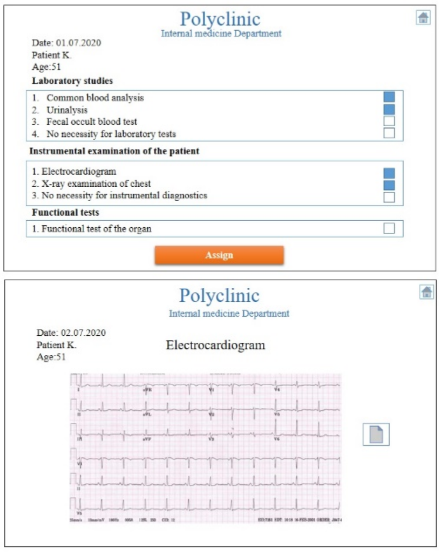


Fig. 2. Laboratory assessment methods.

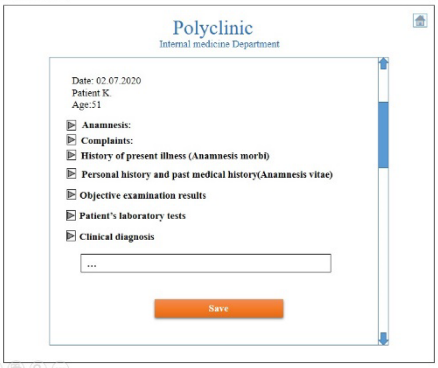


Fig. 3. Consultation and clinical diagnosis.

The student should identify existing deviations from normal ranges and mark them. The next step is to consult the patient with narrow profile specialist(s), which could be selected from a predefined list. In accordance with current medical standards, for certain diseases, these could be a cardiologist, an endocrinologist, a surgeon or other specialists (Fig. 3). The stage ends with creating a clinical diagnosis and recording it into a specially designated field.

3.3 Creating a treatment program

On the next stage the student determines the patient's treatment plan. This is a creative process associated with a specialist qualification. An interface of this page has special fields where you the student can enter all details of the treatment. There is a field for prescribing medications, a recommended regimen, a diet, and other types of treatment (Fig. 4). At this stage, the student has an opportunity to refresh the patient information collected earlier and store it in the database.

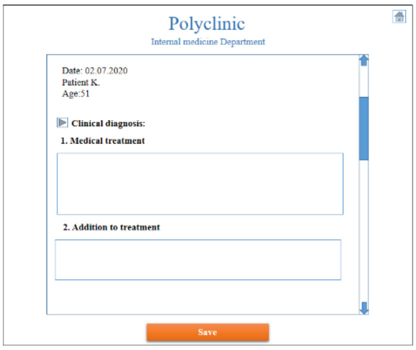


Fig. 4. Treatment strategy.

3.4. Assessment

The student's actions could be assessed by two methods: automatic (conducted by the program) and expert (conducted manually by a teacher). Based on the prescribed algorithm, the program evaluates a completeness of the patient's examination. If the patient information contains empty spaces, the program will require to complete this work.

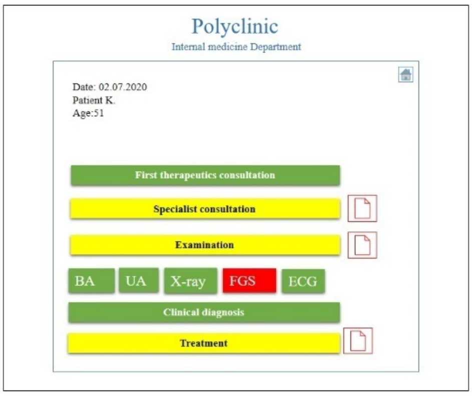


Fig. 5. Mistakes of a student's work.

The program uses color indicators (Fig. 5) as follows. Red color indicates an error, yellow indicates deviations from accepted standards, and green indicates the correct actions. The program allows students to improve knowledge using the interactive buttons that appear near the yellow and red fields. After pressing the button, a student can visit a library in the database and read regulatory documents and standards. A final assessment of the student's work is carried out by the teacher during an individual discussion.

3.5 Example

Today, the problem of treating patients with COVID-19 symptoms at home is of a great social significance. Every graduate medical student must study not only the main symptoms of this disease and know the treatment program, but also take into account nuances of monitoring a potential COVID-19 carrier in an outpatient setting. One of the goals of the program is to combine disparate information (patient data, proposed treatment and additional literature) within a single block to allow transforming theoretical knowledge into practical skills.

We constructed the following example case. The student plays a role of a clinic's entrance department doctor. Patient K., 51 years old, came with complaints of a high fever, a cough, and a general weakness. The student's task is to create an EMR for the patient and determine a medical treatment plan. The program focuses the student's attention on the most typical symptoms of COVID-19. In the "patient's complaints" field main signs of the disease appear in a form of a list. The student should ask the patient more carefully for each symptom (Fig. 6). The preliminary diagnosis of COVID-19 should be confirmed by appropriate laboratory and instrumental assessment (Fig. 7, 8).

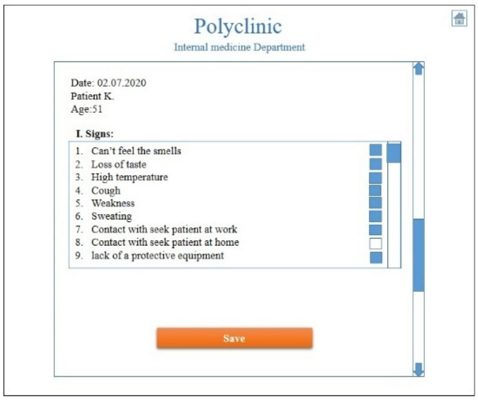


Fig. 6. Patient's complaints.

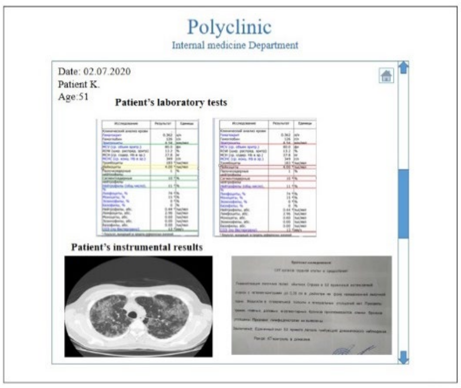


Fig. 7. Laboratory assessment results.

The program generates a list of diagnostic tests that should be used for patients with suspected COVID-19. Next, the student should identify deviations from the normal ranges in laboratory assessment results of the patient. A correct decision is illustrated on the screen with a green background. Errors are marked in red, and incomplete solutions are marked in yellow. In this situation, a special symbol, which appears on the screen, recommends visiting the "library" in the database and studying the standards for laboratory and instrumental assessment of patients with suspected COVID-19. After successfully completing the first two stages and making a clinical diagnosis, the student is asked to make a comprehensive treatment plan for a suspected COVID-19 carrier at home (Fig. 8).

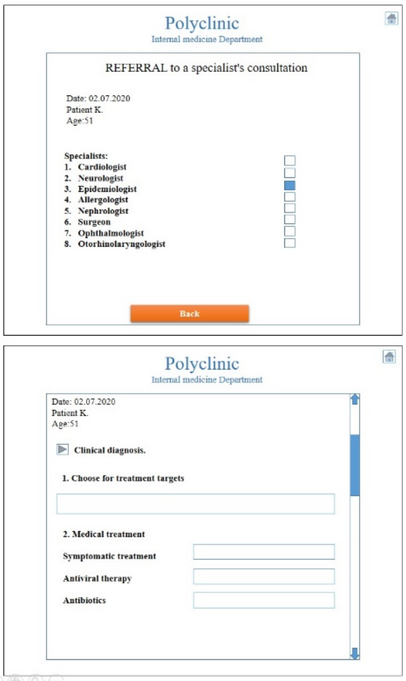


Fig. 8. Final steps to create a clinical diagnosis.

4. Conclusions

The paper presented a preliminary architecture and an interface for the MIS training program for students of higher medical schools. The program simulates main stages of a polyclinic doctor work in monitoring and treating patients at a pre-hospital stage, which opens up opportunities for independent and distance education. It allows creating EMRs and using a database with regulatory documents for outpatients' treatment. The proposed interface is similar to commercial software products currently used in polyclinics of Kazan city, Russia. The program develops practical skills of medical students in modern approach of EMR processing using artificial intelligence-based support for evaluation of outpatient's conditions, diagnosing a disease and selecting its treatment, including a special case of potentially infected with COVID-19 patients.

References

1. L. S. Schneiderand, M. L. Council (eds.), Distance learning in the era of COVID-19, *Arch. Dermatol. Res.,* 2020, pp. 1-2.

2. S. Shah, S.Diwan, L. Kohan, D. Rosenblum, C. Gharibo, A. Soin, A. Sulindro, Q. Nguyen and D. A Provenzano, The technological impact of COVID-19 on the future of education and health care delivery, *Pain Physician,* 23(4S), 2020, pp. 367-380.

3. A. Sagitov, H. Li, N. Schiefermeier-Mach and E. Magid, Extending Gazebo simulator for surgical robotics: tissue and suture modeling. In *International Conference on Artificial Life and Robotics*, 2019, pp. 124-127.

4. H. Lal, D. K. Sharma, M. K. Patralekh, V. K. Jain and L. Maini, Outpatient department practices in orthopaedics amidst COVID-19: The evolving model, *J. Clin. Orthop. Trauma,* 11(4), 2020, pp.700-712.

5. M. Samadbeik, F.Fatehi, M.Braunstein, B. Barry, M.Saremian, F.Kalhor and S. Edirippulige, Education and Training on Electronic Medical Records (EMRs) for health care professionals and students: A Scoping Review, *Int. J. Med. Inform.,* 142:104238, 2020, pp. 1-10.

6. S. I. Karas, I. O. Korneva, M. B. Arzhanik, O. L. Semenova, E. V. Chernikova, O. V. Urneva and A. Y. Grechishnikova, The role and prospects of information and communication technologies in developing of physician competencies, *Doctor and information technology (IDMZ)*, №4, 2018, pp.46-58.

7. N. M. Pageler, C. P. Friedman and C. A. Longhurst, Refocusing medical education in the EMR era, *JAMA,* 310(21), pp. 2249-2250.

8. D. R. Cheng, T. Scodellaro, W. Uahwatanasaku and M. South, An Electronic medical record in pediatric medical education: survey of medical students' expectations and experiences, *Appl. Clin. Inform*, 9(4), 2018, pp. 809–816.

9. A. Rajaram, Z. Hickey, N. Patel, J. Newbigging and B. Wolfrom, Training medical students and residents in the use of electronic health records: a systematic review of the literature, *J. Am. Med. Inform. Assoc.,* 27(1), 2020, pp. 175-180.

10. C. Mountain, R. Redd, C. O'Leary-Kelly and K. Giles, Electronic medical record in the simulation hospital: does it improve accuracy in charting vital signs, intake, and output?, *Comput. Inform. Nurs.,* 33(4), 2015, pp. 166-71.

11. C. M. Yuan, D. J Little, E. S Marks, M. A Watson, R. Raghavan and R. Nee, The electronic medical record and nephrology fellowship education in the united states: an opinion survey, *Clin. J. Am. Soc. Nephrol.*, 15(7), pp. 949-956.

Satisfaction Assessment on the Counseling Service System for Full-Time Teacher-Counselor in Tainan Elementary School

Hsiu-Hao Liu
The Ph.D. Program in Business and Operations Management, College of Management, Chang Jung Christian University
No.1,Changda Rd.,Gueiren District, Tainan City, 711301, Taiwan

Yun-Syuan Jhang
Department Of Counseling and Guidance, National University of Tainan
No.33, Sec. 2, Shu-Lin St., Tainan City, 700301, Taiwan
E-mail: harry2356969@gmail.com, s16784@hotmail.com

Abstract

As the most important case management system in the school counseling work, the counseling service system must meet the needs and expectations of the school teacher-counselors (the main users). This investigation uses the questionnaire survey method aimed at teacher-counselors in elementary schools in Tainan City to understand the degree of teacher-counselors' satisfaction when they use the counseling service system. The results mainly found that counselor-teachers were most satisfied with "The items provided by the system website are completed and various." followed by "The items provided by the system website are necessary items meeting the needs for counseling work." while the least satisfactory item was "Using the system website helps me communicate and interact with others in counseling work." In the five satisfaction aspects of the system, the perceived ease of use and the usage attitude were found to be effective in explaining the system usage intention. Some practical implications of this research as well as topics worthy of further research were discussed at the conclusion of the research.

Keywords: Perceived Usefulness, Perceived Ease of Use, Counseling Service System, Teacher-Counselor

1. Introduction

The counseling service system is the most important tool for case management in counseling work in the junior high and elementary schools in Tainan City. The purpose of the system design is to assist school teacher-counselors to care for students and track their psychological development, and hope that students will have good adaptive growth. The school teacher-counselors are the main users of using the system. In addition to daily counseling affairs, the school teacher-counselors also use the system to record the background information, counseling records, referrals and other operations of the case. In actual use, the school teacher-counselors find that they have encountered difficulties and doubts in the operation of the system, filling in items, and review procedures, and even ethical concerns about the leakage of personal data, which may endanger the relevant rights and protection of the case, and therefore also affect the system usage intention of the school teacher-counselors.

This study is to explore the reasons why the school teacher-counselors accept or reject this system. It is based on the Technology Acceptance Model (TAM) (Davis, 1989) and refers to Ahmad and Khalid (2017), Elena and Detmar (1999), Luarn and Lin (2005), Moore and McElroy (2012), etc., used the five dimensions, these are system quality (SQ), system response time (SRT), perceived usefulness (PU), perceived ease of use (PEOU) and usage attitudes (UA), as the main factors to explore the system usage intention (SUI) of the school teacher-counselors.

The purpose of this study is to explore the current situation of the school teacher-counselors using the counseling service system, and to understand whether the school teacher-counselors are satisfied with the system and the system usage intention due to demographic characteristics (seniority, system using frequency, the average using time), there are differences depending on the difference, and then we will further explore whether the school teacher-counselors satisfaction with the system can effectively explain the system usage intention, and follow-up discussion based on the results.

2. Research Design

The conceptual framework of this study is as follows:

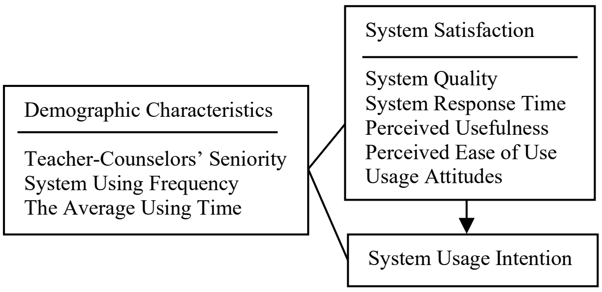


Fig. 1 Conceptual Framework

2-1. Participant

The main objects of this research are full-time teacher-counselors in Tainan Elementary Schools, with a total of 90 teachers who voluntarily participated in answering questionnaires. Finally, 20 teachers submitted responses.

2-2. Data Collection Instrument

This research was based on the self-compiled "Questionnaire on the Status of Professional Counselors in Tainan Elementary Schools Using the Counseling Service System" to conduct a survey, including demographic characteristics (4 questions), system satisfaction (18 questions) and system usage Intention (2 questions) consists of three parts, which are in the form of a Likert four-point scale. The higher the score, the higher the degree of satisfaction or system usage intention.

This questionnaire has a very good internal consistency reliability (α= .92). In terms of validity, 7 senior experts and scholars in the relevant fields, including information, counseling and education, and with seniority between 6 and 20 years are invited to conduct the survey. The content is appropriate for the discussion, so the research tools used in the research should have content validity.

Table 1. Experts and Scholars Background

No.	Present Job	Areas of expertise	Seniority
A	University Assistant Professor	Information Management	10 years
B	University lecturer	Information Engineering	8 years
C	Junior High School Teacher-Counselor	Counseling & Education	20 years
D	Junior High School Teacher-Counselor	Counseling & Education	11 years
E	Elementary School Teacher-Counselor	Counseling & Education	19 years
F	Elementary School Teacher-Counselor	Counseling	6 years
G	Elementary School Teacher-Counselor	Counseling	6 years

3. Result

3.1 System Satisfaction Score

The results of the elementary school teacher-counselors satisfaction items are as follows: "The items provided by the system website are completed and various." (M=2.90), was the highest of all items. "The items provided by the system website are necessary items meeting the needs for counseling work." (M=2.75) was secondly. The least satisfactory

item was "Using the system website helps me communicate and interact with others in counseling work" (M=1.85). The second lowest is "I think the system helps me in the word processing of counseling work is more efficient and convenient" (M=1.90).
From the perspective of the five aspects, except for the system quality, the other aspects have the lowest scores, but none of the aspects have the highest scores. Observing the average score of the aspect again, it can be found that the school teacher-counselors are somewhat dissatisfied with the degree of satisfaction with the system

Table 2. Descriptive Statistics of System Usage Satisfaction

Dimension / Full marks	M	SD	MIN	MAX
SQ/24	15.65	1.79	11	18
SRT/8	4.55	1.43	2	7
PU/8	4.30	1.42	2	6
PEOF/16	8.20	2.48	4	12
UA/16	8.60	2.42	4	12

3.2 System Usage Intention Score

In terms of the performance of the school teacher-counselors in elementary schools, from the highest, lowest and average scores, it can be found that school teacher-counselors have a little low system usage intention.

Table 3. Descriptive Statistics of System Usage Intention

Dimension / Full marks	M	SD	MIN	MAX
SUI/8	4.15	1.50	2	6

3.3 Demographic Characteristics and System Satisfaction

After performing a single-factor variance analysis, it is found that the system satisfaction, system quality, system response time, perceived usefulness, perceived ease of use and usage attitudes of the school teacher-counselors.
There will be no difference due to age, frequency of use and average use time.

Table 4. Summary table of single factor variance analysis (System Satisfaction Dimensions)

Dimension	Seniority	Usage Frequency	Average Use Time
SQ	.61	2.85	.50
SRT	1.13	1.22	.02
PU	1.61	1.89	.04
PEOU	2.03	2.16	.42
UA	.59	1.51	.02

3.4 Demographic Characteristics and System Usage Intention

A single-factor analysis of variance found that the system usage intention of the elementary school teacher-counselors does not depend on seniority ($F=1.48$, $p> .05$), usage frequency ($F= 2.48$, $p> .05$) and average use time ($F= .00$, $p> .05$).

Table 5. Summary table of single factor variance analysis (System Usage Intention)

Dimension	Seniority	Usage Frequency	Average Use Time
SUI	1.48	2.48	.00

3.5 System Satisfaction and System Usage Intention

System satisfaction and system usage intention. are related to the Pearson product difference. It is found that there is a high positive correlation between the two ($r= .89$, $p< .01$), and four of the five dimensions are significantly correlated ($r = .71\sim .87$, $p<.01$).
Only the system reaction time and the willingness to use the system are not significantly correlated ($r=.33$, $p> .05$). It shows that in the usage intention of the school teacher-counselors to use the system, compared with other aspects, the system response time is not the main consideration.

Table 6. Correlations between the Variables of the Study

	SQ	SRT	PU	PEOF	UA	SUI
SQ	-	.55*	.61**	..52*	.66**	.71**
SRT		-	.43	.38	.20	.33
PU			-	.89**	.76**	.87**
PEOF				-	.81**	.83**
UA					-	.85**
SUI						-

**p< .01， *p< .05

Carry out multiple regression analysis based on the five dimensions of system satisfaction and system usage intention, and use stepwise regression as the method of regression model to screen predictive variables. It is found that the two dimensions of perceived usefulness and usage attitude in the degree of system satisfaction can effectively explain the system usage intention (multiple R^2= .84, adjusted r^2= .82, p< .01). It shows that there may be key factors in the system usage intention of the school teacher-counselors in the two aspects.

Table 7. Summary table of multiple regression analysis coefficient

Selected forecast variables and order	Multiple R^2	ΔMultiple R^2	ΔF-value
PU	.76	.76	56.37**
UA	.84	.08	8.74**

**p< .01， *p< .05

4. Conclusion

Based on the above results, it can be found that different seniority, usage frequency, average use time, the elementary school teacher-counselors are more dissatisfied with the degree of system satisfaction. Similarly, elementary school teacher-counselors with different background variables tend to be less system usage intention. In terms of this result, it can be seen that the current counseling service system is not trusted by the elementary school teacher-counselors, and it cannot help the school teacher-counselors to improve their work efficiency or improve the time-consuming trouble of paper processing. It is difficult to obtain the expected benefits in practice.

The main factors influencing the system usage intention are related to the two dimensions of perceived ease of use and usage attitudes. From the items of the two dimensions, it can be found that there may be difficulties in learning to operate the counseling service system, and the security of the system cannot be trusted by the school teacher-counselors. These results show that there was a lack of guidance and communication when the system was launched, and no good consultation channels were provided. What is more worthy of discussion is the management of information security. Since school counseling work has a fairly strict principle of confidentiality, the use of any information is regulated by law.

From this result, it can be found that the school teacher-counselors have great concerns about data preservation or management.

In order to explore the critical factors of the school teacher-counselors system usage intention, this research will conduct qualitative research, integrate the results, and provide the results to relevant agencies to improve the counseling service system.

References

1. Ahmad, S. Z., & Khalid, K. (2017). The adoption of m-government services from the user's perspectives: Empirical evidence from the United Arab Emirates. *International Journal of Information Management*, 37(5), 367-379.
2. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.
3. Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, 35(8), 982-1003.
4. Elena, K., & Detmar, W.S. (1999).The psychological origins of perceived usefulness and ease-of-use. *Information ＆ Management*, 35, 237-250.
5. Luarn, P., & Lin, H. H., (2005), Toward an Understanding of the Behavioral Intention to Use Mobile Banking. *Computers in Human Behavior*, 21(6), 873-891.
6. Moore, K., & McElroy, J. C. (2012). The influence of personality on Facebook usage, wall postings, and regret. *Computers in Human Behavior*, 28(1), 267- 274.

Assess the Critical Factors for the Counseling Service System Usage Intention

Hsiu-Hao Liu
The Ph.D. Program in Business and Operations Management, College of Management,
Chang Jung Christian University
No.1,Changda Rd.,Gueiren District, Tainan City, 711301, Taiwan
E-mail: harry2356969@gmail.com

Abstract

As the most important case management system in the school counseling works, the counseling service system must meet the needs and expectations of the school teacher-counselor (the main users). Semi-structured interviews were used to interview teacher-counselor who have experience in using system in Tainan. The data are recorded and semantic analyzed after interviewing, summarize the critical factors that affect the usage intention of the teacher-counselor. The conclusion was to discuss the implications of this research and suggest several future research issues.

Keywords: Usage Intention, Qualitative Research, Counseling Service System, Teacher-Counselor

1. Introduction

The counseling service system is the most important tool for case management in counseling work in the junior high and elementary schools in Tainan City. The purpose of the system design is to assist school teacher-counselors to care for students and track their psychological development, and hope that students will have good adaptive growth. The school teacher-counselors are the main users of using the system. In addition to daily counseling affairs, the school teacher-counselors also use the system to record the background information, counseling records, referrals and other operations of the case. In actual use, the school teacher-counselors find that they have encountered difficulties and doubts in the operation of the system, filling in items, and review procedures, and even ethical concerns about the leakage of personal data, which may endanger the relevant rights and protection of the case, and therefore also affect the system usage intention of the school teacher-counselors.

Researchers have used self-made questionnaires in previous study. Obtained the result that the school teacher-counselors have low system satisfaction and system usage intention under different background changes. After further analysis, it is found that the two aspects of perceptual ease of use and use attitude will affect the system use willingness. Based on the above research background and qualitative research methods, this study hopes to explore the critical factors that affect the willingness of school teacher-counselors to use the system, and provide suggestions to relevant agencies for improving the system.

2. Research Methods

2-1. Research Design

The structure of qualitative interview questions can be divided into: 1. Structured interviews, which refer to the interview content and questions prepared in advance, also known as standardized interviews; 2.

Unstructured interviews, refer to the undeveloped interview topics, from the focus of the research participants, and ask questions according to the needs of the interview situation, also known as non-standardized interviews 3. Semi-structured interviews, between structured and unstructured, the interview content will be prepared in advance, but the questions will also be adjusted according to the subsequent situation (Lincoln & Guba, 1985).

This research adopts semi-structured face-to-face formal interviews. Through one-to-one in-depth interviews with research participants, the content and data of the interviews are analyzed, reviewed, aggregated, and summarized to determine the dimensions of critical factors that affect the system usage intention.

2-2. Participant

In this study, purposive sampling was adopted, and "information-rich cases" were selected as research participants, and a total of four "information-rich cases" were selected as research participants.

The selection criteria for participants in this study are as follows:

- The current position is as a full-time teacher-counselors in Tainan Elementary Schools (including public and private).
- These teachers come from different schools and still use the counseling service system.
- These teachers are willing to share personal experiences related to research topics.

Through the researcher's own workplace, direct contact with potential participants, a total of four current full-time teacher-counselors who meet the selection criteria and are willing to accept interviews were invited. The basic information of the participants is shown in Table 1.

Table 1. The basic information of the participants

No.	Anonym	Seniority	Interview Date
A	Tea Bag	4 years	2020/11/3
B	Little Buck Teeth	5 years	2020/11/4
C	Agatsuma Zenitsu	5 years	2020/11/6
D	Chibi Maruko Chan	3 years	2020/11/9

2-3. Research Tools

Based on the researcher's previous questionnaire, this study divided the interview content into two parts in view of the critical factors of the system usage intention of the teacher-counselors:

- Explore the main motivations and functions of using system.
- Explore the attraction and resistance factors of using system.

2-4. Data analysis

Before the interview, an interview outline should be drawn up according to the research question and research purpose. After recruiting the research participants, first send the interview outline to the research participants to discuss the direction and content of the interview. During the interview, the whole process was recorded after seeking the consent of the research participants.

After the interview, the recording files of each research participant were compiled into a verbatim manuscript, then the key points were sorted out and analyzed, and the elements and core themes were summarized. Finally, the interview interpretation was written, and the interview interpretation was sent to the research participants In addition, the "Research Participant Interview Interpretation Confirmation Reply Form" is attached. The research participants are asked to check whether the researcher's interpretation of the interview content is incorrect, and correct any doubtful points for further discussion and clarification.

In order to interpret the experience and information of the research participants, the original interviews were interspersed with the description of the thesis when writing the paper. When quoting the content of the interview, if the entire conversation is quoted, it will be presented in italicized block text; if it is a few sentences, it will be framed with " " and changed to italics. The source of the interview content is marked after the quoted paragraph, such as A001: the first English letter is the respondent code, and the last three numbers are the paragraph number of the interviewed content.

When quoting the entire conversation, in order to make the theme more clear, when deleting some interview sentences, "..." will be used. In order to facilitate the smooth reading of the quotation, it will be indicated by [ ] as the text added by the researcher.

3. Discussions and Suggestions

3.1 Discussions

The main purpose of this research is to explore the critical factors of the system usage intention of teacher-counselors. First of all, it can be known from the interview content that the main reasons for the participants to use the system are because of the requirements of the system or regulations, and there is no choice. It is only used under circumstances; and the case record is also because the information is passively registered when a case needs to be referred.

It was because the Student Counseling Center asked us to go to a meeting, and then announced that we would all use the online record to fill in and use it. (A001)

...Because we were actually a crisis incident at that time, we had to ask for the tertiary prevention psychiatrist, which is the part of emergency intervention in school services. Then, [the psychiatrist in the school] he reminded us that we must go to the system, and then make such an application. (B001)

The tertiary prevention part seems to be the Student Counseling Center. If you want to transfer the secondary prevention case to the tertiary prevention. They stipulate that there must be a record that is uploaded. (C002)

Secondly, the background variables of the teacher-counselors are also one of the factors that affect the usage intention. Three participants thought that the frequency of use and the average time of each use would affect the usage intention, because they may become more proficient or unsatisfied as they learn more about the functions of the system.

The reason for frequency of use is closer to the reason for using average time. It refers to the people who use it frequently. He should know more about his advantages and disadvantages. If there are many disadvantages, then of course the less satisfied. If there are many advantages, of course the more satisfaction. (A005)

In addition, two participants believed that age or seniority is also one of the factors that affect the usage intention. Senior teachers with older age or seniority may be less familiar with operating information systems.

Suppose I fill in [the system], or if our senior director or team leader fills in it, they may be less familiar with computer use and it will take longer to fill in. (D002)

Finally, we will discuss the positive and negative opinions of participants on the system. As for the attraction factor, two participants thought that the use of systems would help their work. This is a Perceived Usefulness dimension.

Because before you upload [the Case Record], in order to write more detailed information, you must spend some time to do some case review. I think this is helpful for self-organizing and helping, that is, for self-organizing the case. (B005)

Another participant thought that the system responded smoothly, which is the system response time dimension.

For the resistance factor, the three participants said that the review mechanism of the system is suspicious. One participant reflected the inconvenience of the countdown design and the lack of pre-teaching and post-assistance in the system. All of the above can be summarized in the system quality dimension.

...These team leaders, directors, and principals have seen those cases [records] in everything they reviewed, ..., we would not write them in such detail, because we have the principle of confidentiality and there is a problem of funding. I just think why the director, group leader, and principal would also see it? If it's just a review, then what is the purpose of seeing these things, and what is the purpose of these case records and case conceptualizations? (D010)

I think the countdown timer is very unfriendly. I think, I don't know the purpose of its setting? Then again, for the teacher who uses this system for the first time, I think he is also relatively unfriendly, because he only introduced to the teacher-counselors how to use the system at the beginning, and then I reviewed it again. He actually put a teaching manual on it, but the content is trivial and complicated. (B007)

The four participants agreed that the system cannot protect the safety of case data and that the uploading behavior may cause personal information leakage. This is the usage attitudes dimension.

As far as I know, this system is no longer specifically managed, especially his database for storing data. I heard that there is no special person to manage it. Then how do we know that after we fill in the

information of these cases, Will this information be leaked? Or will it be used? (A009)
As for the confidentiality of this [case records], it should be the unit that requires us to report. They must be themselves, and they must be ethically confidential and protected. Since they have designed this [system], there should be a mechanism. (D012)

Dimensions	Demographic Characteristics	System Quality	System Response Time	Perceived Usefulness	Usage Attitudes
Criterias	1. *Age and Seniority* 2. *System Using Frequency* 3. *The Average Using Time*	1. *Ensure that the audit mechanism is reasonable and confidential* 2. *Provide sufficient time to fill in* 3. *Provide adequate pre-training and post-event assistance*	1. *Smooth system response*	1. *Provide a unified file format* 2. *Contribute to case analysis*	1. *Have a good management authority design* 2. *Ensure the confidentiality of case information and counselling records*

Table 2 The structure of the critical factors for the counseling service system usage intention

3.2 Suggestions

Based on the above, the researcher sorted out the critical factors of the system usage intention of the teacher-counselors. There are five main dimensions and eleven criteria (see Table 2).

Compared with the results of previous questionnaire surveys, the usage attitudes has been mentioned all the time, which shows that this dimension requires special attention. Past research can also confirm that there is a positive correlation between usage attitudes and usage intention (Ahn et al., 2007; Juarez Collazo et al., 2014; Wilson, 2019).

Finally, during the interview, many participants mentioned their concerns about participating in this research. Compared with other teachers, the teacher-counselors are a closed group. If the upper-level authorities use this research to investigate the identity of the participants, it may affect the rights of many teachers. Therefore, it is recommended that researchers pay more attention to the research description and confidentiality when conducting closed group interviews, so that participants can express their own feelings without scruples.

References

1. Ahn, T., Ryn, S., & Han, T. (2007). The impact of web quality and playfulness on user acceptance of online retailing. I*nformation and Management*, *44*(3), 263-275.
2. Juarez Collazo, N. A., Wu, X., Elen, J., & Clarebout, G. (2014). Tool use in computer-based learning environments: Adopting and extending the technology acceptance model. *ISRN Otolaryngology*, *2014*, 1-11.
3. Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
4. Wilson, N. (2019). The Impact Of Perceived Usefulness And Perceived Ease-Of-Use Toward Repurchase Intention In The Indonesian E-Commerce Industry. *Jurnal Manajemen Indonesia, 19*(3), 241-249.

Design and implementation of EMI suppression filter for electronic commutation fan systems

Ching-Chun Chuang [1], Chih-Chian Hua [2], Chung-Wen Huang[2], Chun-Jen Yao[3*]

[1]Department of Computer Science and Information Engineering, National Formosa University, No. 63, Wunhua Rd., Huwei Township, Yunlin 632, Taiwan (R.O.C),
[2]Department of Electrical Engineering, National Yunlin University of Science and Technology, No. 123, University Rd., Section 3, Douliu, Yunlin 640, Taiwan (R.O.C),
[3]Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology, No. 43, Section 3, Keelung Rd., Taipei 106, Taiwan (R.O.C),

*E-mail: *austincc@nfu.edu.tw, huacc@yuntech.edu.tw, wenhung@yuntech.edu.tw, alanyao@mail.ntust.edu.tw*

Abstract

An electronically commutated fan has a complex digital control mode compare to the conventional alternating current fans. Since a serious electromagnetic compatibility situation was produced in the electronic commutation fan system, the filter circuit of a fan system and its mechanical design was significantly important. Because the digital control strategy for an electronically commutated fan was used to reduce the component of the EMI filter so the suppression of conducted and radiated electromagnetic interference will be discussed in this fan system. Moreover, the fan system can be combined with the mechanical design for getting over a series of issues of radiated electromagnetic interference. Thus, fig.1 reveals that a single-phase BLDC motor driver schematic diagram. Eventually, a cause that electromagnetic interference was produced from the electronically commutated fan will be analyzed and its suppression strategy also provided in this paper. The experiment verification for the electronically commutated fan will be made available to the public.

Keywords : Electronically commutated fan, Electromagnetic compatibility, Digital control mode

1. Introduction

The electronic communication fan that combines the frequency conversion function, and the power transistors for frequency modulation to achieve speed control. Because of the higher speed, the motor output efficiency is improved. The Single-Phase BLDC motor system is driven by AC power supply. A DC motor is used inside of this system, so the friction loss is reduced. Thus, the advantages of AC and DC fans are combined to improve the efficiency and life of the overall electronic commutation fan. Fig. 1 show a commercial electronic commutation fan product architecture [1]. A number of switch devices were switched at high speed, and the voltage and current are changes at the same time. The result not only cause more serious electromagnetic interference problems [2] [3] [4], but also affected the electromagnetic tolerance of the internal controller.

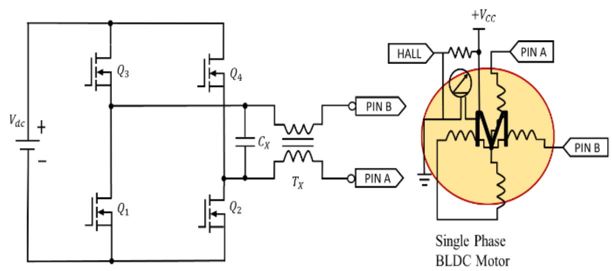


Fig. 1. Single-Phase BLDC motor driver schematic diagram

Therefore, an attention has to consider in the production design for passing relevant electromagnetic compatibility certification.

2. EMI Standard for Electronic Communication System

Generally speaking, Electromagnetic Compatibility (EMC) is divided into two types of tests: Electromagnetic Interference (EMI) and Electromagnetic Susceptibility (EMS) [5][6]. Electromagnetic interference testing refers to measure of the intensity of conductive interference and radiant interference while the electronics device is normal operated under testing. The purpose of electromagnetic interference suppression design is to reduce the intensity of noise generated by the electronics device under testing, and it was not to affect other electronic products in the test environment. On the contrary, the purpose of electromagnetic tolerance is to prevent the DUT from being interfered with by other electronic products in the environment during operations

There are international standards for the electromagnetic interference law of electronic communication fans, which can refer to the European regulations EN 55011, the United States FCC PART 18 and the domestic CNS 13803. Among them, there are two different limit values of class A and class B based on this product are used in commercial or residential households. However, the products analysis in this paper will be used in information products since Europe is the main market. So, the information products for European standards will be adopted. Class B electromagnetic interference standard in EN55022, EN55022[7] and EN55011[8] were the same Class B electromagnetic interference test standard.

3. EMI Suppression and Analysis of Electronic Communication System

3.1 Conduction EMI measurement

To measure the noise values of the live wire (Line, L) and neutral wire (Neutral, N) after a connection of the LISN. The measure results are shown in Fig. 2 and Fig. 3. Observing the two figures, it can be seen that the test value without any filter and other countermeasures far exceeds the standard limit value.

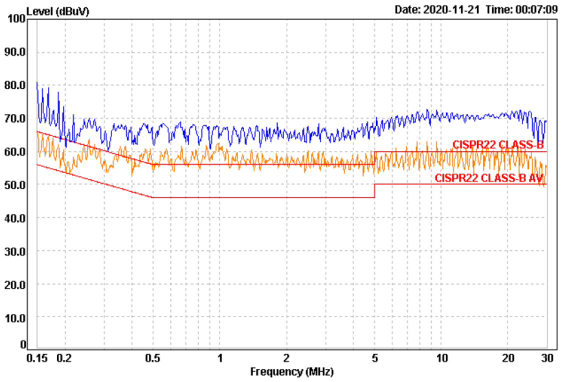


Fig. 2 EMI testing of L Phase conductivity

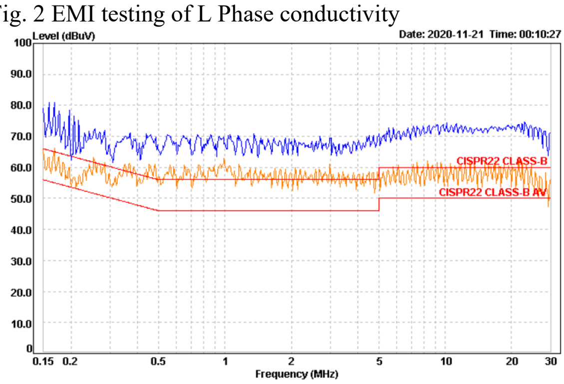


Fig. 3 EMI testing of N Phase conductivity

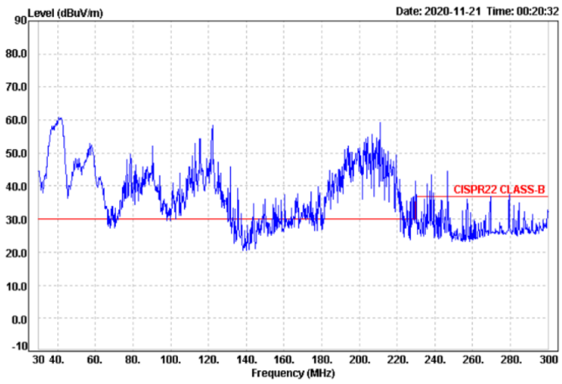


Fig. 4 The vertical polarity test of radiation EMI

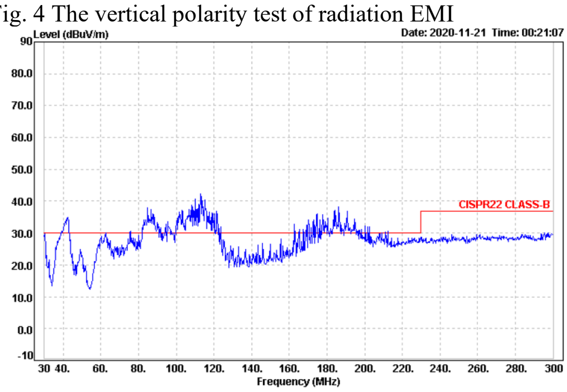


Fig. 5 The horizontal polarity test of radiation EMI

3.2 Radiation EMI measurement

According to section, the noise measurement of the EUT was received through the antenna under two different polarities, vertical and polar. The actual test diagrams are shown in Figure 4 and Figure 5. It can also be seen that the noise radiated through space is very serious.

3.3 The Suppression Strategy for Conduction EMI and Radiation EMI

3.3.1 A Strategy of Conduction EMI

Conducted EMI is mainly used to measure the noise transmitted by the power line through LISN, so the main suppression method is to add an EMI filter to the power line [13]. However, it can be seen that a multi-stage filter with grounding is required to have a wider suppression range from Fig. 2. In this paper, the space of the DUT is limited, and there is no ground wire architecture. The input EMI filter architecture is adopted in the electronic commutation fan systems. The differential mode inductance is 100uH, the X capacitor is 0.1uF, and the common mode inductance is 20mH. Fig. 6 and Fig. 7 show the actual test patterns after adding the input EMI filter and other countermeasures to the EUT internal circuit.

$$Attenuation = 10\log\left(\frac{N \times \delta \times f_c}{f_m}\right) \quad (1)$$

where N is harmonic number

δ is spread spectrum

f_c is switching frequency

f_m is frequency modulation

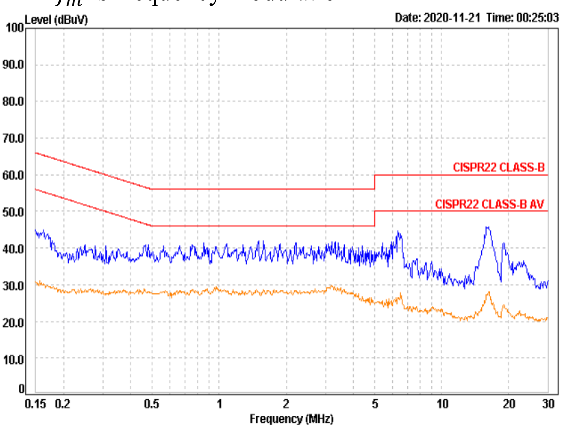


Fig.6 EMI testing of L Phase conductivity after suppression

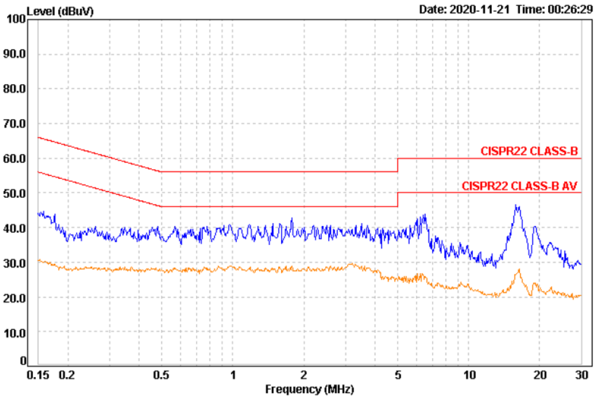


Fig.7 EMI testing of N Phase conductivity after suppression

3.3.2 A Strategy of Radiation EMI

In order to improve the radiation problem between the circuit output and the motor, an output EMI filter design is need in the electronic commutation fan systems. The common mode inductance is 0.95mH and the X capacitor is 4700pF.

The EUT in this paper has two specifications, namely plastic fan blades and metal fan blades. Among them, the series of metal fan blades have the most serious radiated EMI problem. The noise generated by the drive motor will provide a path to the metal fan blade through the metal shaft, which will cause a serious antenna effect. Therefore, an insulating part will be added to the mechanism part, as shown in Fig. 8, which not only saves the cost of motor manufacturing but also greatly improves the problem of radiated EMI. Fig. 9 and Fig. 10 show the actual test patterns after EUT suppression.

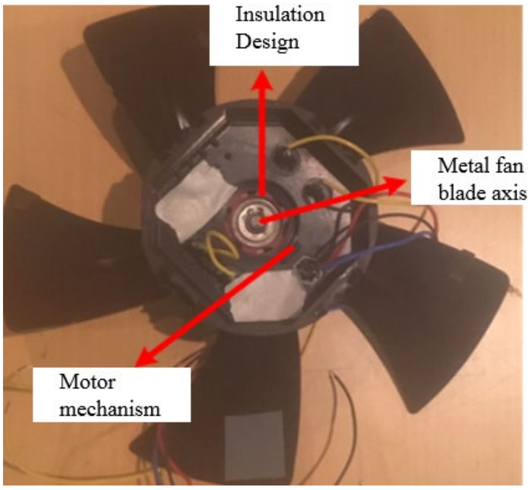


Fig. 8 A strategy for motor shaft and mechanism insulation design

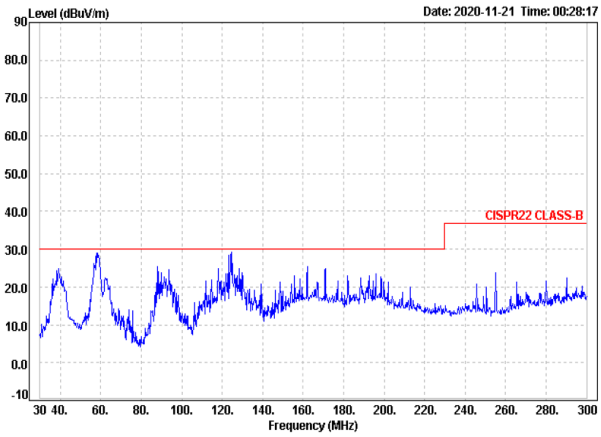


Fig.9 The vertical polarity test of radiation EMI after suppression

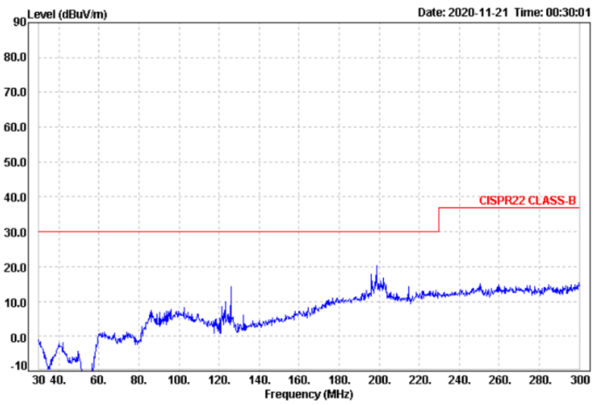


Fig.10 The horizontal polarity test of radiation EMI after suppression

4. Conclusion

This paper proposed an electronic commutation fan that is to perform the actual electromagnetic interference certification process for the EUT. The suppression of conductive EMI and radiate EMI was discussed, and a digital control system was to reduce the EMI filter components in this research. Meanwhile, it combines the mechanism to improve the design to overcome an issue that the system without grounding was operated that causes a serious radiated EMI problem. This method can not only be applied to this product, but also produced the similar effect for the other relative products. The goal that saves time and cost will be implemented after a successful certification of the subsequent relative products

References

1. S. Chen, T. A. Lipo and D. Fitzgerald, "Measurement and Analysis of Induction Motor Bearing Currents in PWM Inverter Drives", IEEE PES Summer Meeting, July 1995.
2. J. Erdman, R. J. Kerkman, D. Schegel and G. Skibinski, "Effect of PWM Inverters on AC Motor Bearing Currents and Shaft Voltages", APEC '95, vol. I, pp. 24-33, Mar. 1995.
3. Henry W. Ott, “Noise Reduction Techniques in Electronic Systems,” John Wiley & Sons, 1988.
4. C. R. Paul, “Introduction to Electromagnetic Compatibility,” John Wiley & Sons, 1992
5. EN 55022, Information Technology Equipment - Radio Disturbance Characteristics - Limits and Methods of Measurement, 2010.
6. EN 55011, Industrial, Scientific and Medical Radio-frequency Equipment – Radio Disturbance Characteristics – Limits and Methods of Measurement, 1998.
7. S. Ogasawara, H. Akagi, “Analysis and Reduction of EMI Conducted by a PWM Inverter-fed AC Motor Drive System Having Long Power Cables,” IEEE Power Electronics Specialists Conference, Vol. 2, pp. 18-23, 2000.
8. M. L. Heldwein, T. Nussbaumer, F. Beck and J. W. Kolar, “Novel Three-phase CM/DM Conducted Emissions Separator,” IEEE Appl. Power Electron. Conf. Expo., pp. 797-802, March 2005.
9. R. L. Ozenbaugh, EMI Filter Design, Marcel Dekker, New York, 1996.
10. J. Kim, P. Jun, J. Kim, “Dithered Timing Spread Spectrum Clock Generation for Reduction of Electromagnetic Radiated Emission from High-speed Digital System,” In Proc., IEEE Int. Symp., Electromagn. Compat., pp 413-418, 2002.
11. Hardin, Fessler and Bush, “A Study of the Interference Potential of Spread Spectrum Clock Generation Techniques,” IEEE International Symposium on EMC, August 1995.
12. A. Santolaria, J. Balcells, D. Gonzalez, J. Gago, and S. D. Gil,“EMI Reduction in Switched Power Converters by Means of Spread Spectrum Modulation Techniques,” Power Electronics Specialists Conference, 2004. PESC 04. 2004 IEEE 35th Annual, vol. 1, pp. 292 – 296, June 2004.

Application of IOT to Forest Management Taking Fushan Botanical Garden as An Example

Shuo-Tsung Chen[1], Chih-Chiang Hua[2], and Ching-Chun Chuang[3*]

[1]Bachelor Program in Interdisciplinary, National Yunlin University of Science and Technology;123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.
[2]Department of Electrical Engineering, National Yunlin University of Science and Technology;123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.
[3]Department of Computer Science and Information Engineering, National Formosa University; No. 63, Wunhua Rd., Huwei Toshnship, Yunlin 632, Taiwan, R.O.C.

E-mail: shough34@yahoo.com.tw, huacc@yuntech.edu.tw, austincc@nfu.edu.tw

Abstract

In recent years, the Internet of Things (IoT) technology has developed rapidly and has been successfully applied in different fields, and has begun to expand the application context of the Internet of Things. This work aims to study how to apply the IoT technology to forestry management, including: 1. Forestry management using wireless network communication technology of Low Power Wide Area Network (LPWAN) such as LoRa and NB-IoT; 2. Apply different sensing technologies to survey resource of forest and monitor the microclimate changes in forest. In order to actually verify the proposed feasible communication technology, sensors, and sensor arrangement, we chose the Fushan Botanical Garden with the most diverse and complex terrain in Taiwan as the experimental site. We actually built LoRa and NB-IoT communication equipment (including repeat equipment) and various sensors to transmit the real-time sensing data. The returned data also verifies the successful operation of various communication devices and sensors.

Keywords: Internet of Things (IoT), Low Power Wide Area Network (LPWAN), LoRa, NB-IoT

1. Introduction

The application of IoT in forestry can be divided into three major aspects: forest environment and resource monitoring planning, forest fire intelligent management and illegal logging prevention. Environmental and resource monitoring should be a more successful part of the application of wireless sensing network (WSN) in the forestry Internet of Things. Zhang et al. [1] proposed a forest environmental factor collection platform based on the ZIGBEE protocol. This platform includes various terminal monitoring equipment such as temperature, humidity, water level, gas, micro-electromechanical system (mems), photoresistor (LDR) and human infrared sensor (PIR), which are all included in the platform's planning and application. Suciu [2] proposed a wireless sensor network architecture with a solution that effectively uses available resources; that is, it tries to optimize the duty cycle of a single sensor (but a single node) or maximize the service life of the network.

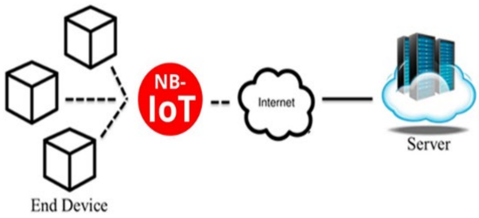
Kelvin Hirschet al. [3] proposed an ecosystem based on Canadian forest fire protection, as shown in Figure 7, which is a practical method of sustainable forest management. The goal is to use forest management practices in an active and planned way to reduce the total area burned by wildfires and the risks associated with the use of designated fires. The authors [4,5] proposed a forest fire detection system, which consists of sensor nodes randomly deployed in the forest area. Each node is equipped with a temperature sensor. Nodes periodically check the environment to determine whether there is an emergency. When some sensor nodes detect a significant change in temperature, they will broadcast a data packet containing their measured value and present the data on the computer web page and mobile phone page. Suguvanam et al. [6] proposed a model to prevent the smuggling of valuable trees such as sandalwood and red sandalwood in forest areas. As shown in Figure 10, their model has three units: 1. Tree unit 2. Sub Server Unit 3. Forest officer.

In this work, we implement the IoT technology for forestry management in Fushan Botanical Garden in Taiwan. In the areas with 3G/4G, we install NB-IoT-support sensors and continuously send out their sensing data by using NB-IoT communication technology. Due to the fact that Fushan Botanical Garden has the most diverse and complex terrain without 3G/4G, the real-time sensing data including illuminance, atmospheric pressure, ultraviolet light, and carbon dioxide are repeated to two repeat stations which is installed on hillside in order. Thus, these sensing data are transmitted to LoRa getway connected with 3G/4G. Finally, the real-time sensing data are transmitted to web server.

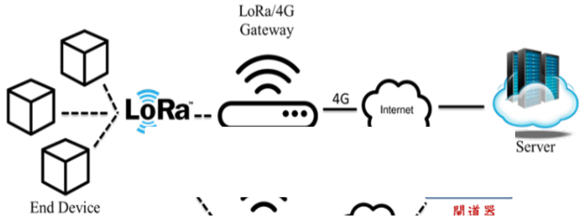
The rest of this work are as follows. Section 2 proposes Forest Sensors and Communication Technology. Section 3 presents the implementation and experiment results. Section 4 lists conclusion.

2. Proposed Forest Sensors and Communication Technology

This section presents proposed forest sensors and communication technology for the Fukuyama Botanical Garden in Yilan County, New Taipei City, Taiwan. A few areas in the Fukuyama Botanical Garden have 3G/4G signals, but most areas do not have 3G/4G signals. Therefore, as shown in Figure 1, we use two wireless network communication technologies of Low Power Wide Area Network (LPWAN), NB-IoT and LoRa, combined with their respective sensors to send the sensing data back to the Forestry Bureau of the Council of Agriculture of the Executive Yuan (Forestry Bureau). The detail is presented in subsections 2.1 and 2.2.



(a) NB-IoT communication technology and its respective sensors



(b) LoRa communication technology and its respective sensors

Fig. 1. Two communication technologies, NB-IoT and LoRa, combined with their respective sensors.

2.1. *NB-IoT communication technology and its respective sensors*

In a few areas with 3G/4G, we use NB-IoT communication technology to continuously return the sensing data from NB-IoT-support sensors.

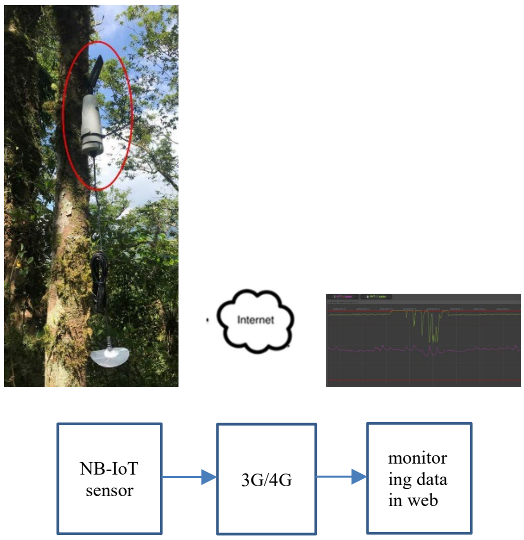


Fig. 2. Flowchart of NB-IoT method for returning the sensing data.

2.2. *LoRa communication technology and its respective sensors*

Due to the complex and diverse terrain of the Fushan site and most areas have no 3G/4G signal, the LoRa signal repeat method of this work will continuously return the sensing data of the LoRa-support solar micro-weather station.

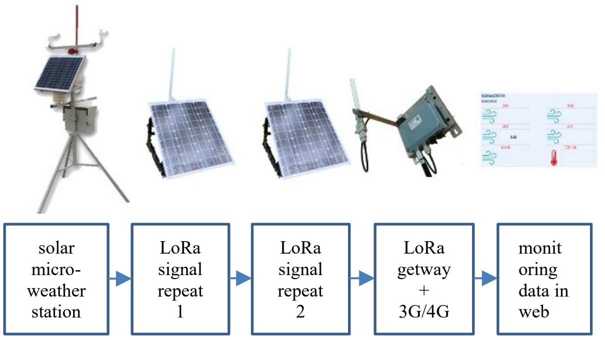


Fig. 3. Flowchart of LoRa signal repeat method for returning the sensing data of the LoRa-support solar micro-weather station.

3. Implementation and Experiment results

This section shows the implementation of the proposed IoT planning applied to forest management and its experimental results.

3.1. *Implementation and Experimental Results of NB-IoT Planning*

In the areas with 3G/4G, as shown in Figure 4, we first install NB-IoT-support sensors as shown in Figure 5 and continuously send out their sensing data by using NB-IoT communication technology. Next, the real-time sensing data including temperature and humidity are connected with 3G/4G and finally transmitted to web server as shown in Figure 6.

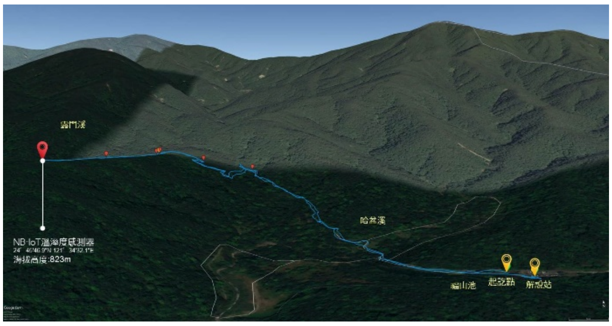


Fig. 4. Implementation of NB-IoT-support sensors and communication technology.

Fig. 5. NB-IoT-support temperature and humidity sensors.

No	Device Type	Device Name	Device MAC	Voltage	Report Time	Bat.low
1	Sensor	-	7D6C603A00002026	5.50	2018-11-13 11:09:21	
2	Sensor	-	7D6C603A00002544	5.50	2018-11-12 23:20:36	
3	Sensor	-	7D6C603600004A2F	4.50	2018-11-02 10:10:59	
4	Sensor	-	7D6C603600004B24	5.50	2018-10-31 18:09:09	
5	Sensor	-	7D6C603D00004D2E	5.50	2018-10-27 03:03:14	
6	Sensor	-	7D6C603C00001B3D	5.50	2018-09-20 12:06:10	

Fig. 6. Real-time data of temperature and humidity.

3.2. *Implementation and Experimental Results of LoRa Repeat Planning*

In the large areas without 3G/4G, we first install LoRa-support solar micro-weather station in the top of some mountain as shown in Figure 7 and continuously send out sensing data by using LoRa communication technology. Next, the real-time sensing data including illuminance, atmospheric pressure, ultraviolet light, and carbon dioxide are repeated to two repeat stations which is installed on hillside in order. Respectively, these sensing data are then transmitted to LoRa getway which is connected with 3G/4G. Finally, these real-time sensing data are transmitted to web server as shown in Figure 8.

Fig. 7. Implementation of LoRa-support solar micro weather station, two repeat stations, and communication technology.

Fig. 8. Real-time data of illuminance, atmospheric pressure, ultraviolet light, carbon dioxide.

4. Conclusions

In this work, we study how to apply the IoT technology to forestry management in Fushan Botanical Garden in Taiwan. In the areas with 3G/4G, we install NB-IoT-support sensors and continuously send out their sensing data by using NB-IoT communication technology. Due to the fact that Fushan Botanical Garden has the most diverse and complex terrain without 3G/4G, the real-time sensing data including illuminance, atmospheric pressure, ultraviolet light, and carbon dioxide are repeated to two repeat stations which is installed on hillside in order. Respectively, these sensing data are then successfully transmitted to LoRa getway which is connected with 3G/4G. Finally, these real-time sensing data are transmitted to web server.

References

1. Zhang Yu, Liu Xugang, Geng Xue, Li Dan, “IoT Forest Environmental Factors Collection Platform Based on ZIGBEE,” CYBERNETICS AND INFORMATION TECHNOLOGIES, Volume 14, Special Issue, 2014, pp.51-62.
2. George Suciu, Ramona Ciuciuc, Adrian Pasat, and Andrei Scheianu, “Remote Sensing for Forest Environment Preservation,” Advances in Intelligent Systems and Computing, 2017, DOI 10.1007/978-3-319-56538-5_23
3. Kelvin Hirsch, Victor Kafka, Cordy TJlmstra, Rob McAlpine', Brad Hawkes, Herman Stegehuis, Sherra Quintilio, Sylvie Gauthier, and Karl Peck, “Fire-smart forest management: A pragmatic approach to sustainable forest management in fire-dominated ecosystems,” THE FORESTR Y CHRONICLE MARCWAPRlL, VOL. 77, NO.2, pp. 357-363, 2001.
4. Pouya Bolourchi and Sener Uysa, “Forest Fire Detection with Wireless Sensor Networks,” 2013 Fifth International Conference on Computational Intelligence, Communication Systems and Networks.
5. Murat Dener, Yunus Özkök, Cevat Bostancioglu, “Fire Detection Systems in Wireless Sensor Networks,” World Conference on Technology, Innovation and Entrepreneurship, Social and Behavioral Sciences, 195 (2015) pp.1846-1850.
6. Suguvanam K, Senthil Kumar, Partha Sarathy, Karthick, Raj Kumar, “Innovative Protection of Valuable Trees from Smuggling Using RFID and Sensors,” International Journal of Innovative Research in Science, Engineering, and Technology, Vol. 6, Issue 3, March 2017, pp.3836-3845.

Development of the IoT module using MQTT Protocol and AES

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 †

Lin Tzu Yuan

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
E-mail: jrhung@yuntech.edu.tw, m10812021@yuntech.edu.tw, khhsia@yuntech.edu.tw

(leave space here should be 4 lines between *www.university_name.edu* and **Abstract***)*

Abstract

The efficiency and safety of the Internet of Things (IoT) have always been the focus of the development of IoT devices. Because the chips used in IoT devices generally have poor computing power, they cannot transmit data quickly and in large amounts, and use more complex security algorithms. Therefore, this thesis is to develop an IoT module with STM32 chip as the main controller. This module uses MQ Telemetry Transport (MQTT)[1] Protocol and AES encryption technology, and this IoT module can be operated directly with a browser. MQTT is a communication protocol for the Internet of Things, which was developed by IBM and Eurotech, and officially became an OASIS international standard in 2014. The purpose of development is to send and receive processing messages under narrow bandwidth and low energy consumption conditions. To ensure the security of IoT communications, we use AES encryption technology. Through this design, the communication of the entire IoT module is more efficient and safe. Finally, we applied this IoT module to the home security system, and the overall efficiency and safety have been verified. In the future, we will continue to improve related software and hardware so that this IoT module can be used in different fields.

Keywords: *Internet of Things (IoT), MQ Telemetry Transport (MQTT), AES, Home Security System.*

1. Introduction

The devices of the IoT have been widely used in our living environment, and these devices have brought many conveniences to our lives. However, the efficiency and safety of these devices are still very hot topics. In previous studies such as Parkhomenko, Anzhelika, et al. [2], the efficiency and safety of Smart House Systems have been studied. It concluded that the safety of Smart House Systems still needs to be strengthened. AHMAD, Farhan, et al. [3] researched the safety of VANET (Vehicular Ad-hoc NET) Because VANET has a great relationship with people's life and property safety, how to ensure the safety of communication and information is

.

very important. MENEGHELLO, Francesca, et al. [4] researched the security issues of the Internet of Things and proposed related solutions. NAIK, Nitin. [5] compares and analyzes the efficiency and security of several commonly used IoT communication protocols such as MQTT, CoAP, AMQP and HTTP.

From previous research, we can find that the communication efficiency and security of IoT are still important research topics. Therefore, this article uses the STM32 chip to develop an IoT module and uses the MQTT protocol as the entire system communication protocol. MQTT is a communication protocol invented in 1999 by Dr. Andy Stanford-Clark of IBM and Dr. Arlen Nipper of Arcom (renamed Eurotech) [6]. This protocol was originally designed to allow IoT devices to effectively transmit messages even with limited bandwidth and computing power. And this agreement also became an international standard of OASIS (Organization Advancement Structured Information Standards) in 2014. Although this standard is designed for communication between IoT devices, large network service providers such as Facebook Messenger[7] and Amazon Iot[8] also use the MQTT protocol. It can be seen that the MQTT protocol is still one of the important protocols for network communication and IoT systems. Therefore, we chose MQTT as the main communication protocol of the Internet of Things system in this paper.

In this paper, we have also modified the MQTT protocol, so that the original MQTT protocol, which can only be transmitted in one direction and cannot be operated, becomes a system that allows users to operate. Punctuation marks are used at the end of equations as if they appeared directly in the text. In order to ensure the security of the entire system, we have also adopted AES to ensure the security of the communication and data of the entire system.

Through the integration of MQTT and AES, and the IoT module developed in this paper using STM32 chips, the relevant software and hardware of this paper can provide a safe and efficient system. And used in different fields.

2. System Architecture

MQTT is a communication protocol invented in 1999 by Dr. Andy Stanford-Clark of IBM and Dr. Arlen Nipper of Arcom (renamed Eurotech). They were in order to provide a lightweight and reliable binary communication protocol between the oil pipeline sensor and the artificial satellite under the premise of the narrow network bandwidth and small power loss. MQTT originally stood for Message Queueing Telemetry Transport. This kind of expression is no longer used. MQTT is MQTT, not an abbreviation of other words.

Because the message content of the MQTT protocol is very streamlined, it is very suitable for IoT devices with limited processor resources and network bandwidth. Many MQTT libraries have been developed for Arduino control boards (C/C++), JavaScript (Node.js, Espruino control boards), Python, etc. There is also an open source MQTT server, which makes it very easy to develop the MQTT IoT and the communication between machines (Machine-to-Machine, M2M).

The advantage of MQTT is that it is a lightweight protocol. Since it is a protocol designed for the Internet of Things, the network bandwidth it requires is very low. And the required hardware resources are also low. It is very suitable for IoT environments with low power consumption and limited network bandwidth. Such as smart appliances or medical devices.

MQTT uses the Publish/Subscribe mechanism for messaging, which contains 4 main elements, Publisher, Subscriber, Topic, and Broker. Among them, the IoT module is the Publisher, and the information that the IoT module needs to send out is the Topic. And these information Topics will not be sent directly to the demand-side Subscriber, but sent through the Broker. So the Publisher and the Subscriber are not directly connected. But because of this, the information and communication security of the entire IoT system must be controlled through a Broker.

At present, most Brokers have SSL (Secure Sockets Layer) encryption mechanism when transmitting data, but SSL is too complicated for IoT devices. Therefore, this paper uses AES (Advanced Encryption Standard) technology to ensure the security of all communication and data in the communication of the IoT module and the communication of the Broker.

Finally, we also modified the MQTT architecture. Because the original MQTT architecture Subscriber can only simply respond to whether it has received a Topic from the Publisher. But in the application of the IoTs system, in many cases, the Subscriber needs to send messages or control the Publisher. Therefore, the revised MQTT architecture of this paper is shown in Fig. 1.

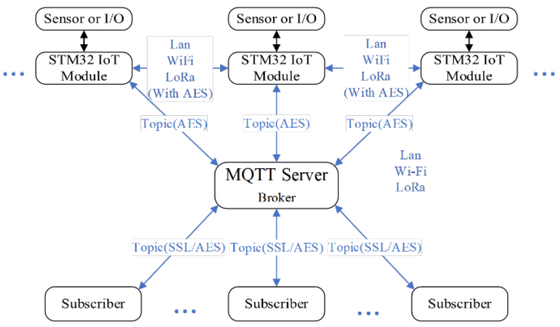


Fig. 1. The Revised MQTT Architecture Diagram of This Paper.

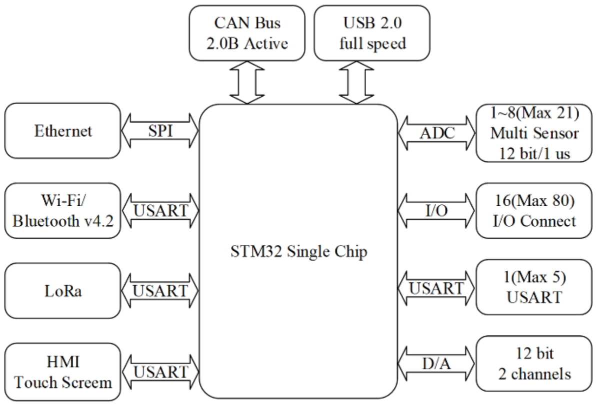


Fig. 2. Base on STM32 Chip IoT Module Block Diagram.

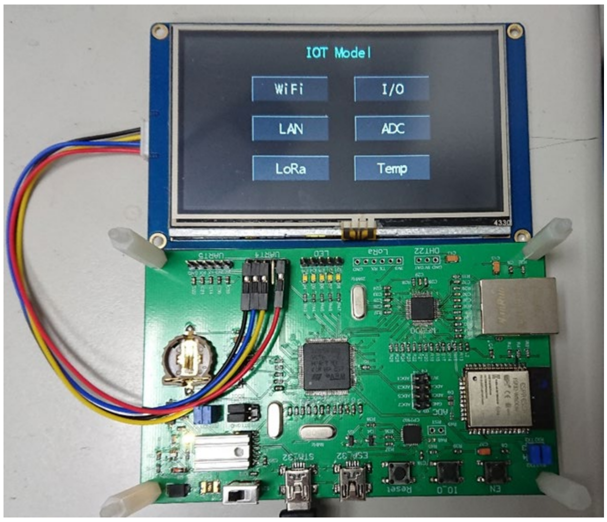
2.1. *Hardware Architecture*

The IoT module developed in this paper uses a 32-bit STM32 series microcontroller. This series of microcontrollers are designed based on ARM Cortex™-M. Mainly to provide MCU users with new development freedom. It integrates high performance, real-time functions, digital signal processing, low power consumption and low voltage operation. At the same time, it is highly integrated and easy to develop. For the IoT module developed in this paper, the microcontroller we chose is STM32F103VET6. It is an integration CPU, RAM, ROM, a variety of I/O ports (including display drive circuit, PWM, analog multiplexer, A/D converter, etc.) and interrupt system, timer/counter and other functions constitute a microcomputer system. Because of its better stability and low price, this paper chooses this chip as the main controller of the IoT module.

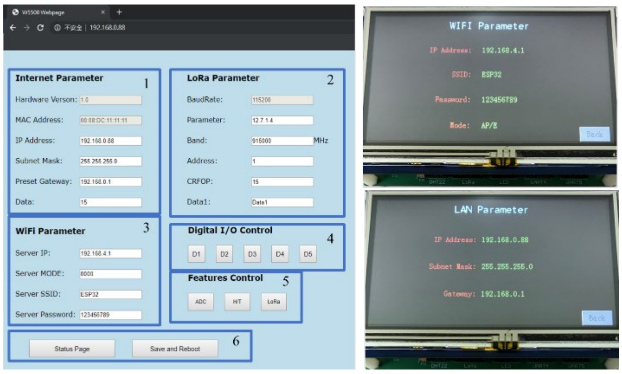
In the communication interface part, the IoT module developed in this paper can connect to Wi-Fi, Ethernet, and USART (Universal synchronous/asynchronous receiver transmitter), LoRa (Long Range) etc. for the communication interface. Moreover, this IoT module adopts a modular design, so you can choose to use different communication modules according to different needs. And in order to ensure the security of data and communication, we use AES encryption technology in this IoT module. Let this IoT module provide secure data communication and transmission. The block diagram of this IoT module is shown in Fig. 2.

3. Experimental Results

We use STM32 chip as the main controller of the IoT module developed in this paper. This module adopts modular design, so most of the communication interface can be replaced according to needs. And we also add a touch LCD to this IoT module. When necessary, we don't need to connect with the host, and directly operate through the touch LCD. In addition, we also added the function of Web Server to this IoT module. Allow users to directly use the browser to control this IoT module. But for the sake of security, the MQTT setting still has to be set through the Broker. Through this design, the IoT module we developed is both convenient and safe. Fig. 3 shows the browser interface of the IoT module developed by us, and the screen of the entity with the touch LCD and the IoT module connected with the browser.



(a)IoT Module.



(b)IoT Module Browser and LCD Screen.

Fig. 3. STM32 chip IoT Module and Screen.

3.1. *MQTT Broker System and Experiment Results.*

In the MQTT Broker, we use Visual Studio 2017 C# to develop our system. The picture of this system is shown in Fig. 4. In this system, users can define the map of the environment and the related attributes of the IoT module. And the I/O or A/D related parameters of which IoT module the Topic corresponds to. In addition, this system also provides an automatic search function. The system can automatically search for a Ethernet or Wi-Fi was the IoT devices having MQTT function. After searching, it will automatically name the I/O or A/D on the device based on the IP. This simplifies the user naming operation, and the system management of MQTT Topics will be more efficient. The rules for this automatic Topics naming are as follows:
Topic Rule:
INDEX,I/O Type,I/O Locate,IP, PORT

Table 1. Topic Rule Description

Field	Description
INDEX	After the IP of the IoT module is found, the sequence of saving it into the system.
I/O Type	I/O category, such as I/O, A/D, D/A, etc.
I/O Locate	I/O Type corresponds to the physical location of the IoT module.
IP	IoT module IP
PORT	The communication port used by the IoT module.

Finally, we apply the IoT module developed in this paper and the MQTT system to home security. We connect the IoT module to the temperature and humidity sensor and integrate it with the MQTT system we developed. The result is shown in Fig. 5.

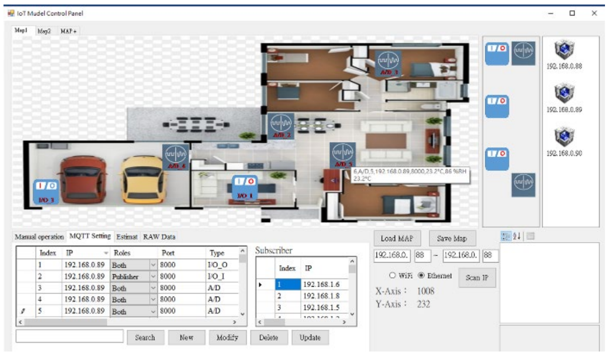


Fig. 4. MQTT Broker Software System with MQTT Topic Names Sample.

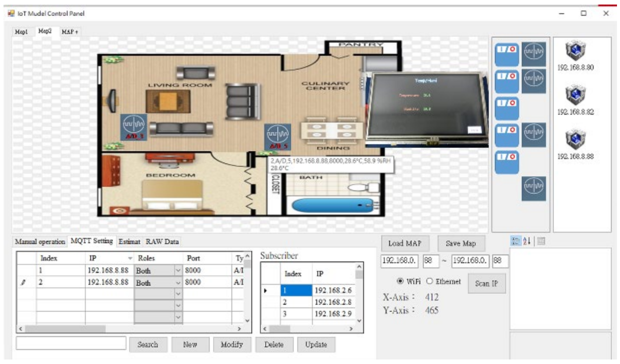


Fig. 5. MQTT Broker System with IoT Module.

4. Conclusions

This paper uses STM32 chip to develop a multi-functional IoT module, and uses MQTT protocol to develop an application system. Finally, the software and hardware are integrated and applied to the home security system. The performance and condition of the entire operation are very good. At present, we put the MQTT management interface and the user's operation interface together. Although this design has its convenience, there may be doubts in practical application and safety. Therefore, in the future, we will separate the management interface from the user's operating interface, and convert this system to a web. I believe that the overall ease of use and safety can be improved. In the future, we will continue to add artificial intelligence and more sensor functions. In addition to being used in home security, this system can also be used in different fields such as smart cities and Industry 4.0.

5. References

1. http://public.dhe.ibm.com/software/dw/webservices/ws-mqtt/mqtt-v3r1.html,2020/12/01
2. Parkhomenko, A., Tulenkov, A., Sokolyanskii, A., Zalyubovskiy, Y., Parkhomenko, A., & Stepanenko, A. (2018, March). The application of the remote lab for studying the issues of Smart House systems power efficiency, safety and cybersecurity. In International conference on Remote engineering and virtual instrumentation (pp. 395-402). Springer, Cham.
3. AHMAD, Farhan, et al. A comparative analysis of trust models for safety applications in IoT-enabled vehicular networks. In: 2019 Wireless Days (WD). IEEE, 2019. p. 1-8.
4. MENEGHELLO, Francesca, et al. IoT: Internet of Threats? A survey of practical security vulnerabilities in real IoT devices. IEEE Internet of Things Journal, 2019, 6.5: pp. 8182-8201.
5. NAIK, Nitin. Choice of effective messaging protocols for IoT systems: MQTT, CoAP, AMQP and HTTP. In: 2017 IEEE international systems engineering symposium (ISSE). IEEE, 2017. p. 1-7.
6. http://www.steves-internet-guide.com/mqtt/,2020/12/01
7. https://www.facebook.com/notes/facebook-engineering/building-facebook-messenger/10150259350998920,2020/12/01
8. https://aws.amazon.com/tw/blogs/aws/aws-iot-cloud-services-for-connected-devices/,2020/12/01

Exploring the Intention to Continuance of Learning Programming at Elementary School of Rural Area by the mBot Robot

Yung-Hsin Cheng
Department of Information Management, National Yunlin University of Science and Technology,
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.

Jia-Ming Hsiao
Bachelor Program in Interdisciplinary Studies, National Yunlin University of Science and Technology,
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.
E-mail: cindy6092624@gmail.com, hsiaojm@yuntech.edu.tw
https://www.yuntech.edu.tw

Abstract

Since Curriculum Guidelines of 12-Year Basic Education implemented by the Ministry of Education in 2018, the program learning courses have been added to junior high school education. However, there are no programming course in the elementary school. This study is proposed to explore the continuity and intentions of the rural area students in the programming course with mBot robot and mBlock programming tool through the Post-Acceptance Model of IS Continuance. It is indicated that enlightenment education of computational thinking should be implemented during the elementary school by means of graphical programming software and robot practice. Through the graphical software and teaching robots, we can cultivate problem solving skills for students' logic, creative thinking and communication, as well as through the task-guided way to train the students' concentration and perseverance.

Keywords: Rural Education, Robot programming, Makeblock mBot, Post-Acceptance Model of IS Continuance.

1. Introduction

In 2018, the Curriculum Guidelines of 12-Year Basic Education had been implemented by the Ministry of Education, Taiwan. One of the major changes is that curriculum development in the field of technology in junior and senior high school should be compatible with information technology and life technology. The programing language courses have been added to junior high school education. However, the programming course is not essential but implemented as a flexible curriculum in the elementary school. It is indicated that enlightenment education of computational thinking should be implemented during the elementary school by means of graphical programming software and robot practice. Through the graphical software and education robots, we can cultivate problem solving skills for students' logic, creative thinking and communication, as well as through the task-guided way to train the students' concentration and perseverance. On the other hand, it is also observed that there exists digital divide in education between urban and rural schools. The digital divide can be reduced by the skill training of software, programming and problem solving. In the mean time, many volunteers provide learning activities of programing and robot education for rural school to assist in teaching and learning in Taiwan.

mBot is an education robot for beginners, that was announced by Makeblock Co., Ltd in 2015. In Taiwan, it is often used in elementary and junior high school to develop the students' computational thinking and programming skills. The programming tool is mblock which is a block-based and code-based programming software developed from Scratch. The mBot robot course is aimed on the Grade 3 to 6 students of elementary students. There are four topics that include logical thinking, sensors, differential driven wheel robot and project practice to cultivate problem solving skills for students' logic, creative thinking and communication, as well as through the task-guided way to train the students' concentration and perseverance.

Expectation-Confirmation Theory was proposed by Oliver[1]. The sustainability theory in the IS research field draws on Expectation-Confirmation Theory in consumer behavior. However, Bhattacherjee[2] believes that the Expectation-Confirmation Theory is controversial and unreasonable in some aspects. He proposed post-acceptance model of IS continuance by the concept that continuance usage intention of information system will be influenced by usage satisfaction and perceived ease of use. Many scholars conduct related research based on the post-acceptance model of IS continuance, but there are not many related researches on education. Based on the post-acceptance model of IS continuance, the students' perceived usefulness of the curriculum can be used as a reference benchmark for confirmation and judgment, which will positively affect their satisfaction with the curriculum[3].

This study is proposed to explore the continuity and intentions of the rural area students in the programming course with mBot robot and mBlock programming tool through the post-acceptance model of IS continuance. The purpose is to explore students' satisfaction with programming courses, and to understand whether it can stimulate students' motivation to continue participating advanced programming courses.

2. Research structure

This study is based on the Post-Acceptance Model of IS Continuance, as shown in Fig. 1. It is implemented by the following steps:

(i) Teaching and learning mBot,
(ii) Example exercise,
(iii) Group project,
(iv) Questionnaire collecting and analysis.

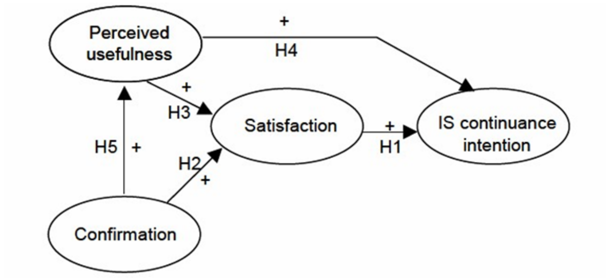


Fig. 1 Post-Acceptance Model of IS Continuance[2]

2.1. *Teaching and learning mBot*

There are four rural elementary schools participated in this study. The course is implemented in the formal course in two schools, and in the student club in the other two schools. There are 85 students participated in the course. The students are divided into groups in which there two members each. Group members are encouraged to show their team-work spirit and knowledge sharing.

2.2. *Example exercise*

Each lesson is arranged to focus on one theme, e.g., motor control, sensor application, etc. A short example is given in class, and the students are asked to test and practice an exercise. Both mechanical and programming discussion are applied to each group. The subject of exercise may be from the missions of MakeX robot competition.

2.3. *Group project*

In the last two or three lessons of the semester, a group project is assigned to each group. The project assigned to Grade 3 and 4 students is RoboRAVE a-maze-ing problem, and Grade 5 and 6 students are assigned to solve the line-following and object avoiding problems.

2.4. *Development of research model and propositions*

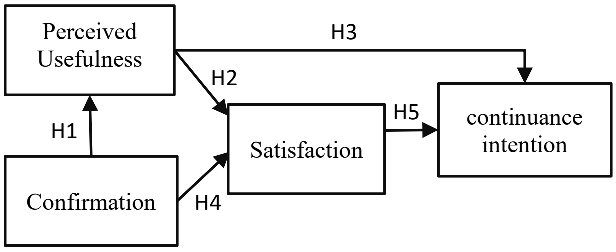


Fig. 2 research framework

Based on the Post-Acceptance Model of IS Continuance, the research framework is shown as Fig. 2. There are five hypotheses made in this study:

H1: Users' confirmation has the significant influence on satisfaction.
H2: Users' perceived usefulness has the significant influence on satisfaction.
H3: Users' perceived usefulness has the significant influence on continuance intention of programming course.
H4: Users' confirmation of programming course has the significant influence on satisfaction.
H5: Users' satisfaction of programming course has the significant influence on continuance intention of programming course.

The four factors of research framework are defined as followings:

- perceived usefulness: Students subjectively feel that programming courses are helpful for learning.
- confirmation: After students experience the programming course, their before and after psychological expectations of the experience are the same.
- satisfaction: How the students feel about the programming course after the experience.
- continuance intention: Students are willing to continue learning programming courses in the future.

3. Main results

After the programming course, a questionnaire is issued to students. There received 68 questionnaire samples, and 65 samples are valid, the other 3 samples are invalid. The students' basic personal information are listed in Table 1.

Table 1 Students' basic personal information

Basic personal information		Count	Percentage(%)
Gender	Male	43	66.2
	Female	22	33.8
Age	<= 10	37	56.9
	11	15	23.1
	12	8	12.3
	> 12	5	7.7
Grade	3	14	21.5
	4	30	46.2
	5	8	12.3
	6	13	20.0
How long to learn mBot?	1 semester	57	87.7
	2 semesters	7	10.8
	3 semesters	0	0
	4 semesters	1	1.5
Experience coding?	Yes	47	72.3
	No	18	27.7
Interested in coding before learning mBot?	Yes	55	84.6
	No	10	15.4
Using computer or tablet after school?	Yes	60	92.3
	No	5	7.7
Average time of using computer or tablet every day	0	5	7.7
	< 1 Hr	27	41.5
	1~2 Hr	14	21.5
	2~3 Hr	7	10.8
	> 3 Hr	12	18.5

After the data analysis, the result of five hypotheses is shown as Table 2. It is shown that hypotheses H1, H2, H4 and H5 have significant influence, and H3 does not.

Table 2 Data analysis result

Hypothesis	Path coefficient(β)	t-value	p-value
H1	0.691	6.335	< 0.001***
H2	0.410	3.273	< 0.001***
H3	-0.005	0.053	0.958
H4	0.518	4.769	< 0.001***
H5	0.925	11.782	< 0.001***

4. Conclusions

From Table 1, it is shown that there are 72.3% students experienced programming course, 84.6% students are interested in coding before learning mBot, and 87.7% students have one-semester experience of learning mBot.

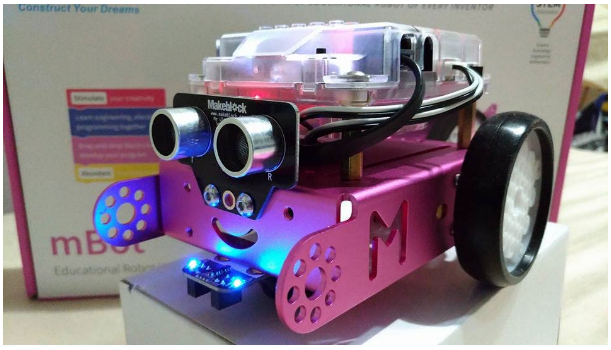
This research believes that after experiencing the mBot programming course, students have more willing and motivation to learn to code, and are willing to learn about other programming after school.

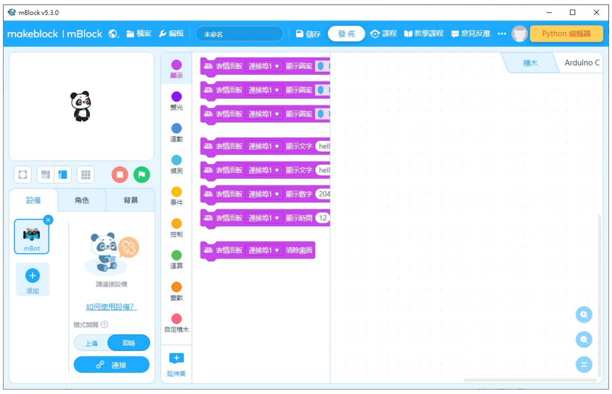
We observed that most of the students are a little afraid of learning the new programming course before the course. For the brand-new programming course, the original expectation was low. After experiencing the mBot programming course, through the combined teaching of mBlock and mBot robot, the expectation after the course was greater than expected. With the mBot robot programming course, students can learn and understand programming applications faster with fun, so as to be satisfied with the mBot robot programming course. With respect to the degree of confirmation of students in the mBot program course, it is indicated that the higher their degree of confirmation, the higher their satisfaction. It is confirmed that the satisfaction of mBot programming courses will positively affect the students' intention to continue participating programming courses. It is observed in this research, students' satisfaction with programming courses makes them willing to spend more time on learning programming, and students will be more willing to actively learn advanced programming courses.

However, students' enthusiasm may decay along with time passed. It could result from that students do not know why they should learn robot programming. Teachers should link the robot programming course to real life applications or examples. With the linkage, students are guided to build and program useful robots to solve problems. It is also important to encourage students to share their ideas of building robots, and then to maintain students' enthusiasm for learning programming and robot building.

References

1. Oliver, R. L. and Richard, L., A cognitive model for the antecedents and consequences of satisfaction decision, *Journal of Marketing Research*, 17(4), 1980, pp. 460-469.
2. Bhattacherjee, A., Understanding information systems continuance: An expectation-confirmation model, *MIS Quarterly*, 25(3), 2001, pp. 351−370.
3. Lee, M.-C., Explaining and predicting users' continuance intention toward e-learning: An extension of the expectation–confirmation model, *Computers & Education*, 54(2), 2010, pp. 506-516.





Basic Research on Parameter Tuning Skills Evaluation Based on Sensor Car Behavior Data in Technology Education

Teruyuki Tamai
Graduate School of Humanities and Social Sciences, Hiroshima University / Faculty of Education, Ehime University
3, Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan

Yoshihiro Ohnishi
Faculty of Education, Ehime University
3, Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan

Kazuo Kawada
Graduate School of Humanities and Social Sciences, Hiroshima University
1-1-1, Kagamiyama, Higashi-hiroshima, Hiroshima 739-8524, Japan
E-mail: tamai.teruyuki.xq@ehime-u.ac.jp, ohnishi@ehime-u.ac.jp, kawada@hiroshima-u.ac.jp
www. ehime-u.ac.jp, www. hiroshima-u.ac.jp

Abstract

The "Future Investment Strategy 2018" presented by the Cabinet Office in 2018 states that there is an urgent need to develop human resources who can use AI to achieve the goal of Society 5.0. As a part of this strategy, programming education in primary/secondary level of education is being promoted. However, there are some issues in the quantitative evaluation of learning and teaching methods based on evaluation. In this paper, sensor car teaching materials will be used to teach programming. Then, a new lesson is proposed to tune the optimal parameters by quantitative evaluation. The results of the university students practice of the proposed lesson was explained.

Keywords: Junior High School, Parameter Tuning Skill, Quantitative Evaluation

1. Introduction

The Cabinet Office published "Future Investment Strategy 2018" in 2018[1]. In this strategy, it states that in the AI era, in addition to the ability to understand and use AI and data with high mathematical capabilities, the ability to set up and solve problems and to create heterogeneous Human resources who can create value with abilities that are difficult to replace by AI, such as the ability to combine things, are needed. AI education at the primary/secondary levels is being strengthened. The development of teaching materials so that programming education can be effectively implemented in elementary schools and the development of local environments so that students can learn more advanced programming are being carried out. Thus, there is an urgent need to improve the education and instruction of programming. This is not limited to Japan, in recent years, there have been many studies on education and lesson in programming. For example, it has been shown that there are many difficulties in learning programming[2] and that beginners in programming are more likely to learn in-class than in take-home learning[3]. Thus, there is an urgent need to improve human resource development, including programming, both domestically and internationally. In

Japan, the study of programming at the elementary/secondary level school is mainly conducted in junior high school technology and home economics (in the technology field) (hereinafter referred to as Junior High School Technology). Teaching materials for teaching programming in Junior High School Technology include teaching materials for programming the timing of lights and sounds, and robotics[4]. Programming skills are generally assessed by taking "Can you use sequential, iterative, and conditional processing?" and "Are there typing errors?". There are problems with these types of evaluation, such as the fact that they take a lot of time and depend on the discretion of the teacher. To solve this problem, quantitative evaluation studies have been conducted. The "effect" of the program is visualized by having the learner complete tasks of varying difficulty[5]. This method of evaluation allows students to feel their programming skills improve. On the other hand, in the learning process to develop the competency of Junior High School Technology, it is important to learn how to optimize the solution of problems. There is no practice of learning programming that incorporates quantitative evaluation into this learning process. In this paper, A new evaluation method to quantitatively evaluate a sensor car's linear motion in terms of time and left-right swing during the motion are proposed. Then, A lesson for beginners in programming to develop the skills to tune the optimal parameters are proposed.

2. Teaching Materials

In this paper, A sensor car (Made by Artec) as shown in Fig. 1 are used. As a microcontroller board, Studuino mini is used, which is an Arduino-compatible microcontroller board with input/output interfaces for education. An infrared photoreflector (RPR-220, Made by ROHM) is used as a sensor to identify the line to be traveled. It was powered by three AA batteries at 4.5V.

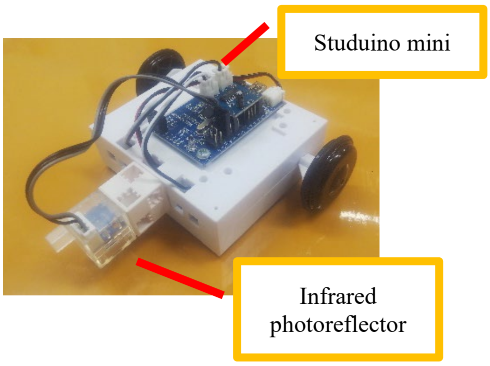


Fig. 1. Sensor car.

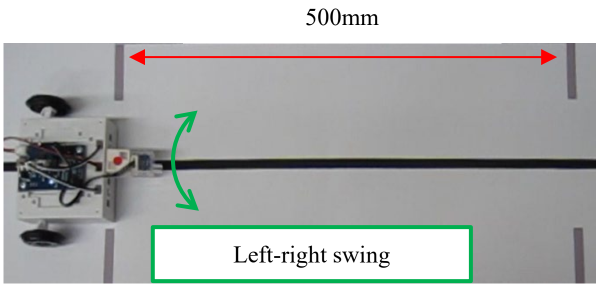


Fig. 2. Field to run.

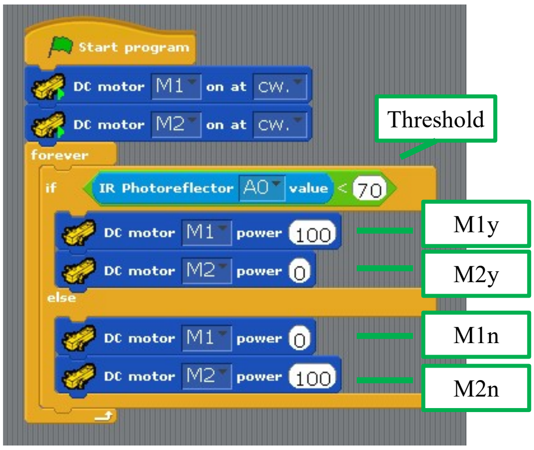


Fig. 3. Studuino mini programming.

3. Learning Task

A learning challenge aimed at the beginning student of programming was devised.

3.1. *Evaluation Items*

As shown in Fig. 2, a black line of 10mm width is drawn on the white surface of the field. The sensor car runs straight 500mm along the black line in this field. The running time and left-right swing are measured. Then, Task 1 is to shorten the running time. Task 2 is to reduce left-right swing. The learning task was to tune the parameters to solve these two tasks.

3.2. *Parameters to be tuning*

As shown in Fig. 3, the Studuino mini software is capable of Scratch-based visual programming. The learner tunes five parameters. It is the threshold of the sensor (threshold), and the ratio of PWM to adjust the speed of the left and right motors (M1y, M2y, M1n, M2n).

4. Parameter Evaluation

The running time t in Task 1 was measured using a stopwatch. The left-right swing of the sensor car in task 2 was recorded by a digital camera (IXY 92015, Made by CANON). Then, it was analyzed with the video analysis software PV Studio 2D ver2 (Made by L.A.B). To measure the left-right swing, markers were attached at two points as shown in Fig.4. The x-axis and y-axis were defined as shown in Fig. 5, and the position of the marker was read. A left-right swing $y12$ was calculated by Eq. (1).

$$y12 = |y1 - y2| \qquad (1)$$

The variance $\sigma^2{}_{y12}$ was calculated in Eq. (2). However, n is the number of data extracted from the video from the start to the goal by the video analysis software. This $\sigma^2{}_{y12}$ was used as the left-right swing.

$$\sigma^2{}_{y12} = \frac{\sum_{i=1}^{n}(y12_i - \overline{y12})^2}{n} \qquad (2)$$

If the running time t and the degree of left-right swing $\sigma^2{}_{y12}$ become small, the learner is able to choose a good parameter. In other words, the parameter tuning skill is high.

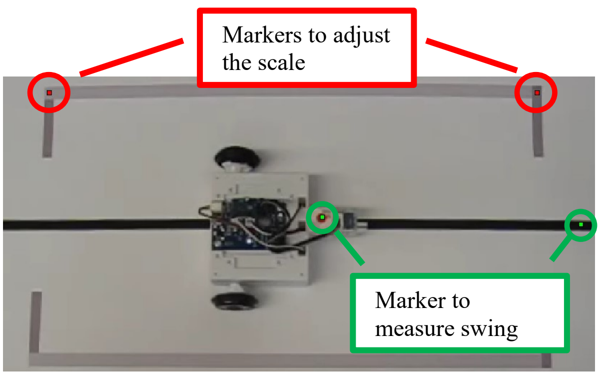


Fig. 4. Marker position.

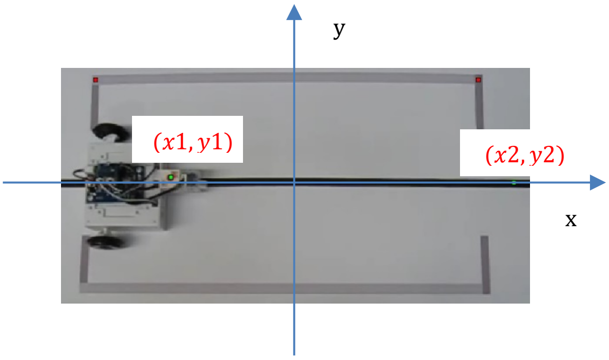


Fig. 5 Set axis.

Table 1. Lesson plan.

	Hours [min]	Main learning contents
Lesson 1	60	Check the relationship between the adjustment of parameters and the movement of the sensor car's wheels.
Lesson 2	30	Know how to program a line trace using a sensor car.
Lesson 3	90	Understand the learning task. Tuning the optimal parameters.

5. Lesson Suggestions

To examine the time required to implement the proposed learning evaluation method and the learning effects, a class was practiced by university students in teacher training. Two people could use one sensor car. eight groups of 16 people were conducted. Table 1 shows the learning plan.

In Lesson 1, students experienced how changing the direction and speed of rotation from left to right changed the behavior of the sensor car. The reason for this was that most of the students had never experienced programming a sensor car before. In Lesson 2, students were shown the program shown in Fig. 3 and were taught the steps of the program to do a line trace. In Lesson 3, the students first predicted the movement of the sensor car and designed the parameters to complete the task. Then, they programmed the sensor car according to the design and made it work. In this way, Lesson 3 incorporated the cycle of design, creation, evaluation, and improvement so that the solution to the problem

Table 2. Sensor car results.

	Group 1							Group 2						
	t[s]	$\sigma^2{}_{y12}$	threshold	M1y	M2y	M1n	M2n	t[s]	$\sigma^2{}_{y12}$	threshold	M1y	M2y	M2n	M1n
First time	4.17	0.085	80	55	100	100	55	4.01	0.32	46	100	20	20	100
Second time	4.12	37.80	60	75	100	100	75	4.05	0.31	45	100	30	30	100
Third time	4.02	0.24	60	80	100	100	60	3.73	0.25	43	100	80	80	100
Fourth time	3.98	0.18	60	90	100	100	75	3.73	0.16	43	100	90	90	100

would be optimal, which is considered important in the learning process.

6. Results and Discussion

Table 2 shows the results of the practice of the university students. In the practice of this study, university students carried out the learning process of Lesson 3 four times. Some of the results of the students' practice are shown in Table 2. For Group 1, the fourth time was the fastest for Task 1. And for Task 2, the second time was the worst. However, it was improved in the third and fourth times. For Group 2, for the third and fourth times is the same for Task 1. However, the left-right swing was improved from 0.25 to 0.16. For this value, the difference in variance was determined by the F test. Assuming that the left-right swing of the sensor car is normally distributed, a one-tailed test with 114,114 degrees of freedom was performed and the F value was 0.0070. Therefore, it can be said that there is a significant difference at 5% level of significance. For the groups shown here, it was concluded from the quantitative evaluation that the parameters were better regulated.

7. Conclusion

In this study, a lesson and an evaluation method were proposed to develop the skills to tune parameters. The proposed lesson and evaluation method were practiced by university students. It was shown that the parameter regulation skills could be evaluated quantitatively. The evaluated data were compared and statistically processed. And the goodness of parameter tuning was demonstrated. In future studies, learning time will be organized and supplementary materials will be developed so that this study can be implemented in junior high schools.

Acknowledgements

The authors are also grateful that this work was supported by the Ehime University Challenge Support Fund.

References

1. Cabinet Office, Future Investment Strategy 2018, 2018. http://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/miraitousi2018_en2.pdf [Accessed December 10, 2020]
2. Y. Bosse and M. A. Gerosa, Why is programming so difficult to learn? Patterns of Difficulties Related to Programming Learning, *ACM SIGSOFT Software Engineering Notes*, **41**(6) 2016, pp.1-6
3. O. Karnalim, Comparing the Impact of Programming Assessmennt Type:In-Class Vs Take-Home, *iJEP*, **10**(4) 2020, pp.125-132
4. Y. Ohnishi, K. Honda, R. Nishioka, S. Mori and K. Kawada, Robotics Programming Learning for Elementary and Junior High School Student, *Journal of Robotics and Mechatronics*, **29**(6) 2017, pp. 992-998
5. Y. Ohnishi, T. Tamai and K. Kawada, Estimation of Programming Learning Achievement by Line Tracing Robot, *Proceedings of the 2019 International Conference on Artificial Life and Robotics,* Beppu, Oita, Japan, Feb. 2019, pp. 431-434

Mini Windmill Generator Kit for Homework for Hiroshima Univ. Monozukuri Junior Doctor Special Educational Program

Hiroyuki Y. Suzuki, Kazuo Kawada, Masayasu Nagamatsu
Graduate School of Humanities and Social Sciences, Department of Educational Science, Hiroshima University, 1-1-1-Kagamiyama, Higashi-Hiroshima City, Hiroshima 739-8524, Japan
E-mail: hiro-suzuki@hiroshima-u.ac.jp, kawada@hiroshima-u.ac.jp, nagamatu@hiroshima-u.ac.jp

Abstract

Hiroshima University Monozukuri Junior Doctor is a special educational program for young (11 – 15 in age) students. It had to be stopped, however, at the midst because of COVID-19 epidemic. We changed it to homework based one. Modified "Mini Windmill Generator Kit" were sent to them with a manual of "Procedure of quantitative research" for self-learning. The result shows that several specific tendencies when the data summarized statistically, although the data collected with different places with different conditions.

Keywords: Technology education, STEM, Quantitative evaluation, Moment of inertia

1. Introduction

Hiroshima University Monozukuri Junior Doctor is a special educational program for young (11-15 in age) volunteered students, aimed for cultivating innovative and outstanding young people as leaders for next generation. Original idea of the program is proposed by Japan Science and Technology Agency (JST).

Our program has started in Jul. 2019, giving 5 seminars in Pre-Stage to around 300 participants (5 times of originally recruited number), and selected to 40 for First-Stage by the end of Sept. The First-Stage also went on smoothly, giving three seminars to the end of 2019. However, following seminars had to be cancelled because of COVID-19 epidemic. We had to change all the following program to self-learning ones. This report presents details of first homework for First-Stage student.

2. Characteristics of Our Program

Our program is composed of three stages of Pre-Stage (recruit of 60 participants), First-Stage (selected to 40) and Second-Stage (selected to 20), in which there are selections at each stage-up timing. This selection policy is based on JST's requirement, that is, only few excellent students will be selected "individually" and continue to study throughout the program. Nonetheless, we introduce distinctive characteristics of "group activity" to the program (approved by JST). We believe that selection of only few "individual" leaders is somehow oversimplified vision to compete with today's science and technology developments. Many of current problems are resolved by, sometimes very large, projects.

We are going to find and cultivate young people of distinctive talents, not only the merely innovative leaders but also capable supporters of innovators, by the program. Proper mixing of diverse talents is a key of success for many kinds of today's projects.

3. Monozukuri Seminar using Mini Windmill Generator Kit

We originally introduced Mini Windmill Generator Kit (Artec co., Japan) in First-Stage's third seminar. The windmill generator has weathercock appearance, that is, it always turns to upwind direction automatically. Torque generated by the wind at the tail fin makes the body turn. Proper size of tail fin and the smoothly moved bearing under the body are the keys for smooth function.

Two themes were given to students (9 groups, 4-5 member each) .

- Finding proper size and shape of tail fin.
- Making smoothly moving bearing using 3D printer.

3.1. *Finding proper size and shape of tail fin*

Plastic sheet of 0.8 mm was used to make tail fins. The sheet was easy to be cut by scissors and set to the rear of the body with a screw (a screw hole was added). Wind was given by electric fans and generated voltages were measured by a tester changing "parameters" of power of electric fan, distance between fan and generator, etc.

We modified the kit to be able to measure the generated voltages by extending thin lead wires from the generator. Additionally, turning torques of the whole body (with fins) were measured using a jig we prepared (Fig. 1).

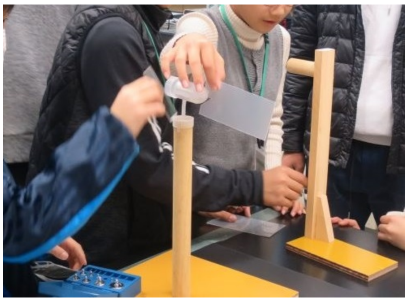


Fig. 1 Jig for measuring turning torque of the windmill generator body.

Generated torque by wind is in proportion to the size of the fin. At the same time, the required torque to turn the whole body is also in proportion to the moment of inertia. That is, the larger the fin is, the larger the torque is required to turn the body. They are trade-off relationship each other. The optimum shape and size of the tail fin will be found at the balanced point of those two factors.

Students were expected to find appropriate sizes and shapes of fins by themselves. We did not give, however, detailed lecture on above mentioned mechanism beforehand. We rather expected that students grabbed overall picture of the mechanisms by themselves through quantitative data collection and arrangement of the data using tables and graphs.

3.2. *Making smoothly moving bearing using 3D printer*

Students were experienced 3D printing as an introduction to digital fabrication. They firstly studied basic commands of 3D CAD (Design spark Mechanical) and "extrude" concentric shapes with a hole, namely bearing, in monitor. Then, they made the bearings data to reality, using 3D printer (XYZ printing PRO).

Majority of the students were successful to draw bearing design but only few were printed, because of limitation of printing speed. Therefore, theme 3.1 was performed using original bearings in the kit.

3.3. *Result of seminar*

Each team could take data of tail fins performance by changing distance between electric fan and the generator and power needed to turn the fins. Group work made easy to perform experiments in shorter time.

They were eager to propose new design of tail fins, to collect data by experiments, to design bearings in 3D CAD by themselves. Nonetheless, sufficient data were not collected for the discussion in the seminar, because of the time limitation (a half-day seminar). We were going to continue the same theme, changing approach rout, in following seminar.

4. Homework

4.1. *Preparation of homework kit*

All meetings were banned by Univ. official, after rising COVID-19 epidemic. In Mar. 2020, we sent a homework kit of mini windmill generator to every student, as supplemental contents of previous seminar.

Three color LEDs are added (with a series resistor) as voltage indicator. A tester is attached to the kit. Manual of "Procedure of quantitative research" in which we roughly explained how to figure the graphs and encouraged them to "find a peak value" in figures. Note that we did not designate any parameters to be changed. All the experimental conditions were decided by oneself.

In homework, we gave two themes of,

- finding best design of tail fin (same as seminar)
- finding best design of windmill.

The windmills could easily made by cutting pet PET-bottles. The kit contained a coupler to unite PET-bottle windmill to generator axle.

4.2. *Reports returned from students*

An example of report returned from a student (female, first grade in junior high school) is shown in Fig. 2.

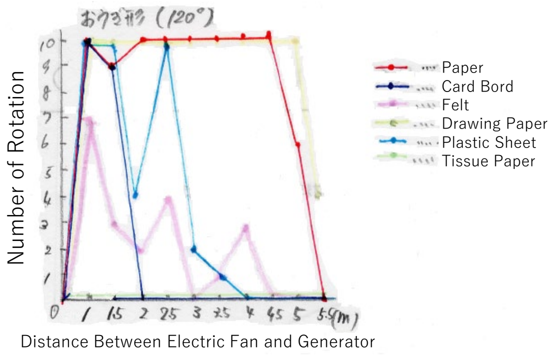


Fig. 2 Research for finding best material for tail fin.

She tried to evaluate the effect of materials on the performance of tail fin. She chose fan shape fins (radius of 100 mm and angle of 120 deg.) made of different materials of, i) copy paper, ii) card-bord, iii) felt, iv) drawing paper, v) plastic sheet and vi) tissue paper.

Number of rotation (reacted to the wind) were counted changing distance between electric fan and the generator. She concluded that the drawing paper showed the best performance. She also checked size of the fins in other part of the report and concluded about 100cm^2 was the optimum size.

Figure 3 is an example of report on windmill experiment from another student (female, first grade in junior high school), showing various windmills. She checked windmill performance from two viewpoints of, i) number of blades (upper table and figure) and ii) length of blade (lower table and figure) and summarized that the optimum number of blades was three and size of blade was 3cm, respectively.

Students made plan of experiment by themselves since we did not give step-by-step procedure in the instruction manual but gave basic idea of appropriated data collection, so that somehow unique trials were performed by many students.

5. Discussions

Windmill generators are one of popular teaching materials for energy conversion since it is easy to make windmills by hands. However, there are hurdles when we try to understand "the essence" of the product, since it

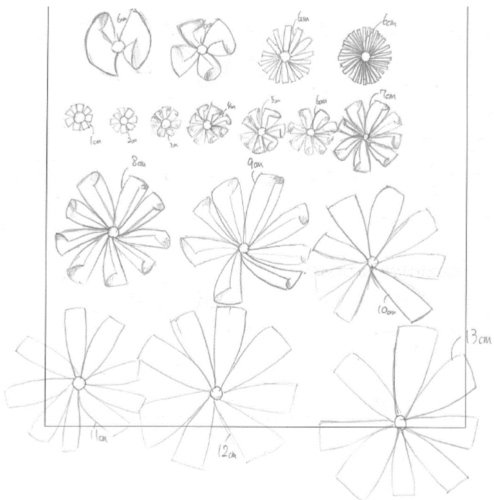


Fig. 3 Performance of many designs of PET-bottle windmills were measured by a student.

contains "complexity" as "real product". It contains all STEM factors. Therefore, some research gave up quantitative evaluations by students[1)] or evaluated only by instructors[2,3)]. Some research came into quantitative evaluations, but they still scarcely used graphs[4,5)].

Our program is forced to introduce, by chance, experiments planned and performed by students "individually". We introduced quantitative evaluation using graphs, but experimental parameters and conditions were all different in each student. Maximum voltages attained were varied from 0.3 V to 6 V, for example, among their experiments. Nonetheless, there

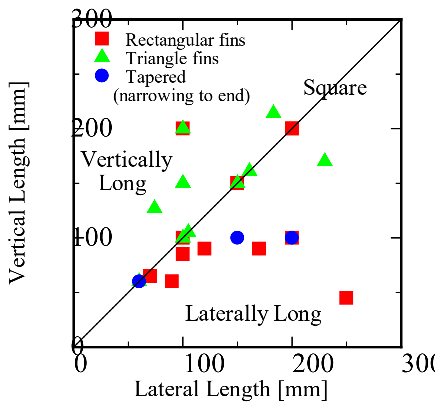


Fig. 4 Optimum sizes and shapes of tail fins found by each student experiment.

are several noteworthy tendencies can be found when we summarize their data statistically.

Figure 4 shows the lateral and vertical sizes of tail fins of the best performance. Although argument for selecting the best fin were merely scattered, there is tendency between shape and aspect ratio of fins. Rectangular and tapered (narrowing to the end) fins are relatively laterally long. In contrast, triangle fins (widen to the end, including fan shape) have vertically long. Higher moment of inertia of triangular ones must be one reason of such difference. Interestingly, on the other hand, sizes of fins are almost indifferent to their shapes.

Materials of fins also have no considerable effect to sizes and shapes of them, as shown in Fig. 5, although the result in Fig. 2 by a student indicates clear difference. Much higher weight of generator body may mask the weight differences of fin materials.

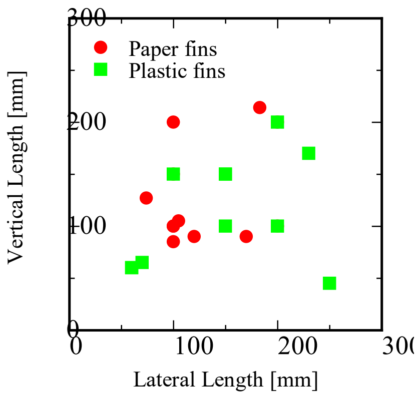


Fig. 5 Effect of materials on sizes and shapes of tail fins.

As for the windmill design, optimum blade sizes are summarized in Fig. 6, showing an optimum of around 40 mm. In contrast, blade number (from 4 to 16) is relatively indifferent among students results (graph is omitted).

These graphs indicate that, the data collection of scattered experimental conditions was also effective to find out optimum conditions of the products, if the member took data quantitively and summarized the data statistically. It is possible that this style of approach is rather matched to improve working conditions of many products since many "real" products essentially contain complexity in themselves.

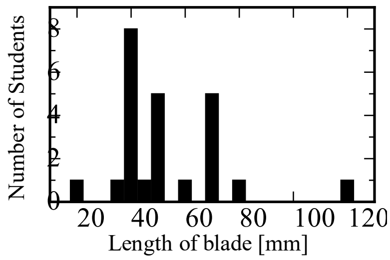


Fig. 6 Optimum blade length of windmills.

6. Summary

Optimum sizes and shapes of tail fins and windmills for mini generator kit were determined by quantitative data collection by students' (11-15 in age) homework. The result shows that several specific tendencies when the data summarized statistically, although the data collected with different places with different conditions.

Acknowledgements

This program was fully supported by JST fund of Junior Doctor Educational program.

References

1. R. Khunyakari, S. Mehrotra, S. Chunawala, C. Natarajan, Design and technology productions among middle school students: an Indian experience, *Int. J. Technol. Des. Educ.* **17**, 2007, pp. 5–22.
2. Y. Kawamura, K. Honda, O. Izutsu, Y. Matsumoto, Development of educational materials savonius type windmill aerogenerator using handmade generator, *J. Jpn. Inst. Energy* **92**(9), 2013, pp. 913-918.
3. R. Hikiji and Y. Fujimoto, Study on improving experiment-teaching materials of wind power generator for junior high school students to get interested in natural energy, *J. JSEE* **65**(5), 2017, pp. 48-53.
4. J. Concannon and P. L. Brown, Windmills by design: Purposful curriculum design to meet next generation science standards in a 9-12 physics classroom, *Science Activities* **54**(1), 2017, pp. 1-7.
5. Y. Ohnishi, S. Mori, T. Deyama, J. Ono, H. Muko, Development of teaching materials featuring wind power generator for 6th elementary science class based on the relationship with technology class of junior high school, *Ehime Univ. Edu. Faculty Kenkyu Kiyo* **61**, 2014, pp. 105-110.

Development of Teaching Materials to Learn the Efficient Use of Energy

Yoshihiro Ohnishi*
Faculty of Education, Ehime University, 3, Bunkyocho
Matsuyama, Ehime 790-8577,JAPAN
E-mail: ohnishi@ehime-u.ac.jp
www. ehime-u.ac.jp

Teruyuki Tamai*
E-mail: tamai.teruyuki.xq@ehime-u.ac.jp

Shinnosuke Mori*
E-mail: mori.shinnosuke.my@ehime-u.ac.jp

Kawada Kazuo**
Faculty of Education, Hiroshima University, 1-1-1, Kagamiyama
Higashihiroshima, Hiroshima 739-8524,JAPAN
E-mail: kawada@hiroshima-u.ac.jp

Abstract

Control technology using the computer is one of the key technologies for improving energy efficiency. PWM control, which is a control technology for DC motors, is widely used as the basis of power electronics with the development of semiconductor devices. On the other hand, programming techniques are required to effectively use these techniques. In this paper, we propose teaching materials that can visualize the performance improvement by programming as the operating time performance index of the wind turbine.

Keywords: teaching materials, electric double layer capacitor, power electronics, control engineering

1. Introduction

Computer-based control technology is one of the key technologies for improving energy efficiency. PWM control, which is a control technology for DC motors, is widely used as the basis of power electronics with the development of semiconductor devices. On the other hand, programming techniques are required to make effective use of these techniques. Control performance assessment[1],[2] which is a measure of how effectively control technology is working. This method originated in Harris's research[3] and is widely used in the field of process control. The authors of this paper are conducting research to apply this control performance assessment method to the achievement evaluation of programming learning[4],[5].

Since the original control performance assessment is a method based on the variance of control error, it is difficult for elementary and junior high school students to understand its meaning. Therefore, presenting more easy-to-understand index to students as a scale may lead to independent learning.

In this paper, we propose to use the operating time of the wind turbine as a performance index. Using an electric double layer capacitor (EDLC), which is used as a new storage device for electrical energy, as a power source, we provide activities to think about how to use that energy efficiently. Teaching materials that can visualize performance improvement by programming are explained.

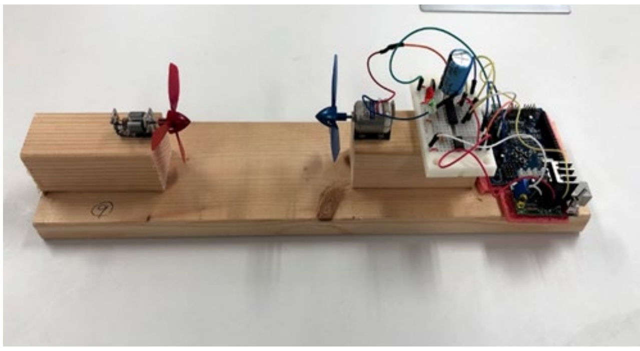


Fig. 1. The Overview of the developed equipment

Fig. 2. The Drive wind turbine which is driven by FA-130RA.

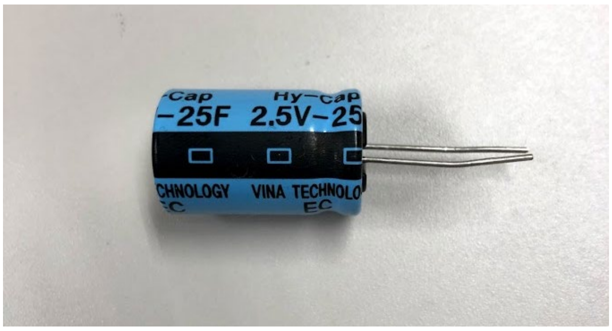


Fig. 3. Power supply Electric Double Layer Capacitor with withstand voltage 2.5V and capacitance 25F.

2. The Overview of the developed equipment

The overview of the teaching device is shown in Fig. 1. The red windmill is rotated by a blue fan. The blue fan is driven by the Mabuchi DC motor FA-130RA shown in Fig. 2. The power source for the blue fan is a charged EDLC. The EDLC used as a power source is shown in Fig. 3, with a withstand voltage of 2.5V and the capacitance of 25F. 2.5V constant voltage power supply was used to charge the capacitance. This is a kit using the variable regulator NJM2397, which is shown in Fig. 4 and can be purchased at Akizuki Denshi. If the regulator for charging is turned on, the motor will rotate with the regulator as the power source instead of the EDLC. Therefore, exclusive control as shown in Table 1 is required. A circuit using NAND was constructed to achieve this exclusive control. The blue fan can be driven by PWM control. The PMW control frequency is 10Hz and is programmed so that the sum of on-time and off-time is 0.1 seconds. PWM control to the motor is realized by MOSFET 2SK4017. The circuit diagram is shown in Fig. 6. A0 in Fig. 5 is the output of the ON / OFF signal from the computer, and A4 is the AD input for monitoring the voltage of the EDLC. The voltage of the A4 channel can be monitored with 3.3V as the maximum value of 100. This relationship is shown as Fig.6.

Fig. 4. Variable regulator.

Table 1. Output signal for FET.

A0	5V	Out
0	0	0
0	1	0
1	0	1
1	1	0

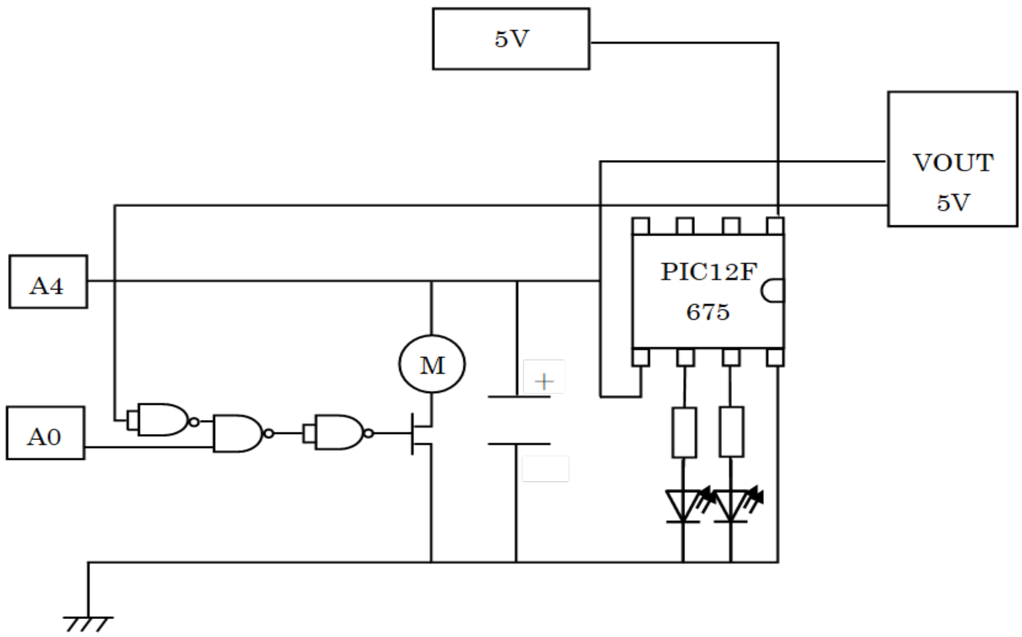


Fig. 5. Overall circuit diagram.

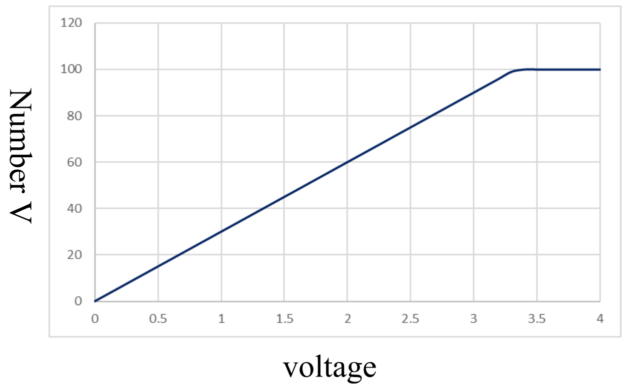


Fig. 6. The relationship between the monitoring voltage and digital number which is loaded by AD input.

3. Experimental evaluation

The purpose of the device experiment described in the previous section is to turn the red windmill as long as possible with a blue fan. At this time, the number of rotations of the red windmill does not matter, as long as it is moving. By adjusting the duty ratio of PWM control by programming, the speed of the blue fan can be changed. The control purpose is to give the minimum energy for the red windmill to move by the blue fan. Table 2 shows the results of the experiment under different conditions. First, when the duty ratio was 100%, the time was the shortest in this experiment. Next, when the duty ratio is 50% at all times, the time is extended a little more. However, not all the energy in the EDLC can be used. Therefore, if the program is switched to a duty ratio of 50% at first and then to a duty ratio of 100% when the voltage drops, the time will be further extended. The longest time could be obtained by increasing this condition. From these results, it can be seen that the time of the result changes by changing the programming. At this time, the performance index by programming is the result time. Students try to improve programming in order to increase the time of results. Teachers can look at time and evaluate student skills.

4. Conclusions

In this study, we created teaching materials aimed at efficiently using the energy stored in EDLC. In order to use energy efficiently, PWM control technology is utilized, and students perform activities to increase the target performance index by programming. In actual products, energy is used efficiently by control technology such as refrigerators, air conditioners, and even hybrid vehicles. This teaching material aims to learn that. In the future, students will be able to carry out activities and their effectiveness will be verified.

Table 2. Experimental result.

Condition(Duty ratio)	Measurement time
D=100%	3m 09s 69
D=50%	3m 31s 99
If V>30, then D=50% else D=100%	3m 56s 83
If V>25, then D=50% else D=100%	4m 17s 48
If V>65, then D=20% else if V>51 then D= 30% else if V>37 then D=50% else if V>27 then D=80% else D=100%	5m49s 51

Acknowledgements

The development of the experimental system was carried out by Mr. Tomoki Yokohama, Mr. Yuji Mukaigawa and Mr. Motoki Nanjyo former students of Ehime University. And the practical evaluation was done with kindly support by Mr. Kimitoshi Honda Junior high school teacher in Ehime prefecture. This work was supported by JSPS KAKENHI Grant Number JP18K02980.

References

1. B. Huang and S. L. Shah: Performance Assessment of Control Loops: Theory and Applications, Springer, London (1999)
2. M. Jelali: An overview of control performance assessment technology and industrial applications, Control Engineering Practice, Vol.14, pp. 441-466 (2006)
3. T. J. Harris: Assessment of Closed Loop Performance, Canadian Journal of Chemical Engineering, Vol.67, pp. 856-861 (1989)
4. Y.Ohnishi: Educational Effect Measurement of Control Technology based on Performance Index, Proceedings of the SICE Annual Conference 2019, pp. 1323-1327(2019)
5. Y.Ohnishi, T. Nakano, T. Tamai, S. Mori and K. Kawada : Practical Learning of Science for Elementary School Students via Programming and Control Experimentation, Journal of Robotics, Networking and Artificial Life, in press (2021)

Networking Integration Application of an Intelligent Production Line for Aerospace Precision Manufacturing

Chau-Chung Song[a,b*], Chun-Chi Wang[c], Geng-Yi Lin[a] and Chung-Wen Hung[d]

[a] *Department of Aeronautical Engineering,*
National Formosa University, Huwei, Yulin, Taiwan
[b] *Smart Machine and Intelligent Manufacturing Research Center,*
Nation Formosa University, Huwei, Yunlin, Taiwan
[c] *Graduate Institute of Aeronautical and Electronic Technology,*
National Formosa University, Taiwan, R.O.C.
[d] *Department of Electrical Engineering,*
National Yunlin University of Science &Technology, Taiwan, R.O.C.
Email: ccsong@nfu.edu.tw[*]

Abstract

In this paper, the networking integration application of an intelligent production line is studied and developed for aerospace precision manufacturing. The technical development of the performance equalization and control for EDM machine is also introduced with the related technologies of the machine network and EtherCAT network. In addition, the development of intelligent production lines is integrated to improve the system stability and production efficiency of production line systems for aerospace component manufacturing. The key technologies of Cyber-Physical System (CPS) and intelligent networking is also integrated to constructs the real-time intelligent system monitoring system (SCADA) for data record and management. The key technologies of the drilling process are developed for intelligent diagnosis and drilling holes inspection of related production processes. The integrated robot arm is responsible for the loading and unloading of the material, and the entire production line is connected by the EtherCAT network. The automatic optical inspection (AOI) is introduced at the rear end of the production line to realize the quality judgment of the finished product. Furthermore, the virtual cloud network and the intelligent factory are achieved with the information collection and intelligent monitoring system.

Keywords: Aerospace Precision Manufacturing, EtherCAT, SCADA, Intelligent Production Line

1. Introduction

In recent years, as the tendency of the popularity in the global aerospace industry, the number of airline passengers has grown year by year. Therefore, there is a strong demand for new civil aircraft development. However, the machine object equipment mostly relied on aerospace manufacturing overseas. Because of the weak confidence level, the domestic aerospace manufacturing industry goes into a declining phase. Therefore, in 2014, the Executive Yuan hopes to introduce the manufacturing line to carry out the aerospace components and then promote the domestic tool machine industry to enter the aerospace processing field. The final goal is to promote domestic machine tools to enter the international aerospace supply chain. In this paper, the intelligent networking system is integrated and developed for production line applications based on the EtherCAT network, including control of the robotic arm.

2. Methods

2.1 *EtherCAT*

EtherCAT (Ethernet for Control Automation Technology) is an Ethernet-based Fieldbus system, invented by Beckhoff Automation. This technology breaks through the system limitations of other Ethernet solutions in the past and has the characteristics of high availability requirements, flexible topology, high-performance, Distributed Clocks for High-Precision Synchronization, and low-cost. The EtherCAT master uses a standard Ethernet port and network configuration information stored in the EtherCAT Network Information file (ENI). The ENI is created based on EtherCAT Slave Information files (ESI) which are provided by the vendors for each device. Slaves are connected via Ethernet, any topology type is possible for EtherCAT networks. EtherCAT architecture is shown in Figure 1.

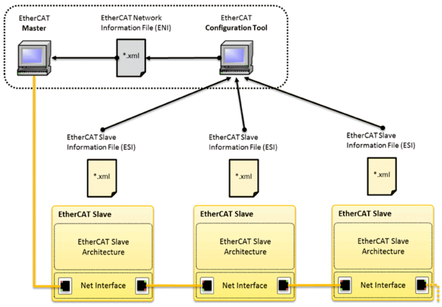


Fig. 1. EtherCAT Network Architecture

2.2 *CANopen over EtherCAT*

Using the CoE application layer protocol, EtherCAT communication has the same standard protocol as CANopen, Including the same Object Dictionary (OD), Service Data Object (SDO) settings, Process Data Object (PDO) data transmission, and the objects refer to CAN in Automation (CiA) 402 for motion control. These objects can define the state control of the motor, basic motor parameter settings, and data feedback objects, etc., providing objects that can be used for basic motor control operations.

3. System Design and Architecture

For the development of an intelligent production line application, the EDM (Electric Discharge Machine) is the core machine and integrates the robotic arm by the EtherCAT network. The SCADA system is to connect the EDM, the robotic arm, and the automated optical inspection system (AOI system) through the communication network to establish an aerospace precision manufacturing, as shown in Figure 2.

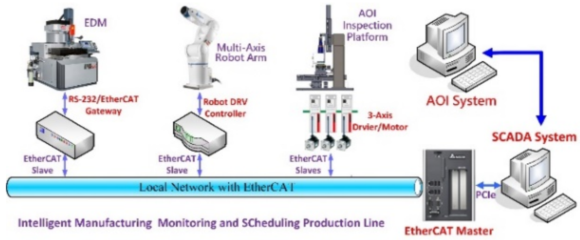


Fig. 2. Aerospace precision manufacturing

4. System Integration and Testing

The experimental method of Intelligent Production Line Application is to run the overall electric discharge machining production line after the EtherCAT local area network and the SCADA system have been constructed and connected to monitor. Review the results of drilling at special angles of the workpiece to verify the effectiveness of the EDM production line automation. To add the Six-Axis articulated robot to the overall machining production line network for control, it will bypass its upper controller of the initial factory and connects to the six motion control cards, which are regarded as nodes. To control the Six-Axis articulated robot, use the kinematic calculation method above to calculate the corresponding value of pulse to the rotation angle of each axis required to reach the target position. Then use the CiA402 moving object of the motion control cards itself for EtherCAT transmission. Then use the value of pulse which is obtained from the feedback object to infer the current angle position information. There are several modes for motion control of the robotic arm, here use interpolated position mode. Also, the control of the electric gripper matched with the robotic arm uses its USB port to communicate and control with the computer via UART. Before the SCADA system starts to connect and control, it is necessary to use TwinCAT to create related corresponding objects and settings through the description of each slave device XML file to construct an EtherCAT network. After the completion of network construction, set the initial parameters of SDO, and set through the objects in the object dictionary, then set the PDO Output object that will be added during subsequent operations. Real-Time numerical control can execute after it into the operational state. Then, SCADA software can be used to connect to the network for reading object information and issue control commands. The SCADA system is to monitors the EtherCAT network status of the entire production line, it is mainly connected to the master station TwinCAT via API. SCADA system can be used to instantly grasp the actions and status of each machine and equipment, and establish a user interface for users to monitor and manage the production line status and to provide the inspection results of the finished workpiece. The SCADA system is shown in Figure 3.

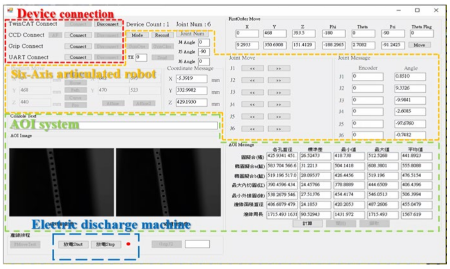


Fig. 3. The SCADA system

According to the function, C# based SCADA system can be divided into four parts, device connection, Six-Axis articulated robot, AOI system, and electric discharge machine. The functions of each part will be introduced below.

4.1 Device connection

The main function of the device connection is to connect with the EtherCAT Master TwinCAT on the production line and communicate and control transmission with each node through this network. In addition to the EtherCAT network connection, this SCADA system also integrates other hardware connections, such as the USB connection of the AOI automatic positioning system, and the connection control of the electric gripper of the robotic arm. On the network connection, the customized RJ45 interface 100BASE-TX is used as the network physical layer, which is connected to the main station by serial connection with each station, and a personal computer is used as the main station. Other devices are connected to a personal computer using USB Port and then controlled by the SCADA system. The interface of TwinCAT master is shown in Figure 4.

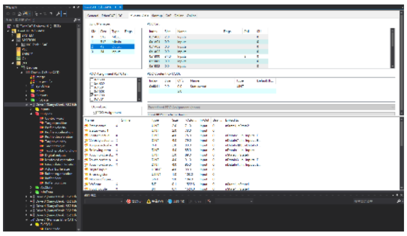


Fig. 4. The interface of TwinCAT master

4.2 Six-Axis articulated robot

After connecting its motion control cards via the network and completing the network communication, the robotic arm can be controlled. It operates on the Cartesian coordinate system and coordinated with single-axis angle numerical control of the rear three axes to adjust the detection angle, and calculated by feedback object to know the current position information and feedback information of each axis. Also, a single-axis control is added for the user to find the grasping point to ease the scheduling design.

4.3 AOI system

This AOI system consists of two parts, the automatic positioning system, and the AOI optical inspection. Connect the MCU and AOI system to obtain the current positioning information and display the result of the finished workpiece inspection on the interface. The analysis information is shown in Figure 5.

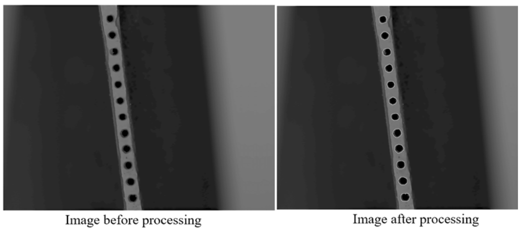


Fig. 5. Images inspected by AOI system

4.4 Electric discharge machine

Given the issues related to industrial safety, the EDM manufacturer (Ching Hung Machinery & Electric Industrial Co., Ltd.) does not provide control methods related to EDM computer host and external connection, like an independent working machine. However, to achieve the goal of automation of the production line, an EtherCAT network node must be constructed separately. Through the external relay circuit with the EtherCAT communication control board that transmits the signal to the internal relay of the EDM, then the existing programs in the EDM computer will be activated, the program including executing, signals of fixed-point feedback, and emergency stop. The control interface is as shown in Figure 5, the red light is matched with the feedback signal. While the electric discharge machine is stably ready, the green light will be on, and the processing program can be started with Start at this time. The workpiece being drilled by EDM is as shown in Figure 6.

Fig. 6. The workpiece being drilled by EDM

As aerospace consumables workpieces are not available, it can only be machining through similar workpieces, such as galvanized iron blocks, stainless steel plates, aluminum plates, and radiator stator scrap for machining and measurement. During the trial run of the overall electric discharge machining production line, the network controller issues work instructions, the robotic arm carries out the reclaiming operation, and the electric discharge machine is processed. After the machining is completed, the robotic arm sends it to the AOI system for quality inspection. The entire automated production line operates smoothly. The workpieces after machining are shown in Figure 7.

Fig. 7. Workpieces after machining

5. Conclusion

In this paper, the networking system integration and development is focused on EtherCAT protocol for an intelligent production line application, Use the EtherCAT network to integrate the Six-Axis articulated robot and EDM, and design a SCADA system to connect the EDM, Six-Axis articulated robot, and AOI system via the communication network to build an automated network control station. During the trial run of the overall electric discharge machining production line, the network controller issues work instructions, the robotic arm carries out the reclaiming operation, and the electric discharge machine is processed. After the machining is completed, the robotic arm sends it to the AOI system for quality inspection. The entire automated production line operates smoothly. On the network control station, the SCADA system can be used to grasp the actions and status of each machine tool and equipment of the production line in real-time, and establish a user interface for users to monitor and manage the production line status, and to provide the inspection results of the finished workpiece. This Networking System Integration and Development of an Intelligent Production Line Application has been completed, which can implement intelligent communication between machine tools and equipment, and achieve the goal of intelligent control of machining equipment, improve machining quality and productivity.

6. References

1. ETG. *EtherCAT The Ethernet Fieldbus*. (2009).
2. H. Y. Tang, “Key Technologies of Realizing Advanced Motion Control Design on CNC Machines through the EtherCAT,” M.S. theses, National Chiao Tung University, Taiwan, R.O.C., 2015.
3. Y. L. DENG, “Design and Implementation of with EtherCAT and Low Power Consumption Sensing System applied in Smart machinery,” M.S. theses, National Formosa University, Taiwan, R.O.C., 2019.
4. ETG. *2200 Slave Implementation Guide V3.1.0*. (2018).
5. ETG. *1000 EtherCAT Specification Part 6: Application layer service definition*. (2017).
6. J. J. Craig, “Manipulator kinematics,” in Introduction to Robotics: Mechanics and Control, 3rd ed., London: Pearson Education Ltd, pp. 66-76, 2017.
7. J. G. JIANG, “The Development of Chopstick Tube Feeding Detection System based on the Visual Angle of Robot Arm,” M.S. theses, National Yunlin University of Science and Technology. Taiwan, R.O.C., 2019.

Acknowledgements

This research was supported by the Ministry of Science and Technology, Taiwan, R.O.C. under Grant: MOST 107-2218-E-150-007 and 108-2218-E-150-004.

Low-Cost Indoor Localization Using Sound Spectrum of Light Fingerprints

Chung-Wen Hung
Department of Electrical Engineering, National Yunlin University of Science and Technology. Douliou, Taiwan

Hiroyuki Kobayashi
Osaka Institute of Technology, Osaka, Japan

Jun-Rong Wu
Department of Electrical Engineering. National Yunlin University of Science and Technology. Douliou, Taiwan

Chau-Chung Song
Department of Electrical Engineering, Department of Aeronautical Engineering, Taiwan

E-mail: wenhung@yuntech.edu.tw, hiroyuki.kobayashi@oit.ac.jp, m10812085@yuntech.edu.tw, ccsong@gs.nfu.edu.tw

Abstract

A low-cost indoor localization system using sound spectrum of light fingerprint is proposed in this paper, an artificial intelligence, AI, algorithm will be implemented in a low-cost micro-control unit, MCU, to perform the localization. The unit, complex and tiny differences of the light fingerprints are caused by the different characteristics of the discrete components used in lighting devices. To reduce the memory size for low-cost MCU, only sound spectrum of light fingerprint is used to identification the lighting device. Moreover, the grid search also is used to optimize the hyperparameters to compact the AI model. The system architecture and algorithm development are discussed in this paper, and the experimental results will be present to show the proposed system workable

Keywords: light fingerprint, machine learning, indoor, localization

1. Introduction

Indoor localization is an issue in may application, such as automated guided vehicle, smart building. However, the traditional Positioning method, Global Positioning System (GPS), is not workable indoors. Many indoor localization literatures based localization caught eyes due to infrastructure-less character. However, how to modularize the localization method is other key point for using the technology in indoor equipment.

A mobile phone was used to sense the light and convert it to the received light strength (RLS). This RLS would be compared with other RLSs on the map, then the position was located by the similar RLS [1]. Four LEDs placed on the ceiling were used to transmit the encoded and modulated optical signals with position coordinates, and the signal received from photodiode receiver was analyzed to calculated the position [2]. In [3], light signals were acquired at high frequency, up to 1 mega-samples/second, then mapped into the frequency domain by Fast Fourier Transform (FFT). The k-Nearest Neighbor (KNN) and Convolutional Neural Net (CNN) classifiers were discussed. The hardware system comprised of a high-frequency light sensor, analog-to-digital converter (ADC) scope, and Raspberry-PI processor to the performance of the classifiers. [4] used the microphone build in the mobile phone to sampling the light signal sensed by a photodiode (PD), then the sampling were sent to the cloud server through the network. The FFT preprocessing and 1-dimensional convolutional neural network CNN were calculation on the server, and classification is performed. [1] and [2] used other characteristics for localization, and never mention about realization. [3] achieved the modularized target, and the proposed system included an extra high sampling rate ADC. The method proposed in [4] only needed the acoustic frequency spectrum of the light, it means lower sampling frequency requirement. Then, all the proposed localization algorithm was performed in cloud sever. The proposed method in this paper is similar to [4] with extension and modification, and how to realize the method into a module will be detailed.

2. Data Acquisition

The text is to be typeset in 10 pt Times Roman, single spaced with interline spacing of 13 pt. Text area (excluding running title) is 6.75 inches across and 8.8 inches deep. Final pagination and insertion of running titles will be done by the publisher, so make sure that no page numbers are given in your paper and only the running titles provided in this template (authors' names and paper title) are used.

A Data Acquisition and feature spectrum extraction

The analog to digital converter, ADC, build in MCU is used to sample the time domain data. First, higher sampling rate and the larger number of sampling points is selected to confirm the feature frequency, first. Then, the suitable sampling rate and point will be optimized, means to minimize these two number but keep the feature. The preprocessor is discrete Fourier transform, DFT, to present the frequency feature of the sampling data. The formula is show in (1), $X[.]$ and x[.] represent respectively the frequency and time domain data; k and n denotes the k-th point frequency and n-th point sampling data, and N is the totally sampling points. The frequency resolution, Δf, of FFT is as (2), here, f_s indicates the sampling rate (frequency). The typical FFT results of two fluorescent lamps are shown in Fig. 1 and 2, the 16,384 points time domain data sampled at 65536Hz.

$$X[k] = \sum_{n=0}^{N-1}\left(x[n] \times e^{-i\frac{2\pi nk}{N}}\right), \qquad k = 0,1,\dots,N-1 \tag{1}$$

$$\Delta f = \frac{f_s}{N} \tag{2}$$

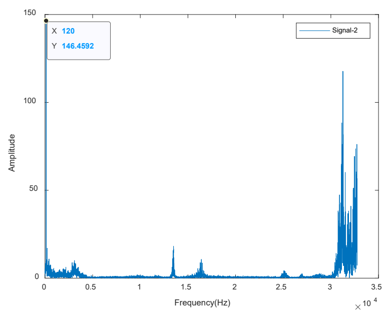


Fig. 1 The spectrum of the fluorescent tubes 1

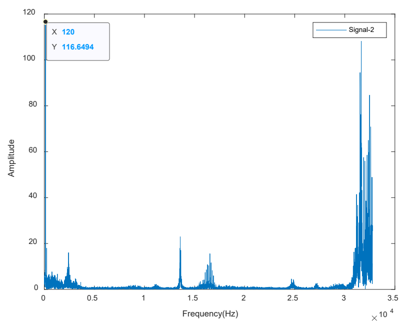


Fig. 2 The spectrum of the fluorescent lamp 2

The spectrums of these two fluorescent lamps are different, and the most features of the spectrum locate at low frequency interval. Note, here is a huge spike at 120Hz no matter lamp 1 or 2, because fluorescent lights is powered by magnetic ballasts with a frequency of 60Hz (50Hz in some country) and lamps flicker with the double frequency. The portion less than 150 Hz will not be used in localization. In the other side, there are also some peaks around 32K Hz. The portion near Nyquist frequency is also ignored. Only the spectrum from 150 Hz to 30K Hz is used to machine learning and localization. The Fig. 3 and 4 are the spectrums of the two lamps in selected frequency range

B Rescaling

For modularization, the artificial intelligence algorithm will be implemented into an MCU, which calculation and memory capabilities are limited. Smaller node number of input layer means less operation and memory requirement. The reduction of the input nodes but without losing frequency feature is important. As shown in Fig. 3 and 4, the low frequency spectrum contains most information, so frequency rescaling equation will be adopted to cover original FFT data to the characteristic vector[4]. The frequency axis will be rescaled to power of F_{factor} as shown in (3), and the F_{factor} is shown as (4). Here, f_{max} and f_{min} indicate the frequency range which is used in machine learning, and N_{F_target} is the target number of reduction. Finally, the amplitude of rescaling spectrum, $A[i]$, is the summation of the amplitude which frequency between two rescaling frequency axis point. (5).

$$F(i) = \left(F_{factor}\right)^i \tag{3}$$

$$F_{factor} = \left(\frac{f_{max} - f_{min}}{f_{min}}\right)^{\frac{1}{N_{F\_target}}} \tag{4}$$

$$A[i] = \sum_{j}^{F_{factor}{}^{i} \le j*\Delta f \le F_{factor}{}^{i+1}} |X[j]| \tag{5}$$

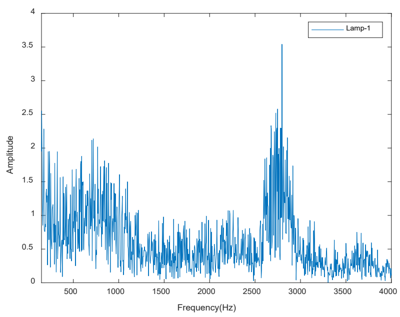


Fig. 3 The spectrum of 150 to 4000 Hz for the lamp 1

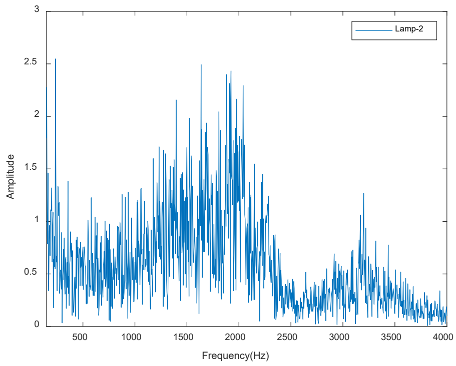


Fig. 4 The spectrum of 150 to 30k Hz for the lamp 2

The Fig 5. presents the rescaling results of FFT for the two lamp. Here, f_{max} and f_{min} are set to 150 and 30k Hz, and rescaling points are 400. The rescaling results keep the original features in low frequency spectrum, and high frequency's is summed into low resolution form. However, the input data are 75 times smaller, this rescaling processing will reduce the requirement of memory and calculation dramatically. In other words, it is very helpful for modularization in a low-cost MCU.

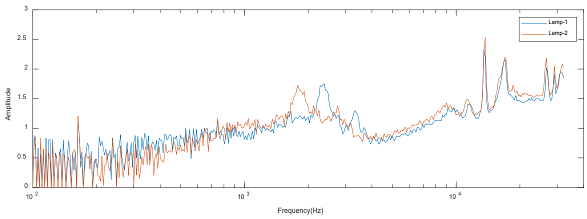


Fig. 5 The rescaling result for the lamp 1 and 2

3. Identification Algorithm and Grid Search for Hyperparameters

CNN is a popular algorithm for classification due to the its ability to learn to extract the features, and less parameter memory and calculation requirement is suitable for modularization [5]-[6]. 1-D CNN is adopted in this paper due to the resource limitation in low-cost MCU, no matter memory or calculation speed. The filter kernel in convolution layer is used to extract certain features from rescaling spectrum, and full connected layers perform the classification, but pooling layer is skipped in this application.

Grid search is standard skill to get the hyperparameter combination of the highest accuracy [7], such as CNN Filter Kernel number and size. Different from general grid search which only focus on accuracy, the memory size of parameter is also a key factor for modularization. The grid search items and the range is expressed in table I, here are totally 240 parameter combination.

Table I. The items and parameter of grid search

Item	*Parameter*
CNN Filter Kernel number	1、2
CNN Filter Size	2、4、8、16、32、64、128、256
Activation Function of Fully Connected layer	ReLU、Tanh、Sigmoid
Node of Fully Connected layer	64、32、16、8、4

Localization is designed for eight lamps position, and 100 averaged rescaling data is collected for each lamp for training. To eliminate random errors caused by initial random weights, the three training results for each parameter combination are averaged. The iteration number set to 1000, and early stopping is adopted to avoid overfitting per training round. The grid search results are sorted by verified accuracy and parameter file size. In the 240 training results for all parameter combination, the top ten results are shown in Table 2.

Table 2. The Top Ten Results Gid Search

Train. Acc.	*Train. Loss*	*Verified Acc.*	*Filter Kernel*	*Filter Size*	*Act. Func.*	*Nodes of Fully Connected Layer*	*Memory Size (kB)*
1	0.000577	1	1	256	ReLU	32/16/16	5.96875
1	0.010547	1	1	256	Tanh	32/16/16	5.96875
1	0.051400	1	1	256	Sigmoid	32/16/16	5.96875
1	0.010734	1	1	128	Tanh	32/16/16	9.71875
1	0.048613	1	1	128	Sigmoid	32/16/16	9.71875
1	0.010579	1	2	256	Tanh	32/16/16	10.4687
1	0.085243	1	2	256	Sigmoid	32/16/16	10.4687
1	0.011359	1	1	64	Tanh	32/16/16	11.5937
1	0.002533	1	1	256	Tanh	64/32/32	12.9375
1	0.010598	1	1	256	Sigmoid	64/32/32	12.9375

The parameters of CNN model are detailed in Table 3, the some parameters are get from the grid search. The settings of fully connected layer number and activation function set to 3 and tanh [4]. The number of output node is eight, its form is likelihood by softmax function.

4. Implementation

The main target in this paper is a low–cost indoor localization module. The RX65N MCU provided by Renesas Electronic Corporation is used to implemented the proposed algorithm. The build-in 2 M bytes of read only memory, ROM, and 640 K Bytes random access memory, RAM, are enough to install the optimized AI model, and the high-speed operation with 32bit floating point could support the calculation requirement of FFT.

Table 3 The Items and Parameter of Grid Search

Item	***Parameter***
CNN Filter Kernel number	1
CNN Filter Size	256*1
Activation Function of CNN	ReLU
Node of Fully Connected layer 1	32
Node of Fully Connected layer 2	16
Node of Fully Connected layer 3	16
Activation Function of Fully Connected layer	Tanh
Node number of Output layer	8
Activation Function of Output layer	SoftMAX

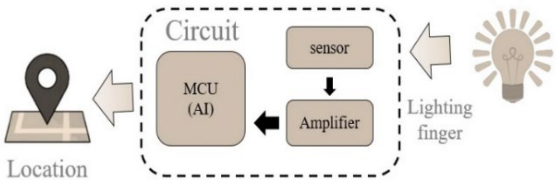


Fig.6 The architecture diagram of system

The proposed system is shown in Fig. 5, it includes a sensor circuit, amplifier block and MCU. The sensor circuit consists of a photodiode and resister, and the amplifier block is implemented with operational amplifier to perform the gain and filter functions. The system hardware is present as Fig. 6. Note, the hardware is only for experimentation but not size-optimization yet.

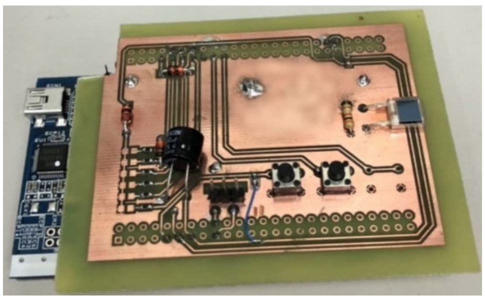


Fig. 7 The photo of proposed hardware

The signal processing flow is dedicated in Fig. 7, moreover, the processing time of every sub processing block is also shown. The longest duration is data acquisition. However, it is necessary, because the data is sampled three times, and there is 16,384 points sampled at 65,536 Hz every time. The transfer time from time domain data to frequency domain may be shorted if fast MCU. And before localization AI algorithm, the rescaling FFT results of three samples are averaged.

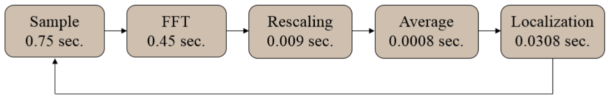


Fig. 8 The signal processing flow is dedicated

5. Experiment

A Setup

The experiment was performed in the EN303 room on third floor of the Engineering sixth hall, National Yunlin University of Science and Technology, as shown in Fig. 8. The proposed module is located about 1.7 meters below lamps. The numbers denote the lamps. The floor plan is revealed in Fig. 9, which also includes the location of lamps, the distance between each lamps and from lamp to hall. Note, the 'X' marks indicate the location when testing the lamp, and the '+' marks shows the position when testing not below the lamps.

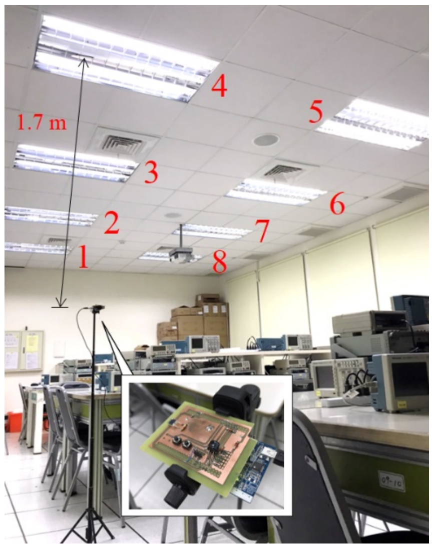


Fig. 9 Experiment in the EN303 room

B Results

The estimated likelihoods of classification are shown in Table 4, the proposed module is placed below the lamps which are marked by 'X' turn by turn. The number in the table indicates the most likelihood lamp number, and the number in the parentheses shows the estimated likelihood. The experiment ran 10 turns, and every 'X' point is tested every turn. The localization accuracy performed 100%. The least likelihood in the table locates at lamp 4 in the turn 8, however the likelihood still achieved 75.8%, it demonstrated the localization is correct.

The Table 5 present an interesting estimation results, the average likelihood of 10 turns is expressed by percentage, when the proposed module is placed at mark '+' position. The two adjacent likelihoods shown in bold are expected to be biggest two number, because the testing position is located between the two lamps indicated in the first line of table. However, the expectation was correct such as position 9, 11 and 13, but incorrect for other points. The reason should be that the sample located the position 9 to 14 never be adopted for AI model training and the estimated likelihood is not the interpolation of any two training points. The future work is to adopt such position sample to machine learning to evaluate the localization effect if the light sampled not below the lighting source.

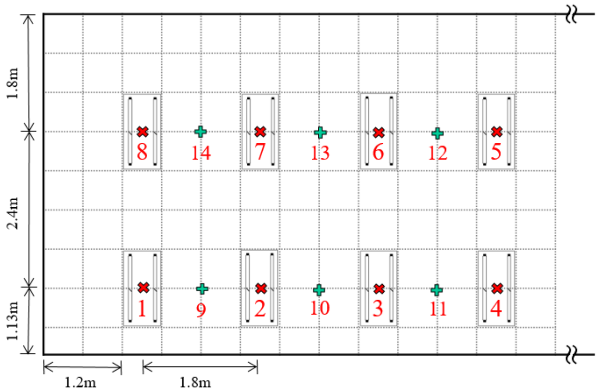


Fig. 10 Floor Experiment in the EN303 room

Table 4. Like hoods of Classification of the Turns

Turn	Lamp1	Lamp2	Lamp3	Lamp4	Lamp5	Lamp6	Lamp7	Lamp8
1	1(98.8)	2(98.9)	3(98.9)	4(97.6)	5(96.2)	6(98.9)	7(98.4)	8(99.1)
2	1(98.6)	2(98.9)	3(97.2)	4(97.6)	5(98.1)	6(98.8)	7(96.5)	8(99.1)
3	1(98.2)	2(98.9)	3(98.2)	4(95.3)	5(98.2)	6(98.8)	7(98.5)	8(99.0)
4	1(98.5)	2(98.9)	3(98.6)	4(93.1)	5(96.9)	6(98.9)	7(98.3)	8(99.1)
5	1(97.4)	2(99.0)	3(91.3)	4(95.1)	5(97.3)	6(98.9)	7(98.5)	8(98.9)
6	1(98.9)	2(99.0)	3(98.9)	4(96.4)	5(98.2)	6(98.8)	7(98.5)	8(99.1)
7	1(98.6)	2(99.0)	3(98.5)	4(94.2)	5(98.2)	6(98.9)	7(98.5)	8(99.1)
8	1(98.8)	2(98.9)	3(96.4)	4(75.8)	5(98.2)	6(98.9)	7(98.5)	8(99.0)
9	1(98.7)	2(98.9)	3(98.5)	4(93.9)	5(98.0)	6(98.8)	7(98.5)	8(99.1)
10	1(98.9)	2(98.9)	3(95.0)	4(90.8)	5(98.1)	6(98.8)	7(98.3)	8(98.8)

Table 5. Likehood Estimation Between Two Lamps

Pos.	Lamp1	Lamp2	Lamp3	Lamp4	Lamp5	Lamp6	Lamp7	Lamp8
9	**28.333**	**26.879**	1.589	3.464	36.366	1.612	1.506	0.251
10	0.113	**80.002**	**7.090**	0.288	0.033	10.378	1.790	0.306
11	0.138	0.054	**38.109**	**60.799**	0.490	0.056	0.300	0.055
12	0.033	0.652	0.017	0.123	**0.509**	**97.521**	0.863	0.283
13	0.011	1.321	0.046	0.267	0.757	**22.136**	**74.460**	1.001
14	0.080	0.034	0.004	0.193	0.657	0.976	**1.443**	**96.620**

6 Conclusion

In this paper, a low-cost indoor localization system using sound spectrum of light fingerprint is proposed. To use this light fingerprint based localization indoor mobile equipment, modularization is the key consideration. So, the memory and calculation of the selected MCU are limited due to cost. Only the rescaling frequency features between 100 Hz to 30k Hz are used in classification, and a simplified 1-D CNN is performed to classify the lamps. Grid search is adopted to optimize the hyperparameters to find the highest-accuracy and smallest-size model. The experiment shows the system could be implemented in a lost-cost MCU with few components. The on-line tests present the localization accuracy achieve 100% when the proposed device located below the lamps. However, not as expected, when testing position locates between lamps, the likelihood is not an interpolation of the nearest two lamps'. It may be the future works of this topic.

Acknowledgment

Part of this work was financially supported by the "Intelligent Recognition Industry Service Research Center" from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan. And it is also supported by Ministry of Science and Technology, Taiwan, under contract MOST 108-2221-E-224-045- and 108-2218-E-150-004-.

References

[1] Y. Hu et al., "Lightitude: Indoor Positioning Using Ubiquitous Visible Lights and COTS Devices," 2015 IEEE 35th International Conference on Distributed Computing Systems, Columbus, OH, 2015, pp. 732-733.

[2] C. Guo, J. H. Shao, W. Ke and C. Y. Zhang, "A Visible Light Indoor Positioning Algorithm Based on Fingerprint," 2018 4th Annual International Conference on Network and Information Systems for Computers (ICNISC), Wuhan, China, 2018, pp. 71-77.

[3] S. Hamidi-Rad, K. Lyons and N. Goela, "Infrastructure-less indoor localization using light fingerprints," 2017 IEEE International

[4] H. Kobayashi, "CEPHEID: the infrastructure-less indoor localization using lighting fixtures' acoustic frequency fingerprints," IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society, Lisbon, Portugal, 2019, pp. 6842-6847.

[5] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, ‘‘Real-time motor fault detection by 1-D convolutional neural networks,’’ IEEE Trans. Ind. Electron., vol. 63, no. 11, pp. 7067–7075, Nov. 2016.

[6] S. Kiranyaz, T. Ince, and M. Gabbouj, “Real-time patientspecific ecg classification by 1-d convolutional neural networks,” IEEE Transactions on Biomedical Engineering, vol. 63, no. 3, pp. 664–675, 2016

[7] Sun, Y.; Wang, Y.; Guo, L.; Ma, Z.; Jin, S. The comparison of optimizing SVM by GA and grid search. In Proceedings of the 2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), Yangzhou, China, 20–22 October 2017.

Air Valve Fuzzy Control Combined with Sheet Music Recognition Techniques Applied to Autoplaying Soprano Recorder Machines

Chun-Chieh Wang[1], Guang-Ming Jhang[2]

[1] *Department of Automation Engineering and Institute of Mechatronoptic Systems, Chienkuo Technology University, Taiwan*
(Tel: 886-4-7111111ext3350, Fax: 886-4-7111164)
[1]*jasonccw@ctu.edu.tw*
[2] *Department of Electronic Engineering, Chienkuo Technology University, Taiwan*

Abstract

In the past research, there are many disadvantages to score recognition and flute performance. Therefore, this paper would improve on these two points. For the part of music score recognition, we use the y-axis projection method to detect the staff position and remove it to replace the erosion and expansion in morphology. It was found that the notes have a specific writing style on the staff. Therefore, this feature is used to distinguish the notes. For the soprano recorder playing part, the finger-shaped electric arm is used to press the blow hole, it will cause the situation that the speed of the score cannot be kept up. Therefore, the motor is changed to a solenoid valve to facilitate the pneumatic cylinder to smoothly press the blow hole. In addition, since the difference in pitch of the soprano recorder requires different air pressure, we increase one valve to three valves. Moreover, the range is divided into bass, midrange, and treble. Not only that, fuzzy control theory is used to control the air valve to greatly improve the sound distortion caused by the original single air valve. Experiments prove that the air valve fuzzy control combined with sheet music recognition techniques can fully realize the functions of autoplaying soprano recorder machines.

Keywords: Autoplaying Soprano Recorder Machines (ASRM), LabVIEW, Arduino, Air Valve Fuzzy Control, Pneumatic Cylinder

1 INTRODUCTION

Nowadays, robots are no longer just used in industrial production, and they can also be seen in medical and artistic fields. The International Robotic Art Competition, which started in 2016, has also been held three times. Many works created by robots have reached a level comparable to human artists. As for performance robots, there have been significant advances and improvements due to artificial intelligence. Among them, the research on music score recognition technology has been proposed in many documents. Although the music notation is limited and there are restrictions on the writing position on the staff. However, there are still many blind spots to be overcome and improved in identifying it with artificial intelligence technology, even if it is only for non-handwritten scores.

In 2018, Yang S.F. [1] proposed sheet music detection techniques applied to fluted recorder robots. The disadvantage is that the mechanism is poorly designed, which often leads to inaccurate beats. In 2004, Tsai Z.W. [2] proposed the recognition system of printed music score. Due to the addition of a variety of staff removal and note recognition methods, the overall score recognition time was extended. Lee H.W. [3] and Peng Z.X. [4] use optical symbol recognition (OMR). Since it must build a symbol database, it is possible to recognize errors as long as there are slight errors. In 2009, Lan S.M. [5] presented the musical notation recognition system for two-wheeled robot. The disadvantage is that the recognized scores are only digital scores. In 2014, Wang B.R. [6] proposed a musical score recognition system for iOS devices, because its recognition process requires human assistance, so there is no way to achieve the effect of automatic recognition.

Based on the shortcomings of the above-mentioned scholars, this article proposes air valve fuzzy control combined with sheet music recognition techniques to improve the recognition of music scores and improve the phenomenon of out of sound. Experiments prove that presented method can fully realize the functions of ASRM.

2 ASRM

To improve the problem that the finger-shaped robotic arm cannot keep up with the rhythm of the music, the motor is changed to a solenoid valve and combined with the Arduino UNO board to control the air cylinder to press the blow hole. The finished ASRM is shown in Fig. 1.

Fig.1. Finished product of the ASRM

2.1 Control Systems

The Arduino UNO board is used as the control core of the ASRM. LINX in LabVIEW Graphical Programming environment is the bridge between LabVIEW program and Arduino UNO board. Like LIFA, LINX must first write a set of communicable commands in Arduino. The difference is that LIFA can provide users with the ability to directly use Arduino connection operation in LabVIEW without writing Arduino code. The LINX kit is an open source kit developed by MakerHub. Its feature is that it cannot only be connected to Arduino, but also supported by many development platforms such as myRIO/Digilent/. Moreover, LINX can directly burn the firmware required for LINX execution on the Arduino through this LINX Firmware wizard. The control system architecture diagram is shown in Fig. 2. Fig. 3 is the Arduino UNO control board.

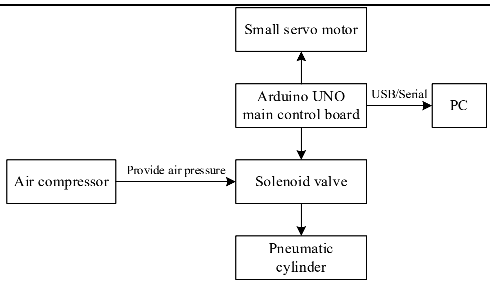


Fig.2. The control system architecture diagram

Fig.3. Arduino UNO

2.2 ASRM Platform

The physical architecture of the ASRM platform is shown in Fig. 4. The main feature is that the part of the original flute that presses the blow hole is pressed by the solenoid valve controlled pneumatic cylinder. Significantly improved the original inability to play music faster than the response speed of the motor. Not only that, the blowing control part has two more solenoid valves. It solves the problem of broken sound when blowing.

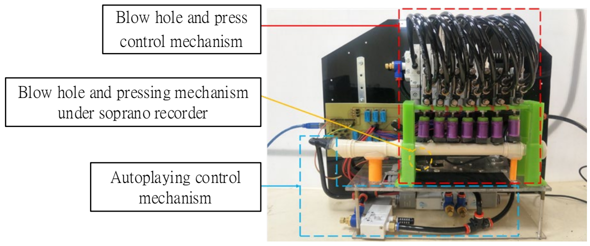


Fig.4. The physical architecture of the ASRM platform

3 MUSIC SCORE RECOGNITION SYSTEM

First, we remove the parts of the score that are not related to performance, as shown in Fig.5. The processed music score is input into the system. After that, binarization and scale recognition are performed through LabVIEW. Fig.6 is its program flow chart.

Fig.5. Score preprocessing

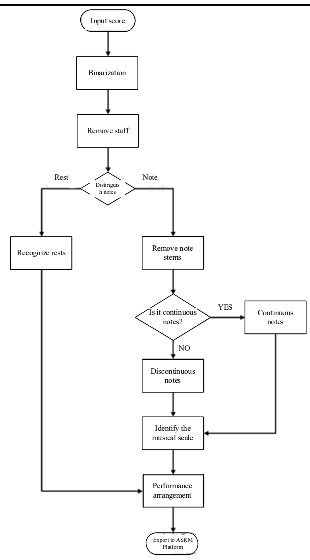


Fig.6. Program flow chart

We use the y-axis projection method to detect the staff position and remove it to replace the erosion and expansion in morphology. It was found that the notes have a specific writing style on the staff. Therefore, this feature is used to distinguish the notes. First, the note stems are distinguished and removed by the x-axis projection detection method. Second, the center point of the head of the note and the position of the staff are identified by the pixel clustering method. Finally, the encoding of the scale is arranged in sequence from bass to treble.

3.1 Binarization

Binarization is to divide the grayscale image into black and white according to the threshold set by the user. When the grayscale value of the pixel is greater than the threshold, it is judged as a white point, otherwise it is a black point. The grayscale image can be converted into a binary image through binarization, as shown in Fig.7.

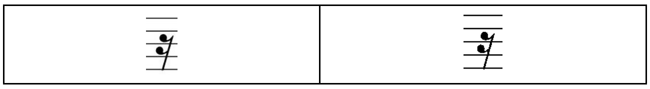
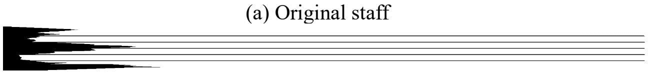


Fig.7. Difference before and after binarization

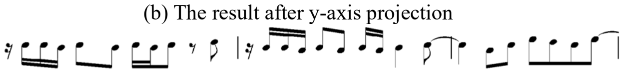
3.2 Remove staff

We use the y-projection in the orthogonal projection method to project the music score to be identified onto a single axis, which will form a graph called the projection profile, as shown in Fig.8. We can clearly find the position of the staff and remove it. This method greatly improves the original use of erosion and swelling in typology to cause unclear symbols, as shown in Fig.9.

(a) Original staff

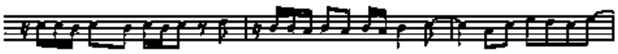


(b) The result after y-axis projection

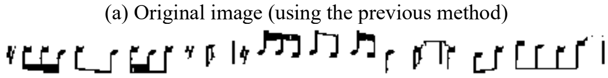


(c) The result after removing the staff

Fig.8. The process of removing the staff



(a) Original image (using the previous method)



(b) The result after removing the staff (using the previous method)

Fig.9. The process of removing the staff(using the previous method)

3.3 Symbol distinction

For rests, this article only deals with the commonly used full rests, two-quarter rests, four-quarter rests, eighth rests, and sixteenth rests. General notes are also processed for commonly used whole notes, half notes, quarter notes, eighth notes, and sixteenth notes. Generally speaking, the height of most rests is not greater than the height of the notes. For example, the height of a quarter note on the staff is approximately equal to the height of the staff, as shown in Fig. 10. The height of the quarter rest is only 3/4 of the height of the staff, as shown in Fig.11. Based on this, the notes are preliminarily divided into rest notes and general notes to facilitate subsequent identification.

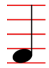
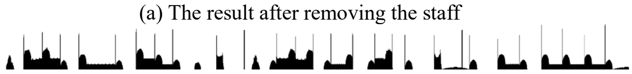


Fig.10. Quarter note

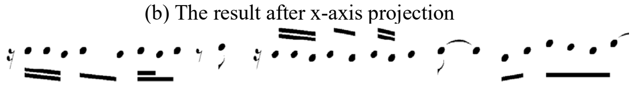
Fig.11. Quarter rest

To distinguish between continuous notes and discontinuous notes, the image is projected on the x-axis using the x-projection in the orthogonal projection method. The note stems exceeding a certain value are removed, as shown in Fig.12. Moreover, the pixel clustering method will be used to distinguish the following three types: discontinuous notes, continuous notes, and discontinuous notes but with tails, as shown in Fig. 13.

(a) The result after removing the staff

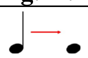


(b) The result after x-axis projection



(c) The result after removing note stems

Fig.12. The process of removing note stems



(a) discontinuous notes

(b) continuous notes

(c) discontinuous notes but with tails

Fig.13. Comparison chart after removing note stems

The pixel clustering method uses the IMAQ Count Objects 2 VI component in the Vision development module in LabVIEW. The function of this component is to cluster the binarized pixels. Let the user set the pixel threshold to distinguish the number of pixel groups. Fig.14 is the result of pixel cluster identification.

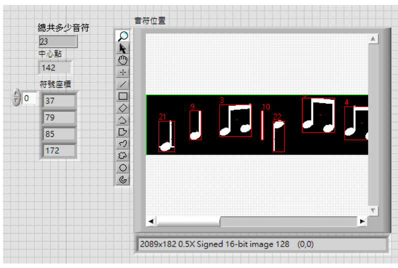


Fig.14 The result of pixel cluster identification

3.4 Scale recognition

While the scale recognition is performed, the note timing is also recognized. Therefore, the head and tail of the note must be distinguished first. First, the discontinuous notes have been separated out after removing the note stem, so there is no need to deal with it. The part with discontinuous

notes but with tails is distinguished by the proportion of black pixels in the image extracted by pixel clustering. The pixel ratio calculation is based on the range enclosed by the red frame to calculate the image size and the ratio of black pixels, as shown in Fig.15. Continuous notes cannot be identified by this method because the proportions of black pixels of note stems and note heads are too close. Therefore, we use the aforementioned note stems to account for more than 2/3 of the overall note width to distinguish. The position of the beam will change due to the way the music theory is written. So we divide the continuous note from the center point into the upper and lower parts, as shown in Fig. 16. ((a) Original image with Note head at the bottom, (b) Original image with Note head at the top)

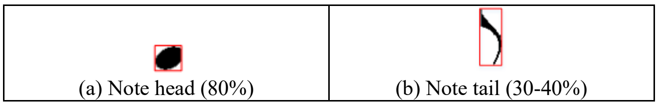


Fig.15. Distinguish of discontinuous notes but with tails

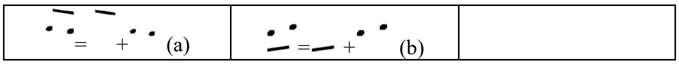


Fig.16. The process of cutting continuous notes

For judgment of discontinuous notes, the difference between discontinuous notes is whether the beam and the head are hollow. We use the ratio of black pixels in the red box to distinguish half notes from quarter notes, as shown in Fig.17. Because we have already recorded the position of the staff while removing the staff. Therefore, we only need to extract the center point of the talisman through the pixel clustering, and then compare it with the previously recorded staff position to know which line or interval the talisman is located on. Put it into the scale table of the recorder to get the scale of the note, as shown in Fig.18 and Fig.19.

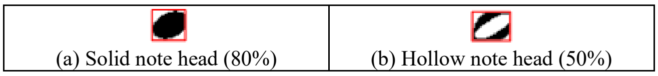


Fig.17. The solid and hollow of the note head

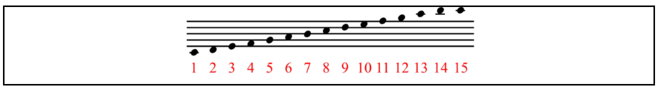


Fig.18. Scale table of the soprano recorder

3.5 Rest identification

The identification of rests is double identification using orthogonal projection method and pixel clustering. According to the appearance of commonly used rests, they are divided into three types: (1) whole rest and half rest, (2) quarter rest, (3) 8th rest and 16th rest. First, we use the pixel clustering method to capture the rest image, and use the black pixel ratio in the red frame to make a preliminary judgment, as shown in Fig.20. Second, make a detailed distinction for the above three types. (1) The image coordinates of the y-axis projection of the full rest and the bipartite rest are captured and compared with the staff position. If the image coordinates are closer to the fourth line of the staff, it is judged as a whole rest. If the coordinates are closer to the third line, it is a half rest. (2) Compare the height value captured after y-axis projection with the highest point value after x-axis projection. If the values are similar, it is judged as a quarter rest, as shown in Fig.21. (3) The 8th rest has only one peak in the y-axis projection image. The 16th rest has two peaks, as shown in Fig.22 and Fig.23.

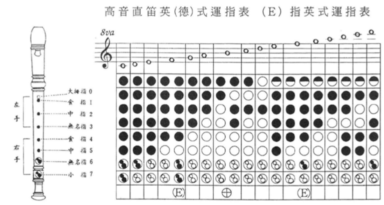


Fig.19. Scale press fingering table of the soprano recorder

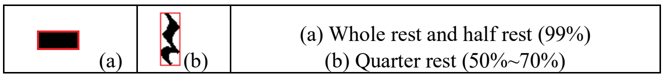


Fig.20. Distinguish between rests

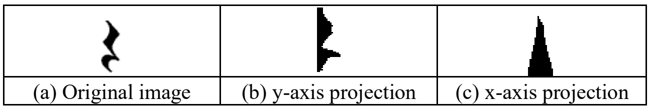


Fig.21. Quarter rest

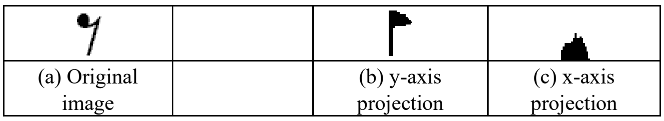


Fig.22. 8th rest

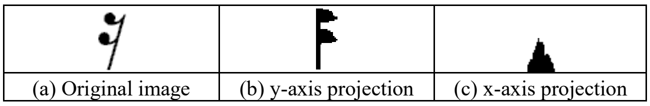


Fig.23. 16th rest

4 AIR VALVE FUZZY CONTROL DESIGN

To provide a smoother air pressure with different sound ranges, this article increases the number of solenoid valves to 3. In addition, we use the fuzzy control law to write the valve control program. The range (R) is divided into bass, midrange, and treble, as shown in Fig.24.

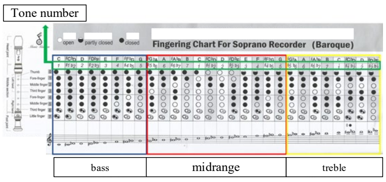


Fig. 24. The range

Define the opening degree of air pressure valve be V. TS1, TS2, and TS3 belong to Takagi-Sugeno type. Fuzzy IF-THEN rules are expressed as follows.

IF R is bass, THEN V is TS1.

IF R is midrange, THEN V is TS2.

IF R is treble, THEN V is TS3.

5 EXPERIMENTAL RESULTS

In view of the large differences in the way of notation for tuplets (Fig.25(a)) and dotted notes (Fig.25(b)) by musicians around the world, it is impossible to effectively identify all these two symbols with a general rule. Therefore, this article will ignore it during the experiment. In addition, the score used for testing is taken from a web site made by netizens and provided free of charge [7, 8]. The scores we used to test included 5 Mandarin pop songs, 2 Japanese pop songs, and 3 movie theme songs. Please refer to the following URL directly for the experimental results. https://www.youtube.com/watch?v=1Z-kPIPQG-U

 (a)	 (b)	**Fig. 25(a).** Tuplets **Fig. 25(b).** Dotted notes

6 CONCLUSION

This paper uses the pixel clustering method, the x & y axis projection method to successfully improve the score recognition results. Moreover, the finger-shaped electric arms are changed to solenoid valves to facilitate the pneumatic cylinder to smoothly press the blow hole. This method has also successfully improved the phenomenon of air leakage. Not only that, fuzzy control theory is used to control the three air valves to greatly improve the sound distortion caused by the original single air valve.

REFERENCES

[1] Yang S.F. (2018), Sheet Music Detection Techniques Applied to Fluted Recorder Robots, Master Thesis of Chienkuo Technology University.

[2] Tsai Z.W. (2004), Recognition System of Printed Music Score, Master Thesis of National Sun Yat-sen University.

[3] Lee H.W. (2016), Automatic Music Score Recognition, Master Thesis of National Tsing Hua University.

[4] Peng Z.X. (2014), Automatic Music Score Recognition System Using Digital Image Processing, Master Thesis of Chung Yuan Christian University.

[5] Lan S.M. (2009), Musical Notation Recognition System for Two-Wheeled Robot, Master Thesis of National Taiwan University of Science and Technology

[6] Wang B.R.(2014), A Musical Score Recognition System for iOS Devices, Master Thesis of National Kaohsiung First University of Science and Technology

[7] Egg Head Liu https://www.youtube.com/channel/UCC6UAv477s0YiQIfioIj0Vw (Search date: 12/15/2020)

[8] TSAI,JU-HSIN https://www.youtube.com/channel/UChE9nxA6DKQUvsJ-HBSgUJQ (Search date: 12/15/2020)

Image Inpainting Techniques Combined with Isolated Pixel Filtering Applied to Multifunctional Drawing Robots

Chun-Chieh Wang[1], Zhan-Xian Ye[2]

[1] Department of Automation Engineering and Institute of Mechatronoptic Systems, Chienkuo Technology University, Taiwan
(Tel: 886-4-7111111ext3350, Fax: 886-4-7111164)
[1]jasonccw@ctu.edu.tw
[2] Department of Electronic Engineering, Chienkuo Technology University, Taiwan

Abstract

The purpose of this thesis is to assist street painters to improve their work efficiency. That is, the simpler part of the drawing is drawn by the robot. Once the initial picture is completed, the artist will take over the painting. In this paper, HSI color space is used to improve the effect of color simplification such that the recognition of image processing results is enhanced. Moreover, the isolated pixel filtering is used to replace the less-affected isolated point color with the surrounding color. Furthermore, image inpainting techniques are utilized to reduce the distortion caused by the isolated pixel filtering. Besides, we adjusted the path planning as well as reduced isolated points to dramatically reduce drawing time. At the same time, LabVIEW can issue commands directly to control the robot by adding the communication function. In addition, to achieve the multifunctional drawing robotic mode, this system adds the sketch function. Through the image resolution adjustment as well as the shortening of the spacing of the drawing lines, the robot can draw more detailed pictures in the same size of drawing space. The measured results confirm that the application of the technology in this paper can shorten the drawing time by about 55% on the multifunctional drawing robot system.

Keywords: Isolated pixel filtering, Image inpainting, Watercolor, Sketch, Multifunctional drawing robots

1 INTRODUCTION

In view of that many street painters usually takes a long time to draw a picture such that it is impossible to produce a large amount in a short time. Therefore, we hope to develop a drawing robot that can assist painters. With the help of robots, the artist can complete more paintings in the same time and improve drawing efficiency.

In the literature of robot drawing in the past, the image detection applied in a watercolor painting robot proposed by Hung YY [1] in 2017. It is characterized by using HSV color simplification method with a self-made drawing mechanism, and then using a limited variety of colors to complete the drawing. Although the recognition of the image is acceptable, the long drawing time is a major disadvantage. In 2016, Hong M.J. [2] proposed robotics artistic colorful picture drawing and painting using visual feedback control system. Two different painting styles were used. 2017 Columbia University Creative Machines Lab / PIX18 [3] uses Gantry multi-axis robotic arm to draw pictures with artificial intelligence. In the same year, Massachusetts Institute of Technology Dane Kouttron et al. [4] used online PNG to SVG software to simplify the colors used Linux CNC with a multi-axis robotic arm for drawing.

For the sketch drawing part, Lin L.C. [5] used edge detection to generate contour images and used mobile robots to draw on a flat surface in 2017. In 2009, Mac T.T. [6] and Chuang L.W. [7] used PI and PID controllers to adjust motors to control multi-axis robotic arms to draw portraits, respectively. The above three documents are all drawn with the edge of the image. In 2016, Jhang G.M. et al. [8] used multi-layer binary images to overlap drawing. It focused on the performance of shadows to achieve the effect of light and shadow drawing.

In order to improve the above shortcomings, we apply image inpainting techniques combined with isolated pixel filtering to multifunctional drawing robots.

2 RESEARCH METHOD

2.1 HSI color model and color simplification

Digital images often use the RGB color space to define colors. The parameters of the RGB color space are all connected, which makes image processing difficult. In contrast, the parameters of the HSI color space are independent of each other, namely Hue (H), Saturation (S), Intensity (I). Although the HSI color model is very similar to the HSV color model, there is a difference in the representation of the Intensity of HSI and the Value of HSV, as shown in Fig.1. Fig. 1(a) is a chromatogram image created by using the S and V parameters in HSV. Fig. 1(b) is a chromatographic image created by using the S and I parameters in HSI. From these two color spectrums, it can be found that the overall color change of the HSI color model is relatively smooth. In contrast, the HSV color model has more compact changes between light colors and primary colors, and it is more difficult to set the color reduction range. Therefore, this article uses the HSI color model with smoother color changes for color simplification. The advantage is that the parameter range setting is more flexible.

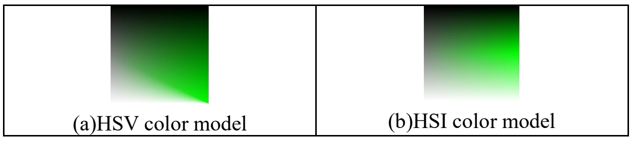


Fig.1. Comparison of HSV and HSI Green Chromatogram

In this paper, the scanning method from left to right and top to bottom is used to perform HSI color simplification on the color of each pixel. First, the RGB value of the pixel color is captured through the system. Since the color reduction method uses the HSI color space for processing, it is necessary to convert the RGB value to the HSI value before further processing, as shown in Fig.2. Since the hue representation in HSV and HSI color models is the same, the hue part continues the setting of reference [1], and its hue color distribution is shown in Fig. 3.

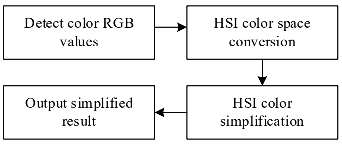


Fig. 2. The process of color simplification

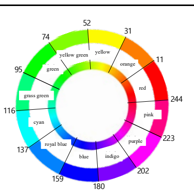


Fig. 3. Distribution of 12 hues

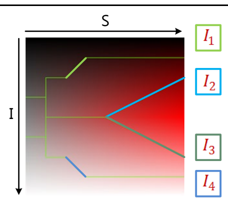


Fig. 4. Distribution of Color simplification range

After determining the range of hue, the range of 6 colors (Dark, light, primary, black, gray and white) is set. Moreover, the color spectrum established by the HSI color space is used for adjustment. Since everyone's perception of color is different, this article is based on the author's perception of color to simplify the range of adjustments.

$$I_1 = 128 - S \quad (1)$$

$$I_2 = \left|\frac{24448 - (64 \times S)}{127}\right| \quad (2)$$

$$I_3 = \left|\frac{8064 - (64 \times S)}{127}\right| \quad (3)$$

$$I_4 = 128 + S \quad (4)$$

In Fig. 4, in the diagonal part, we use Eq.1~Eq4 to import the S value to get the relative range of I value. The scope of HSI color simplification is summarized in a rule table. Through the H, S, and I parameters in the captured target pixel, we can determine from the rule table which shade of color the pixel belongs to. In order to avoid errors in recognizing pictures with the naked eye due to too close shades, the 39 simplified colors are fine-tuned. The color codes used are shown in reference [9].

2.2 Isolated pixel filtering for watercolor system

In the process of image thumbnailing, the intersection of different colors is prone to color mixing. This change in color mixing is usually not too large. We want to remove the mixed color as much as possible and replace it with the surrounding colors. In addition, filtering out isolated pixels with low influence in the image can reduce the time it takes to repeatedly raise and lower the pen. Therefore, this paper proposes to filter out the isolated pixels and replace them with surrounding colors to eliminate the isolated pixels, so as to reduce the time consumed for lifting and lowering a single pixel. The method of filtering out isolated pixels starts from the upper left corner of the picture. A 3 × 3 matrix is used to scan from left to right and top to bottom. The center color of the matrix is used as the target for comparison. When the neighboring points have the same color, the original color of the target is maintained. If there is no same color, it is judged as an isolated pixel and replaced with the color with the largest surrounding area. As shown in Fig. 5(a), the central color is an isolated pixel, which will be replaced with the color of the largest area around it. When the surrounding eight pixels have the same color, they are not isolated points, as shown in Fig. 5 (b).

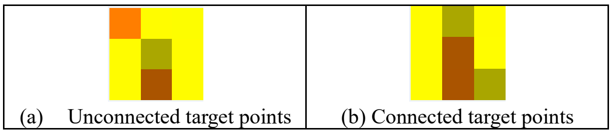


Fig. 5. Judgment of target point connectivity for watercolor

When there are two or more colors with the largest pixel area around the target, the system will scan from left to right and then from top to bottom according to the scanning order, starting from the top left pixel, as shown in Fig. 6. After that, the first and second ranking colors are taken out. The HSI color space is used to convert these two colors to the target color. At the same time, H is used to determine which color is closer to the target color. After that, the color with a closer hue replaces the target color.

(1)	(2)	(3)
(4)	Target	(5)
(6)	(7)	(8)

Fig. 6. Scan order of colors around the target

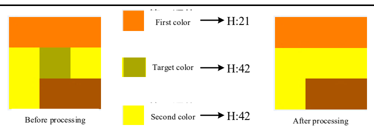
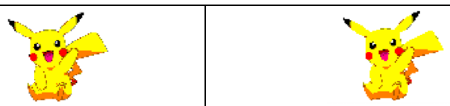


Fig. 7. Judgment method of the two largest color pixel areas around the target

Taking Fig. 7 as an example, since the number of orange and yellow pixels is 3, it is impossible to determine which color replaces the target color. Therefore, it is necessary to judge by hue. First, it is determined by the scanning order that orange is the first color and yellow is the second color. Secondly, convert the color to HSI color space, and compare the H parameters with each other. The hue of the target color in the figure is the closest to the hue of the second-ranked color, so the target color is replaced by yellow. Fig. 8 is a comparison diagram before and after the isolated point filtering process.



(a) Color simplified image (b) Image after isolated pixel filtering

Fig. 8. Before and after filtering isolated pixels

2.3 Feature inpainting for watercolor system

After the isolated points are filtered out, some features in the image may be affected and cause distortion. The right-hand line features are distorted, as shown in Fig. 8(b). Therefore, a new feature inpainting method is proposed in this article. That is, a 3×3 array is used to determine whether it needs to be inpaintinged. First, the images before and after the isolated points are filtered out are subtracted from each other. In the subtraction result, the part where the pixel value is not equal to 0 is the position filtered by the isolated point. The color of the pixels before the filtering of the processed position points is captured from the image before the isolated points are filtered. At the same time, in the image after the isolated points are filtered, these location pixels are found and centered on it. Then the 9 × 9 matrix around it is extracted. If the color of the remaining 8 cells except the center point in the matrix is the same as the color of the pixel before filtering, it is judged to be inpaintinged. This inpaintinging action is to restore the "pixel in the image after filtering out the isolated point" to the "color of the pixel before filtering out".

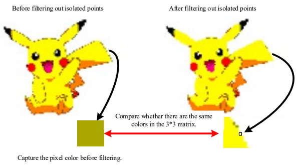
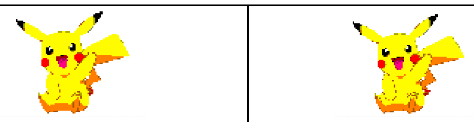


Fig. 9. Judgment method of image inpainting



(a) After filtering out isolated points (b) Results of image inpainting

Fig. 10. Comparison before and after image inpainting

2.4 Optimized path movement for watercolor system

The area of the watercolor painting system consists of three parts: drawing, paint dipping, and brush cleaning. When moving in the drawing area, you only need to move the pen up and down by 4mm. In the paint picking and brush cleaning area, the brush needs to be raised to 16mm. Adjust the pen lift height for different areas to improve efficiency. Fig. 11 is a schematic diagram of the movement of the brush.

2.5 Grayscale

Most common images are color images in RGB color space. However, the common sketches are painted in black, white and gray. Therefore, we convert the RGB image into a grayscale image. The conversion formula is as follows:

$$Gray = 0.299R + 0.587G + 0.114B \tag{5}$$

2.6 Binarization

We use binarization to divide the grayscale image into black and white. Black represents the area not to be drawn, and white represents the area to be drawn. By setting 4 different thresholds, the specific part of the grayscale image is captured, as shown in Fig. 12.

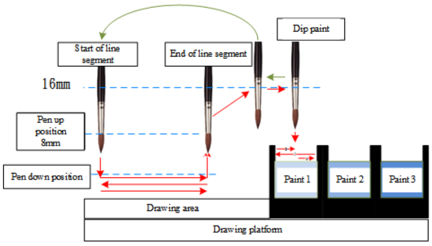


Fig. 11. Schematic diagram of the movement of the brush

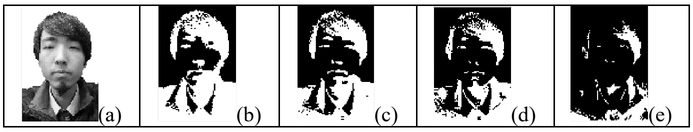


Fig. 12. Results of multiple binarization: (a) grayscale; binarization-(b) the shallowest (c) the second shallow (d) the second deep (e) the deepest

2.7 Isolated pixel filtering and image inpainting for sketch system

The isolated point filtering and image inpainting in the sketch system will process the black and white pixels in the binary image, as shown in Fig. 13. If there is the same color around the target point, it is judged as connected. Conversely, if there is no same color around the target point, it is judged as an isolated point. The image inpainting method is the same as in section 2.2

.

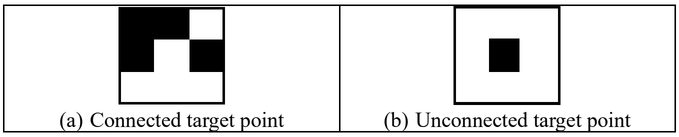


Fig. 13. Judgment of target point connectivity for sketch

2.8 Drawing method for sketch

We use lines in different directions for multi-level depiction. Draw the four shades of images mentioned earlier with lines in different directions, as shown in Fig. 14. The more the lines are stacked, the darker the color. In this way, the changes in depth are expressed.

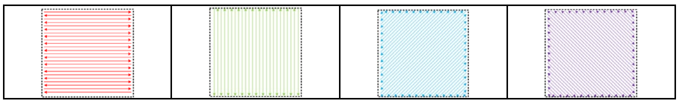


Fig. 14. Drawing trajectory in four directions

3 SYSTEM ARCHITECTURE OF DRAWING ROB OTS

3.1 Watercolor system

The process of watercolor drawing system is shown in Fig.15. The direct connection between the drawing platform and the computer replaces the previous steps of manually loading data into the drawing platform to complete the automation of the watercolor painting system.

3.2 Sketch system

The process of the sketch system is shown in Fig.16. By adjusting the image resolution and the line spacing of the drawing track, the drawing results are more detailed.

3.3 Communication system

To improve the convenience of use, we have added a communication part to the program. The user can decide whether to draw directly with LabVIEW through the buttons in the HMI. After the program converts the processing result and saves it as a G-Code file, if the button in the human-machine interface is not pressed, the program will be directly terminated, allowing the user to transfer data through the SD Card. If the user presses the button, LabVIEW will connect with the robot and send the drawing path data for drawing.

3.4 Configuration of drawing platform

Fig. 17 (a) is the plane configuration of the drawing area. The physical drawing of the multifunctional drawing robot is shown in Fig. 17 (b).

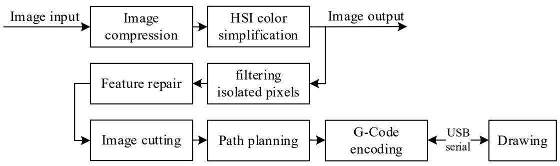


Fig. 15. The process of watercolor drawing system

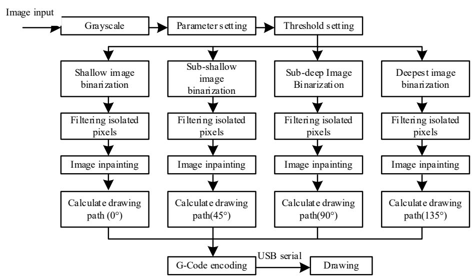


Fig. 16. The process of the sketch system

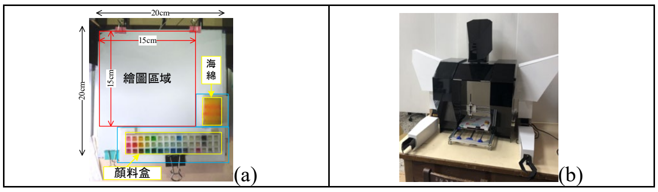


Fig. 17. (a) The plane configuration of the drawing area (b) The physical drawing of the robot

4 EXPERIMENTAL RESULTS

4.1 Watercolor drawing results

Fig. 18 shows the drawing results of Bill Gates photo.

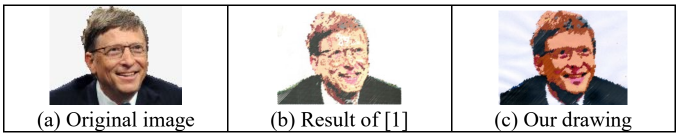


Fig. 18. The drawing results of Bill Gates photo

4.2 Sketch drawing results

Fig. 19 shows the drawing results of different line spacing. Although the drawing time with denser lines will increase a bit, the recognition of the portrait will be significantly improved.

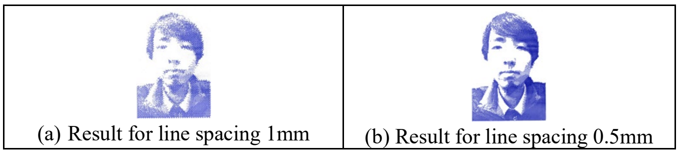


Fig. 19. The drawing results of different line spacing

5 CONCLUSION

In this paper, HSI color space method strengthens the recognition of image processing results. Moreover, we use the isolated pixel filtering, image inpainting techniques and path planning optimization to effectively shorten the drawing time. In addition, through the image resolution adjustment as well as the shortening of the spacing of the drawing lines, the robot can draw more detailed pictures. Experimental results prove that image inpainting techniques combined with isolated pixel filtering have been successfully applied to multifunctional drawing robots.

REFERENCES

[1] Hung Y.Y. (2017), Image Detection Applied in a Watercolor Painting Robot, Master Thesis of Chienkuo Technology University.

[2] Hong M.J. (2016), Robotics Artistic Colorful Picture Drawing and Painting Using Visual Feedback Control System, Master Thesis of National Taiwan University.

[3] Columbia University Creative Machines Lab PIX18 https://www.hodlipson.com/robotic-art.html (Search date:12/28/2017)

[4] Massachusetts Institute of Technology BABOT http://transistor-man.com/bluebot_revival.html (Search date:12/28/2017)

[5] Lin L.C. (2017), Trajectory Planning and Tracking Control of a FPGA-based Mobile Robot in a Picture Drawing Application, Master Thesis of National Taiwan University of Science and Technology.

[6] Mac T.T. (2009), Autonomous Drawing Manipulation of Humanoid Robot Arm, Master Thesis of National Taiwan University of Science and Technology.

[7] Chuang L.W. (2009), Human Portrait Generation System for Humanoid Robot Arm Drawing, Master Thesis of National Taiwan University of Science and Technology.

[8] Jhang G.M., Li C.J. (2016), Multilevel Based Image Recognition Applied in Automatic Sketch XYZ Platform, Graduation monograph of Chienkuo Technology University.

[9] Ye Z.X. (2018), Isolated Pixel Filtering and Image Inpainting Techniques Applied in System Improvement of Multifunctional Drawing Robots, Master Thesis of Chienkuo Technology University.

Mobile Robot with Image Recognition -- Using LabVIEW and KNRm

Kuo-Hsien Hsia
Bachelor Program in Interdisciplinary Studies
National Yunlin University of Science and Technology
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.

Bo-Jung Yang
Graduate School of Engineering Science and Technology
National Yunlin University of Science and Technology
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.

Jr-Hung Guo
Bachelor Program in Interdisciplinary Studies
National Yunlin University of Science and Technology
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.

Chang-Sheng Xiao
Department of Electrical Engineering
National Yunlin University of Science and Technology
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan, R.O.C.
E-mail: {khhsia, D10810211, jrhung, M10712003}@yuntech.edu.tw
www.yuntech.edu.tw

Abstract

The main purpose of this paper is to use the image recognition of LabVIEW software to construct mobile robots with various functions, and make the robots applicable to the industry and have web monitoring applications. The core of the robot is mainly the KNRm controller. This controller is suitable for beginners, and can be connected to DC servo motor, RC servo motor, infrared, ultrasonic and camera to achieve various functions of the robot. The structure of the robot uses metal parts sold by Studica company, which can be in accordance with the desired function to assemble the robot. Since the company is a designated equipment sponsor company for World Skills competitions, it can also be in line with international standards. Finally, PID control and sensors are added to make the robot movement and position more accurately.

Keywords: Image Recognition, Mobile Robot, Web monitoring, KNRm.

1. Introduction

With the increasing advancement of science and technology, the products of robots are becoming more and more diversified, especially in terms of mobile robots. Mobile robots are generally closely related to people's lives, such as sweeping robots and robots used in restaurants. Many countries are focused on the robot industry. And the growth of mobile robots will be faster than industrial robots. The design and development of mobile robots has also received increasing attention in the field of education and competition [1, 2, 3]. Taking the high school industrial student skills competition held by the Ministry of Education, which is the most important high school competition in Taiwan, as an example, the number of participants has grown from 86 to 102 in 3 years [4]. For Taiwan, which is facing a declining birthrate, this is a big growth rate. This shows the booming development of robot education.

2. Problem Description

This paper takes the project of the 49th National Skills Competition in Taiwan as the subject of problem solving. The competitors are asked to design a toy organizing robot. This robot can help the children put the toys into categories after the game. The layout of the competition field is shown in Fig. 1. Zone A in Fig. 1 is the starting area of the robot, and the size of the robot must be completely within the range of the starting area. There is a height-limit gate on the passage as shown in Fig. 1 marked as G, and the robot must be below this height to pass the gate. Zone B is the golf ball sprinkling area. Before the start of the evaluation, 5 to 6 golf balls of 5 colors are sprinkled via the sprinkler, as shown in Fig. 2, to Zone B. Zone C is the target ball-collection area. There are five slots in Zone C. A barcode is hanged on the wall of each grid to indicate balls of what color should be placed in the grid. At the beginning, the robot starts from the starting area, and firstly reads the barcode of each grid in Zone C, and then passes through the passage and gate to the golf sprinkling area to pick up the golf ball. After gripping, the robot returns to the ball-collection area and places the balls according to the color assigned by the barcode. The robot must collect the balls as much as possible and place them to the correct slots within the limited time duration, and return to the departure area before the end of the time. In order to test the design of the robot mechanism under limited resources, the number of motors and sensors are limited as mentioned in the test project.

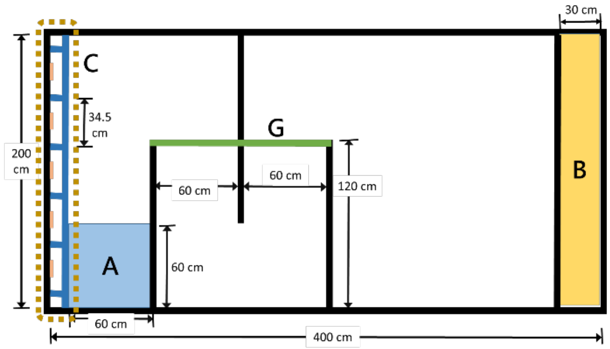


Fig. 1. Field of the considered problem.

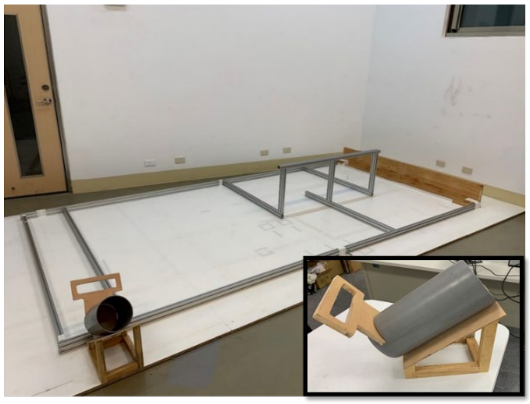


Fig. 2. The field with the sprinkler.

3. Hardware Design

We mainly use the integrated kit, including Tetrix metal blocks, sensors and motors packed by Studica, to design and assemble the robot according to the assigned functions. As for the controller, the KNRm developed by KKITC is used, and its main core is NI myRIO-1950. The main difference from KNRm to myRIO is that a single myRIO needs to be wired by oneself or a suitable expansion board needs to be made by oneself, and the useful pins can be wired out. Most of the pins that need to be used in KNRm have been designed around the controller for user convenience. Like myRIO, KNRm can be programmed with LabVIEW, developed by National Instruments (NI) Company. Different from traditional programming language, LabVIEW uses G language, a data-flow design, that allows users to programming according to the desired functions in conceptualization. In addition, G language has the ability of parallel processing [5].

In terms of mechanism design, because the golf balls have to be gripped out from the sprinkling area, a lifting mechanism and a ball clamping mechanism are necessary. In addition, the picked golf balls need to be stored on the robot. Hence a ball storage mechanism is required. It requires the use of 3 RC servo motors to meet above requirements, so an RC expansion box must be used. Due to the use of the RC expansion box will occupy one DC motor-port and there are only 4 motor-ports in a KNRm, only 3 DC motors with omni wheels can be used for the chassis. With omni wheels, it can achieve the functions of forward, backward, left and right motion, rotation and diagonal movement of the chassis.

The analysis about the robot body design is as following.

1. The ball clamping mechanism: Since the golf balls have to be moved to the collection area in a limited time, the design of the clamp mechanism is to clamp as many golf balls as possible. However, the blocks provided by Studica are not suitable for use. We design a customized part by 3D printing.
2. The ball storage mechanism: Due to the need of storing multiple golf balls on the robot without increasing the number of RC motors, a ramp with end-block is designed to allow the balls to roll into for storage.
3. The lifting mechanism: In order for the balls rolling smoothly into the ball storage mechanism, it is necessary to raise the clamp mechanism to a certain height. In addition, beside the lifting mechanism, a mechanism that can move back and forth must also be connected.

The overall design of the mechanism of the robot is shown in Fig. 3.

For obstacle avoidance, we use ultrasonic sensors as the main sensor, and infrared sensors as an auxiliary. The reason is to improve the accuracy of travel and overcome the lack of infrared rays that can pass directly through when encountering certain obstacles. The advantage of ultrasound is that the distance is farther than infrared. The ultrasonic sensor used in this paper is PING))), and the infrared sensor used is SHARP GP2Y0A21YK0F.

In the motion control part of the robot, we use trapezoidal acceleration and deceleration and PID control for control.

4. Image Recognition

In the problem to be solved in this paper, barcode reading and color determination are needed. Therefore, we use Logitech C310 camera as the image sensing device. In LabVIEW, users can combine their own image processing programs with built-in VIs, or use the Vision Assistant function of LabVIEW to achieve quick start processing. Vision Assistant is an integrated program for LabVIEW image processing, which has many image processing functions. The Vision Acquisition program can be used to set up the camera, including Acquisition Source, Acquisition Type, camera resolution and frame

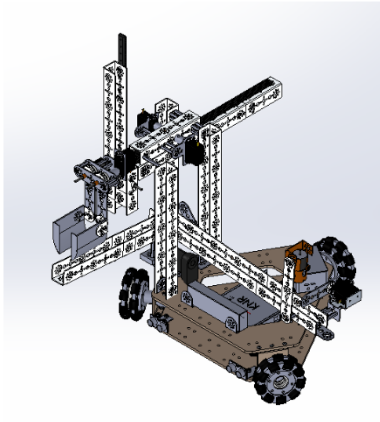


Fig. 3. The mechanism of the designed robot.

number, etc. Then use Vision Assistant to process the image obtained by the camera. In Vision Assistant, one can set and process the options of Image, Color, Grayscale, Binary, Machine vision, Identification and other items. Because the images we deal with in this paper are mainly the interpretation and recognition of bar codes and golf balls, and bar code interpretation is a built-in function in LabVIEW, we use Find Circular Edge and color comparison directly instead of the technique of image gray-scale processing in the paper.

5. Experimental Results

We conducted field test with the robots we designed. After the robot leaves the starting area, it reads the barcodes first, as shown in Fig. 4, and stores the color specified for each ball slot. After that, the height of the robot must be lowered so that it can pass through the gates in the passage smoothly. After the robot has completely passed the passage, it uses the side wall to correct the robot's posture, and turns to make the gripper face to the golf sprinkling area, and then goes to the sprinkling area. In order to facilitate the gripping of golf balls scattered in the sprinkling area, the robot will sweep the balls into the same area after reaching the sprinkling area, and then perform the action of gripping the balls. After performing multiple clamping, the robot then returns according to the path it came. When passing through the gate, the clamping mechanism should be lowered to pass it. After passing through the gates, the gripper returns to the original position, and place the golf ball in the corresponding grid when the robot passes through each collection area. Finally, after releasing all the balls, the robot returns to the starting area to complete the experiment. After actual testing, the robot we designed can be successfully used to achieve the tasks required by the problem and return to the starting area, as shown in Fig. 5.

6. Discussions and Conclusions

In this paper, a KNRm is used as the core controller of the robot, combined with the Tetrix metal building blocks, three motors, three RC servo motors and appropriate sensing devices, to design a solution that can solve the project of Mobile Robotics of 49th National Skills Competition in Taiwan. The designed robot has been verified on the field and found that it can indeed achieve the specified task.

Because it is composed of metal building blocks, the designed robot will be insufficient in terms of mechanical rigidity. When designing robots to solve other application problems, metal building blocks similar to this paper can also be used to design the prototype of the robot to test whether the function can meet the requirements or not. When the designed robot can meet the requirements, one can proceed with the appearance design of the robot to be used as a commercial product. At the same time, the mechanical rigidity of the robot is included in the material consideration. This can greatly reduce the time for product development, and can more accurately demand for application products.

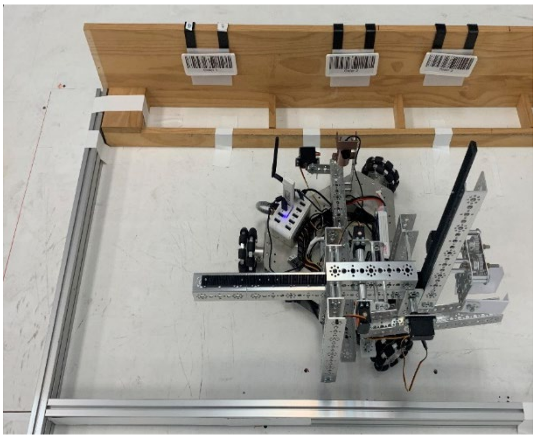


Fig. 4. The robot starts and reads the barcodes.

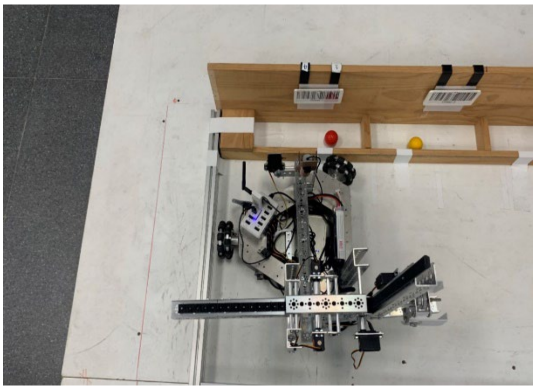


Fig. 5. Task completed.

References

1. S.E. Jung and E.S. Won, Systematic Review of Research Trends in Robotics Education for Young Children, *Sustainability* **10**(4), 2018, 905.
2. D. Karahoca, A. Karahoca, H. Uzunboylu, Robotics Teaching in Primary School Education by Project based Learning for Supporting Science and Technology Courses, *Procedia Computer Science* **3**, 2011, pp. 1425-1431.
3. D. Bazylev, A. Margun, K. Zimenko, A. Kremlev, E. Rukujzha, Participation in robotics competition as motivation for learning, *Procedia - Social and Behavioral Sciences* **152**, 2014, pp. 835-840.
4. National High Schools Skills Competition. (in Chinese) Available: http://sci.me.ntnu.edu.tw/Contest/ Retrieved: 11/30/2012.
5. C.H. Tseng, W.H. Guan, W.H.Wu, *Advanced Teaching Manual of Robot with LabVIEW - Building Robots with Both Intelligence and Interest*, 2nd edn., GOTOP Press, Taipei, 2014. (in Chinese)

Nonlinear Internal Model Controller based on Local Linear Models, and its Application

Shinichi Imai[*]
Graduate school, Tokyo Gakugei University,
4-1-1, Nukuikita-machi, Koganei, Tokyo, 184-8501, Japan.

E-mail: shimai@u-gakugei.ac.jp
http://www.u-gakugei.ac.jp/[†]

Abstract

In this paper, nonlinear internal model controller based on local linear models, and its Application. The internal model control has a simple structure and has a high robustness for system uncertainties. However, there are few studies of internal model control schemes for nonlinear systems. On the other hand, many controlled systems have the nonlinearity. The effectiveness of the newly proposed control scheme is numerically evaluated on experiment examples in comparison with the conventional control methods for nonlinear systems.

Keywords: Control, . Local Linear Models, Internal Model Controller

1. Introduction

In recent years, with the development of computer technology, design methods for control systems with higher-order compensators have been considered with the aim of improving control performance [1]. In particular, it is difficult to obtain a desirable control response with a simple controller using fixed parameters because the characteristics of the system change significantly due to changes in the system with non-linearity, operating conditions, and changes in the environment.

On the other hand, a model-driven control method [2] has been proposed in which the controlled object is described by the most detailed mathematical formula and the model of the controlled object is incorporated into the control system. Internal model control (IMC) is one of the model-driven control methods [2,3]. IMC is characterized by a simple control system structure and high robust stability against uncertainty of the controlled object. However, there are few examples of applying IMC to nonlinear systems. For example, data-driven IMC that uses a database to extract data similar to the current data as a neighborhood and control it according to the required point. Systems have been proposed, but these methods require enormous amounts of time for computational processing [4].

By the way, the authors have previously proposed a method for calculating control parameters using the concept of the local linear model method [5]. In this method, multiple local linear models can be constructed for a nonlinear system, system parameters corresponding to each local linear model can be obtained, and weights can be applied to design a control system. For nonlinear systems. It is possible to control with fast calculation processing.

Therefore, in this paper, as a method of designing an internal model for a nonlinear system, we propose a method of designing a control system by individually calculating the system parameters corresponding to each local linear model and weighting them. In this method, the system parameters are determined for the local linear model divided into multiple parts, so it can be expected that more appropriate tuning can be performed for the nonlinear system. Since it is not necessary to build the database required by the data-driven IMC system, the time required for construction can be reduced. As a result, the load and processing time can be significantly reduced in terms of memory capacity and calculation time.

2. IMC design using a local linear model

In this paper, we construct a local linear model around the equilibrium point that differs depending on the characteristics of the static characteristics of the nonlinear model. Calculate the control parameters corresponding to each local linear model and obtain the estimated value of the system output. Then, the non-linear control is realized by calculating the distance between the actual system output and the estimated value and adjusting the load of the system parameters according to the distance. Fig.1 shows the block diagram of the proposed control system.

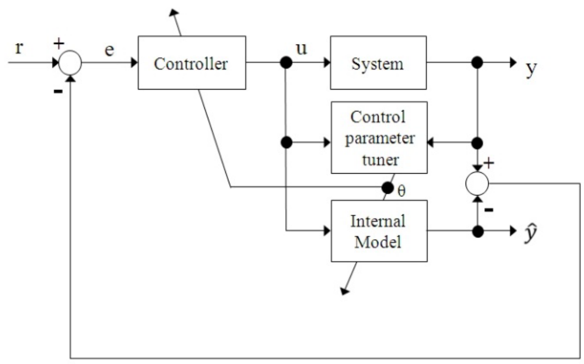


Fig. 1. Block diagram of the proposed method

2.1. *System description*

First, it is assumed that the system dealt with in this paper is given by the following equation.

$$\mathrm{y(t)} = \mathrm{f}(\varphi(\mathrm{t}-1)) \tag{1}$$

where, $\mathrm{y(t)}$ represents the system output and $\mathrm{f}(\cdot)$ represents the nonlinear function. In addition, $\varphi(\mathrm{t}-1)$ represents the state (historical data) before the time $\mathrm{t}-1$ of the system, and is called an information vector. The information vector $\varphi(\mathrm{t}-1)$ is defined by the following equation.

$$\varphi(\mathrm{t}-1) := [\mathrm{y(t-1)},\ \mathrm{y(t-2)}, \cdots, \mathrm{y}(\mathrm{t}-n_y) \\ \mathrm{u(t-1)}, \mathrm{u}(t-2), \cdots, \mathrm{u}(\mathrm{t}-n_u)\ ] \tag{2}$$

Furthermore, $\mathrm{u(t)}$ is the control input, and n_y and n_u are the output and input orders, respectively. Now, suppose that the nonlinear system represented by Eq. (1) can be locally represented by a linear model like the following Eq.

$$\mathrm{A}(z^{-1})\mathrm{y(t)} = z^{-(k_m+1)}B(z^{-1})u(t) \tag{3}$$

where, z^{-1} represents a time-delayed operator that means $z^{-1}y(t) = y(t-1)$. In addition, k_m represents the minimum estimate of wasted time. In many process systems represented by chemical processes, it is often impossible to clearly specify the dead time. Therefore, if the dead time is known, set k_m to that value, and if the range of dead time is unknown, set $k_m = 0$.

In addition, $\mathrm{A}(z^{-1})$ and $\mathrm{B}(z^{-1})$ are given by the following equations.

$$\mathrm{A}(z^{-1}) = 1 + a_1 z^{-1} + \cdots + a_{n_y} z^{-n_y} \tag{4}$$

$$\mathrm{B}(z^{-1}) = b_0 + b_1 z^{-1} + \cdots + b_{n_u} z^{-n_u} \tag{5}$$

where, if the dead time is unknown or ambiguous, the insufficient dead time information is supplemented by securing the order of $\mathrm{B}(z^{-1})$ in n_u dimensions.

2.2. *Designing an internal model controller using a local linear model*

In this paper, we consider the control law (IMC) of the following Eq.

$$\mathrm{u(t)} = \mathrm{Q}(z^{-1})e(t) \tag{6}$$

$$\mathrm{Q}(z^{-1}) = \frac{A(z^{-1})}{B(1)}\left(\frac{1-\lambda}{1-\lambda z^{-1}}\right)^n \tag{7}$$

where, λ is the design parameter of the filter used in the range of $0 \leq \lambda < 1$, and n is the order of the filter. In addition, by using $\mathrm{B}(1)$, even if $\mathrm{B}(z^{-1})$ contains an unstable zero point (non-minimum phase system), it can be dealt with because pole-zero cancellation is avoided.

In addition, $\mathrm{e(t)}$ is a control error signal and can be defined as follows with $\mathrm{r(t)}$ as the target value.

$$\mathrm{e(t)} := \mathrm{r(t)} - \{\mathrm{y(t)} - \hat{y}(\mathrm{t})\} \tag{8}$$

where, $\hat{y}(t)$ is the internal model output and is shown by the following Eq.

$$\hat{y}(t) = -A(z^{-1})y(t) + z^{-(k_m+1)}B(z^{-1})u(t) \tag{9}$$

Since many real systems have non-linearity, it is difficult to always obtain good control results when the system parameters are fixed. Therefore, in this method, the system parameters included in equations (4) and (5) are self-adjusted according to the characteristics of the system based on the local linear model, so they are replaced as follows.

$$\widehat{A}(z^{-1}:t) = 1 + \hat{a}_1(t)z^{-1} + \cdots + \hat{a}_{n_y}(t)z^{-n_y} \quad (10)$$

$$\widehat{B}(z^{-1}:t) = \hat{b}_0(t) + \hat{b}_1(t)z^{-1} + \cdots + \hat{b}_{n_u}(t)z^{-n_u} \quad (11)$$

where, $\widehat{A}(z^{-1}:t)$ and $\widehat{B}(z^{-1}:t)$ represent the control target at time t, and the above assumptions are inherited. Also, n_y is the order of the output. Along with this, equations (7) and (9) are described as the following Eq.

$$Q(z^{-1}:t) = \frac{\hat{A}(z^{-1}:t)}{\hat{B}(1:t)}\left(\frac{1-\lambda}{1-\lambda z^{-1}}\right)^n \quad (12)$$

$$\hat{y}(t) = -A(z^{-1}:t)y(t) + z^{-(k_m+1)}B(z^{-1}:t)u(t) \quad (13)$$

After the above preparations, design an internal model controller using a local linear model. The specific algorithm is summarized below.

[STEP1] Construction of multiple linear models

For the nonlinear model, multiple linear models are constructed, the system is identified by the collective least squares method, and the parameters of $A(z^{-1})$ and $B(z^{-1})$ included in the linear model of the following equation are estimated.

$$A_i(z^{-1})y(t) = z^{-(k_m+1)}B_i(z^{-1})u(t) \quad (i = 1,2,\cdots,N) \quad (14)$$

where, N represents the number of divisions of the local linear model, and $i = 1,2,\cdots,N$ or less Unless otherwise specified, i takes these values. In addition, $A_i(z^{-1})$ and $B_i(z^{-1})$ are given by the following equations.

$$A_i(z^{-1}) = 1 + a_{i,1}z^{-1} + \cdots + a_{i,n_y}z^{-n_y} \quad (15)$$

$$B_i(z^{-1}) = b_{i,0,} + b_{i,1}z^{-1} + \cdots + b_{i,n_u}z^{-n_u} \quad (16)$$

[STEP2] Weight calculation

Next, for each local linear data calculated in [STEP1], the estimation error $\epsilon_i(t)$ is calculated for each model, and the weight ω_i is calculated based on this. $\epsilon_i(t)$ is the error between the system output value y(t) and the estimated output value $\tilde{y}_i(t)$ of each linear model. Here, $\tilde{y}_i(t)$ is calculated by the following equation based on equation (14).

$$\tilde{y}_i(t) = -A_i(z^{-1})y(t) + z^{-(k_m+1)}B_i(z^{-1})u(t) \quad (17)$$

where, $A_i(z^{-1})$ and $B_i(z^{-1})$ use the system parameters of each linear model estimated in [STEP1].

$$\epsilon_i(t) = |y(t) - \hat{y}_i(t)| \quad (18)$$

$$\omega_i(t) = \frac{1/\epsilon_i(t)}{\sum_{i=1}^{N} 1/\epsilon_i(t)} \quad (19)$$

Fig. 2. Process system

Furthermore, $\omega_i(t)$ is the weight corresponding to the selected i-th information vector. The smaller the difference between the output value of the actual system and each linear model, the larger this weight becomes. Note that the following equation is satisfied when $\omega_i(t)$ is calculated based on equation (19).

$$\sum_{i=1}^{N} \omega_i(t) = 1 \quad (20)$$

[STEP3] Determining system parameters

Using the weights obtained in [STEP2] and$A_i(z^{-1})$ and $B_i(z^{-1})$ in Eqs. (15) and (16), the system parameters are calculated by the following equation.

$$\hat{A}(z^{-1}:t) = \sum_{i=1}^{N} \omega_i A_i(z^{-1}) \quad (21)$$

$$\hat{B}(z^{-1}:t) = \sum_{i=1}^{N} \omega_i B_i(z^{-1}) \quad (22)$$

With this system parameter, equations (10) and (11) are updated to obtain the output $\hat{y}(t)$ of the local linear model.

3. Experimental example

The effectiveness of the proposed method will be examined through application to the thermal process system shown in Fig.2. This system uses an incandescent light bulb (40W) as the control target, and controls the surface temperature of the light bulb by changing the voltage applied to the light bulb by controlling the Joule heat of the filament. A heat transfer pair (R52-CA10AE) sensor is attached to the top of the light bulb. Also, measure the temperature of the light bulb with a thermocouple. Furthermore, the temperature of the thermocouple is converted into a voltage by the thermocouple conversion IC, and after A/D conversion,

the data is output to the computer. The control input is calculated using the output data.
A PWM signal with a duty ratio according to the control input is output through D/A conversion, and a current flows through the heater by a solid state relay (SSR).
Therefore, the control input u(t) in this experiment is the duty ratio (0 to 100%) of the PWM signal given to the SSR, and the control output y(t) is the temperature of the surface of the light bulb.
First, the target value r(t) is given as follows.

$$\mathrm{r(t)} = \begin{cases} 50(0 \le t < 50) \\ 70(50 \le t < 100) \\ 100(100 \le t < 150) \\ 130(150 \le t < 200) \end{cases} \tag{23}$$

Next, a local linear model is constructed in the control input range shown below. The number of divisions was N=2.

$$\begin{cases} 0 \le u_1 < 30 \\ 20 \le u_2 < 100 \end{cases} \tag{24}$$

Here, the input / output data in the range of u is saved as the initial database. In Eq. (24), there is a place where the area of u overlaps, but this avoids that a good response cannot be obtained by selecting the database when the request point is selected near the division of each database. It is provided for this purpose. The values of various design parameters included in the proposed method are $n_y = 2, n_y = 2$, and $k_m = 0$. Furthermore, the parameters of IMC are $\lambda = 0.5$ and $\mathrm{n} = 1$. Here, λ was designed to have the desired rising characteristics.
First, for comparison with the conventional method, the fixed PID control method widely used in the industry is applied. However, for the PID parameter, the value calculated based on the CHR method is used. Its PID parameters are shown below.

$$K_P = 3.66, K_I = 0.38, K_D = 5.98 \tag{25}$$

First, Fig.3 shows the control results of the fixed PID method and the control results of the proposed method. In addition, Fig.4 shows the temporal change of the weight by the proposed method in this case. From the results in Fig.3 and Fig.4, it can be seen that the weight of the proposed method changes according to the characteristics of the system, and the responsiveness is greatly improved.

4. Conclusion

In this paper, we proposed a new design method for an IMC system that constructs multiple local linear models for a nonlinear system and adjusts the system parameters corresponding to them. As a result, it was verified through experiments that the system parameters were adjusted appropriately according to the characteristics of the system and good control results could be obtained. In the future, we plan to study the method of dividing the local linear model of this method.

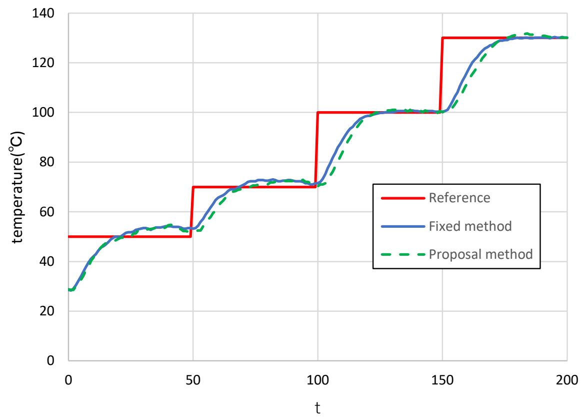


Fig. 3. Experimental result

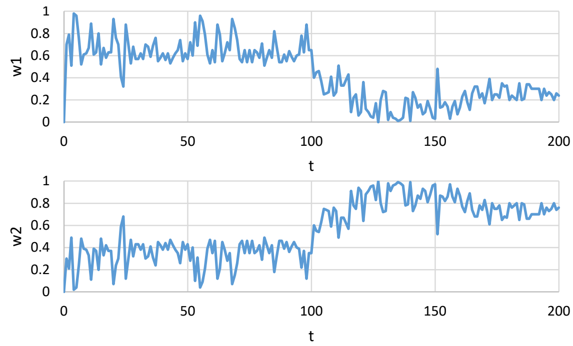


Fig. 4. Change in weight

References

1. Y.Ashida, S.Wakitani and T.Yamamoto : Design of a Data-Driven Multi-loop Self-Tuning PID Controller, IEEJ, Vol.139-C, No.4, 2019, pp.356-36.
2. H.Kimura: Is the model a good controller?, 39th IEEE Conference on Decision and Control, Plenary Lecture I , 2000.
3. T.Shigemasa, M.Yukitomo and R.Kuwata: A model-driven PID control system and its case studies, Proc.CCA, 2002, pp.571-576.
4. K.Tsutsumi and T.Yamamoto: Design of a Data-Driven Internal Model Controller, IEEJ, Vol.131-C, No.4, 2011, pp.781-787.
5. S.Imai and T.Yamamoto: "Design of a Self-Tuning Controller using Multiple Local Linear Models for Nonlinear Systems," IEEJ,Vol.132-C, No.6, 2012, pp860-865

Design of a Data-Driven Control System for Reference Model Design using Predicted Signals

Yuki Nakatani*, **Takuya Kinoshita, Toru Yamamoto**
Graduate School of Advanced Science and Engineering, Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima
Higashi-Hiroshima-shi, Hiroshima-ken, 739-8527, Japan†
E-mail: {nakatani-yuki, kinoshita-takuya, yama}hiroshima-u.ac.jp
www.hiroshima-u.ac.jp

Abstract

In recent years, data-driven control schemes that do not require system identification have been actively studied. Generally, it is easy to give a reference model focusing only on the output response. In contrast, it is difficult to give a reference trajectory considering the input signals based on the controlled system's characteristics. Furthermore, it is necessary to consider the output signal and the input signal since there is a limit of the actuator performance in the control design of the actual machine. This paper proposes a data-driven control scheme that can predict the input/output response of an unknown system in offline using operating data. The effectiveness of the proposed scheme is numerically verified by a simulation example.

Keywords: Data-driven control, extended output, predicted data, off-line tuning.

1. Introduction

In industrial systems, data-driven control methods[1, 2] have been actively studied to achieve the desired control performance for the controlled system with unknown structure and parameters. These methods use only a set of experimental data to design a controller that realizes the desired reference output offline. However, the control system design should consider the input/output response because there is a limit of the actuator performance of the actual machine.

The ERIT (Estimated Response Iterative Tuning) method[3] has been proposed to predict the input/output response. However, this method can only be applied to two-degree-of-freedom (2DOF) control systems[4]. On the other hand, it is useful to design a controller to consider input signals in one-degree-of-freedom (1DOF) because there are also many 1DOF control systems in the industry.

Therefore, in this paper, the new data-driven control scheme is proposed to design a 1DOF controller considering input signals for unknown structure system. The features of this scheme are as follows.

(i) 1DOF control system design that predicts input signals in advance, even for unknown controlled system

(ii) Adjusting the desired input/output response with one parameter

2. Overview of the proposed scheme

Fig. 1 shows an overview of the proposed data-driven control scheme. In the proposed scheme, the controller is designed by the following procedure.

(I) Obtaining the desired predicted data $\hat{y}^*(t)$ and $\hat{u}^*(t)$

(II) Constructing the reference model $G_m(z^{-1})$ based on $\hat{y}^*(t)$ and $\hat{u}^*(t)$ obtained in (I)

(III) Designing a controller using the reference model $G_m(z^{-1})$ obtained in (II)

The detailed procedure is explained in the next section.

3. Design of the controller

3.1. *Generating predicted the data using the ERIT method*

In this section, the ERIT method[3] is described to generate the predicted data. This method is theoretically only applicable to 2DOF control system[4]. Therefore, in the proposed scheme, the ERIT method is utilized for only predicting the input/output data in Fig. 2. Note that the controller is not designed by the ERIT method.

First, the initial output $y_0(t)$ is obtained in a 1DOF control system. In the case of the feed-forward controller $C_{FF}(z^{-1}) = 0$ in Fig. 2, $y_0(t)$ is given as follows:

$$y_0(t) = \frac{G(z^{-1})C_{FB}(z^{-1})}{1 + G(z^{-1})C_{FB}(z^{-1})} r(t). \quad (1)$$

Then, extending to a 2DOF control system for predicting the data, the predicted output $\hat{y}(t)$ is defined as:

$$\hat{y}(t) = \frac{G(z^{-1})C_{FF}(z^{-1}) + G(z^{-1})C_{FB}(z^{-1})}{1 + G(z^{-1})C_{FB}(z^{-1})} r(t). \quad (2)$$

However, system identification is required to obtain an exactly predicted output $\hat{y}(t)$ because Eq. (2) contains a controlled system $G(z^{-1})$. Therefore, substituting Eq. (1) into Eq. (2) yields the following equation:

$$\hat{y}(t) = \frac{C_{FF}(z^{-1})}{C_{FB}(z^{-1})} y_0(t) + y_0(t). \quad (3)$$

Consequently, the predicted data $\hat{y}(t)$ can be derived offline without $G(z^{-1})$ from the initial data $y_0(t)$ by changing $C_{FF}(z^{-1})$. $\hat{u}(t)$ can also be derived in the same procedure using $u_0(t)$ as follows:

$$\hat{u}(t) = \frac{C_{FF}(z^{-1})}{C_{FB}(z^{-1})} u_0(t) + u_0(t). \quad (4)$$

From Eqs. (3) and (4), it is possible to obtain the predicted data $\hat{y}(t)$ and $\hat{u}(t)$ offline corresponding to the various $C_{FF}(z^{-1})$ by using the initial data $y_0(t)$ and $u_0(t)$. Note that the 2DOF control system was only implemented virtually off-line to obtain the predicted data.

3.2. *Obtaining the desired predicted input/output data*

In this section, obtaining the desired predicted data $\hat{y}^*(t)$ and $\hat{u}^*(t)$ is described by using the predicted data $\hat{y}(t)$ and $\hat{u}(t)$. Specifically, $\hat{y}^*(t)$ and $\hat{u}^*(t)$ are calculated by solving the following minimization

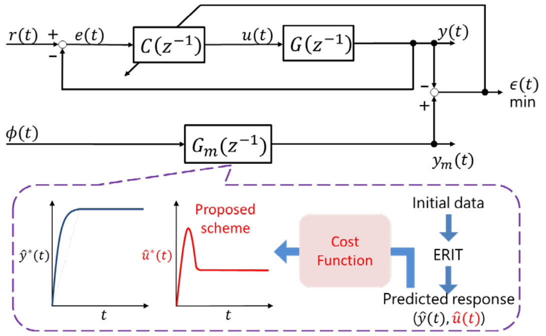


Fig. 1. Overview of the data-driven control system by the proposed scheme.

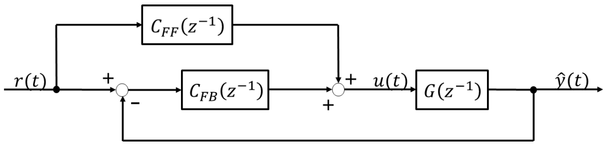


Fig. 2. 2DOF control system.

problem of the evaluation norm based on the ITAE (Integral of Time squared Absolute Error) method[5] for $C_{FF}(z^{-1})$:

$$J_{des} = \sum_{t=0}^{N} t|r(t) - \hat{y}(t)| + \lambda \sum_{t=0}^{N} t|\Delta\hat{u}(t)|, \quad (5)$$

where $r(t)$ and $\Delta(:= 1 - z^{-1})$ denote the reference signal and the difference operator.

Firstly, $r(t) - \hat{y}(t)$ in Eq. (5) shows the difference between the reference signal and the predicted output signal. The desired predicted response is obtained by reducing it. Secondly, $\Delta\hat{u}(t) = \hat{u}(t) - \hat{u}(t-1)$ in Eq. (5) represents the difference of input signal. Therefore, a small value of λ results in a highly responsive response, while a large value of λ results in a less responsive response by depending on the tunable parameter λ. Thus, the input/output response can be easily adjusted with λ. Additionally, it is easy for the user to select the optimal $\hat{y}^*(t)$ and $\hat{u}^*(t)$ by changing λ because predicted data of unknown system can be calculated in the ERIT method.

The adjustment of λ is currently a trial and error process. However, this process is very easy because it is only necessary to adjust λ offline.

3.3. *Constructing $G_m(z^{-1})$ based on the desired predicted data $\hat{y}^*(t)$*

In this section, constructing $G_m(z^{-1})$ is described based on the desired predicted data $\hat{y}^*(t)$. The evaluation function for constructing $G_m(z^{-1})$ is defined as follows:

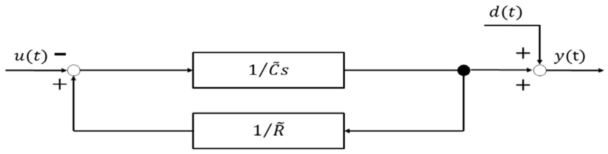


Fig. 3. Block diagram of a temperature control device.

$$J_{ref} = \frac{1}{N}\sum_{t=0}^{N}\left(\tilde{y}^*(t) - y_m(t)\right)^2 \qquad (6)$$

$$y_m(t) = G_m(z^{-1})r(t), \qquad (7)$$

where N denotes the number of data. $G_m(z^{-1})$ is constructed by minimizing the evaluation J_{ref}.

3.4. *Design of a PID controller based on extended output*

In this section, a 1DOF controller $C(z^{-1})$ is designed as an I-PD controller[6] by using the reference model $G_m(z^{-1})$ that considers the input response obtained in the previous section.

I-PD controller is given as follows:

$$\Delta u(t) = K_I e(t) - K_P \Delta y(t) - K_D \Delta^2 y(t), \qquad (8)$$

where K_P, K_I, and K_D are the proportional gain, integral gain, and derivative gain respectively. In this paper, PID gains are tuned by using the extended output. In this method, $C(z^{-1})$ is designed by minimizing the evaluation norm based on the property that the extended output $\phi(t)$ and the reference signal $r(t)$ are equal in Fig. 1. In other words, the evaluation norm is defined as:

$$J = \frac{1}{N}\sum_{t=0}^{N}\epsilon(t)^2 \qquad (9)$$

$$\epsilon(t) = G_m(z^{-1})\phi(t) - y(t). \qquad (10)$$

where $\epsilon(t)$ denotes the difference between the output $y(t)$ of the closed-loop transfer function and the output $G_m(z^{-1})\phi(t)$ of the reference. $C(z^{-1})$ can be designed by minimizing the evaluation J. For reasons of space, the details are omitted, refer to the PID control method based on extended output[7].

4. Numerical example

4.1. *Controlled system*

The effectiveness of the proposed scheme is verified using the numerical simulation. The controlled system is the thermal experimental apparatus for simulating bag-and-bound welding owned by our laboratory. Fig. 3 shows the model of the system, and the coefficients were derived by system identification. The details are omitted due to space limitation.

Table 1. Obtained parameters in the simulation.

	$\lambda = 1$	$\lambda = 35$
α_1	2.23×10^{-2}	1.71×10^{-4}
α_2	1	1
α_3	-2.35×10^{-2}	-1.57×10^{-2}
α_4	-4.07×10^{-2}	-9.75×10^{-1}
θ_1	2.23×10^{-4}	1.72×10^{-6}
θ_2	8.92×10^{-4}	6.89×10^{-6}
θ_3	1	1
θ_4	-1.42	-1.99
θ_5	4.20×10^{-1}	9.93×10^{-1}
K_P	4.24×10^{2}	4.27
K_I	8.16×10^{-1}	5.60×10^{-3}

Then, the proposed scheme is applied to the system in Fig. 3. The reference signal is $r(t) = 100$, the sampling time is $T_S = 0.02$ s, and the room temperature is $d(t) = 20$ [°C]. In this numerical example, I-P controller is used, and the proportional gain and integral gain are set as follows:

$$K_P = 1.5, K_I = 1.0 \times 10^{-3}. \qquad (11)$$

Next, the feed-forward controller $C_{FF}^*(z^{-1})$ was calculated by the proposed scheme as follows:

$$C_{FF}^*(z^{-1}) = \frac{\alpha_1}{\alpha_2 + \alpha_3 z^{-1} + \alpha_4 z^{-2}}. \qquad (12)$$

The order of the denominator was set to second order because the order which significantly affects the output response of controlled system is generally low order. Furthermore, the value of α_i $(i = 1, 2, 3, 4)$ is determined to minimize J_{des} in Eq. (5), for example, using the Matlab/Simulink Ver.9.8.0.1396136(R2020a), Optimization Toolbox 'fminsearch.m '.

Furthermore, $G_m(z^{-1})$ was determined as follows:

$$G_m(z^{-1}) = \frac{\theta_1 + \theta_2 z^{-1}}{\theta_3 + \theta_4 z^{-1} + \theta_5 z^{-2}}, \qquad (13)$$

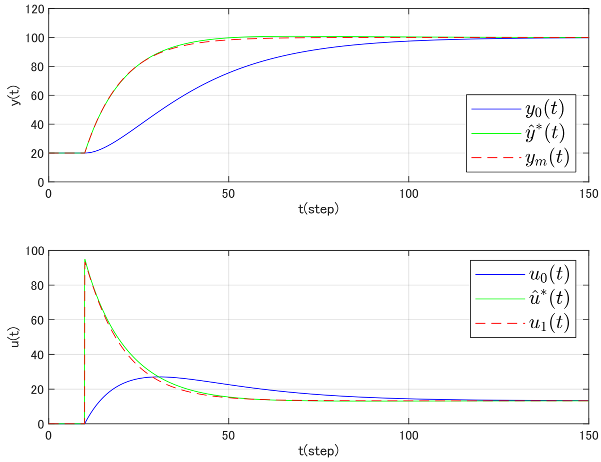


Fig. 4. Simulation results by applying the proposed scheme where $\lambda = 1$.

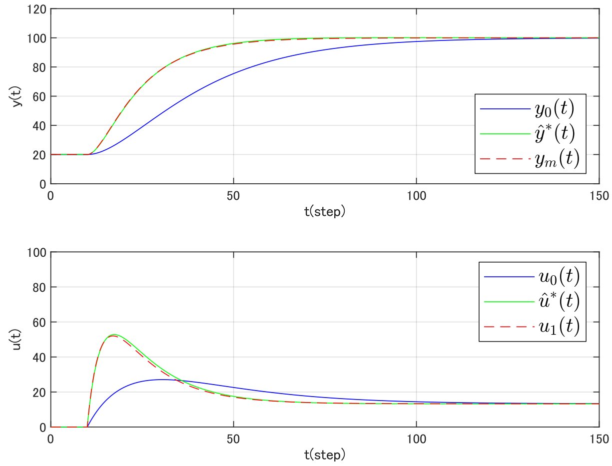


Fig. 5. Simulation results by applying the proposed scheme where $\lambda = 35$.

where the value of θ_i $(i = 1, 2, 3, 4, 5)$ is determined by using the least-squares method to minimize J_{ref} in Eq. (6).

In this numerical example, the results are compared by changing λ in Eq. (5), and Table 1 shows the feed-forward controller $C_{FF}^*(z^{-1})$, the reference model $G_m(z^{-1})$, the proportional gain, and the integral gain respectively in the cases of $\lambda = 1, 35$.

4.2. *Control result*

Figs. 4 and 5 show the initial data $y_0(t)$ and $u_0(t)$, the desired predicted data $\hat{y}^*(t)$ and $\hat{u}^*(t)$, and the results of implementing a 1DOF controller $y_1(t)$ and $u_1(t)$ in the cases of $\lambda = 1, 35$. The input-output response can be adjusted with one parameter according to λ because the responsiveness is high because $\lambda = 1$ in Fig. 4, and the responsiveness is low because $\lambda = 35$ in Fig. 5. Furthermore, it is possible to design a controller that considers input signals because the input signals $\hat{u}^*(t)$ and $u_1(t)$ are almost identical in Figs. 4 and 5.

5. Conclusion

In this paper, the new data-driven control scheme has been proposed to design a 1DOF to consider input signal for unknown structure system. The effectiveness has been verified by the numerical example. It is possible to predict the appropriate amount of input signal to correspond the performance of the actuator by predicting the input in advance. In the future, the method how to choose λ and the controller design considering the input saturation will be studied.

References

1. T. Yamamoto, O. Kaneko, S. Wakitani, T. Kinoshita, Y. Ohnishi, H. Kugemoto, K. Koiwai, Data-oriented PID control, Morikita publish, 2020.
2. T. Yamamoto, O. Kaneko, A Direct Design Method for Control Systems Based on Closed-Loop Data, Journal of the SICE **52**(10), 2013, pp. 841-846
3. O, Kaneko, H. Nakamura, T. Ikezaki, A New Approach to Updating Feedforward Controllers in Two-Degree-of-Freedom Control Systems Estimated Response Iterative Tuning (ERIT) is proposed, Transactions of the SICE **54**(12), 2018, pp. 857-864.
4. Y. Chida, Design Method of Two-Degree-of-Freedom Control System Using Characteristics of Feedback Controller, Transactions of the Institute of Electrical Engineers of Japan **117**(4), 1997, pp.439-445
5. M. Masubuchi, Revised Basic Theory of Automatic Control, Corona Corporation, 2005, pp. 181-184
6. T. Kitamori, Principles and Design Methods of I-PD Control Systems (Special Issue on PID Control - Theory and Application), Transactions of SICE **54**(12), 1998, pp. 7-17
7. A, Yoichiro, S. Wakitani and T. Yamamoto, Design of an Augmented Output-Based Multiloop Self-Tuning PID Control System, Industrial & Engineering Chemistry Research **58**(26), 2019, pp. 11474-11484

Design of a Database-Driven GMV Controller Using the Nelder-Mead Method*

Liying Shi
Graduate School of Engineering, Hiroshima University, Japan
1-4-1 Kagamiyama HigashiHiroshima ,739-8527, Japan
E-mail: m192383@hiroshima-u.ac.jp

Zhe Guan
KOBELCO Construction Machinery Dream-Driven Co-Creation Research Center, Hiroshima University, Japan
1-4-1 Kagamiyama HigashiHiroshima ,739-8527, Japan
E-mail: guanzhe@hiroshima-u.ac.jp

Toru Yamamoto
Graduate School of Advanced Science and Engineering, Hiroshima, Japan
1-4-1 Kagamiyama HigashiHiroshima ,739-8527, Japan
E-mail: yama@hiroshima-u.ac.jp

Abstract

This paper presents a design scheme that can obtain the optimized generalized minimum variance (GMV) control parameters by applying the Nelder-Mead (NM) method based on proportional-integral-derivative (PID) controller. The NM method is used to find the most suitable parameter λ. The application of NM method can optimize the calculation of λ. Furthermore, the estimation of closed-loop response method is introduced in database-driven approach. The effectiveness is verified by using a simulation example.

Keywords: PID controller, GMV, Nelder-Mead method, Database-driven approach

1. Introduction

A Proportional-Integral-Derivative (PID) controller[1,2] is a three-term controller that has a long history in the automatic control field, and it is certainly the most widely used control scheme today owing to its intuitiveness and its relative simplicity. Over the last half-century, a great deal of academic and industrial effort has focused on improving PID control. In previous controller, using the minimum variance (MV)[3,4] as the benchmark can claim to be successful. The basic idea behind the MV index is to only consider the controller error variance. However, few of the available techniques in use take the control effort or the manipulative variable activity into account. Then, the GMV [5] is proposed that takes into account controller error variance as well as the manipulating variable variance. Furthermore, an adjustable parameter λ (a weighting factor penalty on the manipulating variable) is included in the GMV. In the previous GMV controller, the PID parameters are calculated by simply changing λ manually. Therefore, it is hard to get desirable control performance. In order to improve the control performance, the Nelder-Mead method[6] is introduced. It can optimize the calculation of the most suitable parameter λ.

In this paper, a new GMV control scheme is proposed in which the controller parameters can be calculated without any model of the process by applying the Nelder-

Mead method[6] and method of the estimation of closed-loop response[7]

The rest of this paper is organized as follows: the problem is formulated in section 2; the proposed scheme is the main topic in section 3 and section 4; the section 5 provides the simulation and analysis, and the section 6 concludes the paper.

2. Overview of the Data-Driven GMV controller

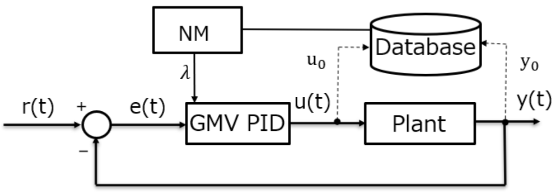


Fig. 1. The Block diagram of database-driven GMV controller.

In Fig. 1, an adjustable parameter λ is included in the GMV, and PID parameters are calculated by simply changing λ. Using the method of estimation of closed-loop response to predict new output, then the application of NM method can optimize the calculation of the most suitable λ according to the predict output and get optimal PID parameters without any model of the process.

3. THE GENERALIZED MINIMUM VARIANCE CONTROL OF PARAMETERS

3.1. *System description*

Consider the following process model and design of PID controller:

$$A(z^{-1})y(t) = z^{-1}B(z^{-1})u(t) + \frac{\xi(t)}{\Delta} \qquad (1)$$

where $A(z^{-1})$ and $B(z^{-1})$ are given by the following polynomials:

$$A(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} \qquad (2)$$

$$\mathrm{B}(\mathrm{z}^{-1}) = \mathrm{b}_0 + \cdots + \mathrm{b}_\mathrm{m}\mathrm{z}^{-\mathrm{m}} \qquad (3)$$

where $u(t)$ and $y(t)$ are respectively the input and the corresponding output; ξ(t) is the Gaussian white noise which has zero mean; z^{-1} , the backshift operator which implies $z^{-1}y(t) = y(t-1)$; Δ, the differencing operator ($\Delta = 1 - z^{-1}$); and m, the order of $B(z^{-1})$.

In this paper, the PID controller is introduced as $\frac{C(z^{-1})}{\Delta}$ in the following equation:

$$u(t) = \frac{C(z^{-1})}{\Delta}\{r(t) - y(t)\} = \frac{C(z^{-1})}{\Delta}e(t) \qquad (4)$$

where $r(t)$ is the reference signal; $e(t)$, the control error.

3.2. *Benchmark of the Generalized Minimum Variance Controller*

The GMV control law for the system (1) can be derived by minimizing the following cost function:

$$J = E[\phi^2(t+k+1)]. \qquad (5)$$

Here, $\phi(t+k+1)$ is the generalized output given by following equation:

$$\phi(t+k+1) = P(z^{-1})y(t+k+1) + \lambda\Delta u(t) - P(1)\omega(t) \qquad (6)$$

where ω(t) denotes the step reference signal, k is the minimum time-lag which is estimated by an operator (k is set as 0 when the time-lag is unknown). In addition, λ is the weight coefficient for the variance of the control input and it is set by a user arbitrarily. Next, the following Diophantine equation is introduced.

$$P(z^{-1}) = \Delta E(z^{-1})A(z^{-1}) + z^{-(1+k)}F(z^{-1}) \qquad (7)$$

$$E(z^{-1}) = 1 + e_1 z^{-1} + \ldots + e_k z^{-k}$$
$$F(z^{-1}) = f_0 + f_1 z^{-1} + f_2 z^{-2} \qquad (8)$$

Moreover, $P(z^{-1})$ is the design polynomial and it is designed based on the following equation. where, the order of $E(z^{-1})$ and $F(z^{-1})$ are set to decide these coefficients uniquely from $\Delta A(z^{-1})$ and $P(z^{-1})$.

$$P(z^{-1}) = 1 + p_1 z^{-1} + p_2 z^{-2} \qquad (9)$$

$$p_1 = -2\exp\left(-\frac{\rho}{2\mu}\right)\cos\frac{\sqrt{4\mu-1}}{2\mu}\rho \qquad (10)$$

$$p_2 = \exp\left(-\frac{\rho}{\mu}\right) \qquad (11)$$

$$\rho := T_s/\sigma \qquad (12)$$

$$\mu := 0{,}25(1-\delta) + 0.51\delta \qquad (13)$$

σ and δ are parameters which express the rise time and the damping property which are set by an operator respectively. Moreover, (σ denotes the rise time that the system output attains about 60% of the value of the step reference signal. From (1), (4) and (5), the predictive value of the generalized output after $k+1$ at t can be obtained as following equation:

$$\phi(t+k+1) = F(z^{-1})y(t) + \{E(z^{-1})B(z^{-1}) + \lambda\}\Delta u(t) - P(1)\omega(t) + E(z^{-1})\xi(t+k+1). \qquad (14)$$

The control law is described as the following equation:

$$\Delta u(t) = \frac{P(1)}{G(z^{-1})+\lambda}\omega(t) - \frac{F(z^{-1})}{G(z^{-1})+\lambda}y(t). \qquad (15)$$

By replacing the polynomial $G(z^{-1})$ in (15) by the steady-state term $G(1)$, the following equation can be obtained:

$$\Delta u(t) = \frac{P(1)}{G(1)+\lambda}\omega(t) - \frac{F(z^{-1})}{G(1)+\lambda}y(t). \qquad (16)$$

Therefore, PID parameters can be calculated as follows:

$$K_P = \frac{f_1+2f_2}{G(1)+\lambda} \tag{17}$$

$$K_I = \frac{f_1+f_2+f_0}{G(1)+\lambda} \tag{18}$$

$$K_D = \frac{f_2}{G(1)+\lambda} \tag{19}$$

In this method, the control performance is strongly depended on λ. Then the user-specified parameters λ is calculated by using the NM method. The predicted output adopts the estimation of closed-loop response method without any model. The detail of this approach will not be given because of pages limitation. Please refer to the Ref. 7.

4. Evolutionary Computation Based on The Nelder-Mead method

The Nelder-Mead method[6] is briefly explained is to calculate the user-specified parameter λ. In this paper, the objective function is determined by

$$H(S_i) = \sum_{t=1}^{M} \{y_m(t) - y_d(t)\}^2 \tag{20}$$

where $y_m(t)$ denotes the reference model output and M denotes the evaluation horizon. $y_d(t)$ is the predicted output corresponding to S_i. The objective function can be determined. The top with the largest target function value $H(S_i)$, the second large top and the smallest top are respectively assumed to be S_H, S_M and S_L. Furthermore, the center between S_M and S_L is determined as S_G. Four operations are briefly explained as follows:

- Reflection

  S_R is determined as $S_R = (1+\beta)S_G - \beta S_H$, where β is set as $\beta > 0$, and it corresponds to the ratio of distances $S_R S_G$ and $S_H\ S_G$.

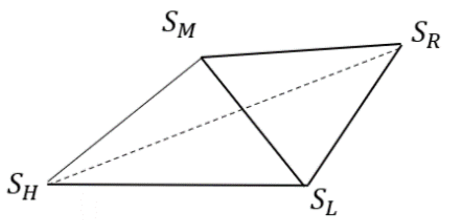


Fig. 2. Reflection of the Nelder-Mead method

- Expansion

  S_E is determined as $S_E = (1-\gamma)S_G + \gamma S_H$, where γ is set as $\gamma > 1$, and it corresponds to the ratio of distances $S_E S_G$ and $S_R\ S_G$.

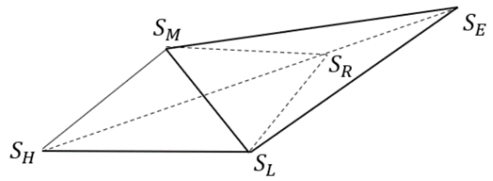


Fig. 3. Expansion of the Nelder-Mead method

- Contraction

  S_C is determined as $S_C = (1-\eta)S_G + \eta S_H$, where η is set as$1 > \eta > 0$, and it corresponds to the ratio of distances $S_C S_G$ and $S_H\ S_G$.

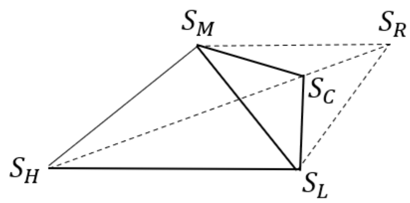


Fig. 4. the contraction of the Nelder-Mead method

- Reduction

  S_H and S_M are moved in the direction of S_L.

The proposed control system can be designed by calculate the most suitable λ by the step based on the mentioned procedure.

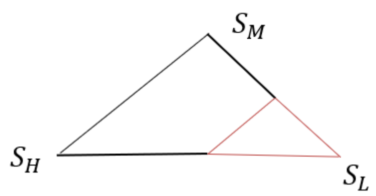


Fig. 5. Reduction of the Nelder-Mead method

The proposed control system can be designed by calculating the most suitable λ by the step based on the mentioned procedure.

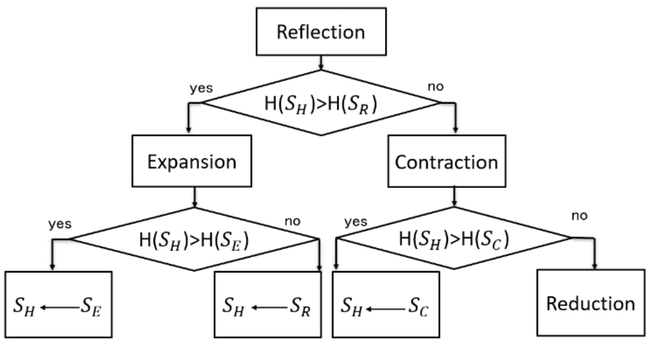


Fig.6. Procedure of the Nelder-Mead method

If $H(S_H) > H(S_R)$, accept expansion, and then if $H(S_H) > H(S_E)$, S_E replace S_H. If $H(S_H) < H(S_E)$, S_R replace S_H. If $H(S_H) < H(S_R)$, accept contraction, and then if $H(S_H) > H(S_C)$, S_C replace S_H. If not, accept reduction.

5. Simulation Example

The effectiveness of the newly proposed design scheme is evaluated on a numerical simulation example.

$$y(t) = -1.575y(t-1) - 0.654y(t-2) + 0.023u(t-1) - 0.019u(t-2) + \frac{\xi(t)}{\Delta} \quad (21)$$

Firstly, Fig. 7 shows the control result by using manually adjusted λ.
when $\lambda = 0.09$

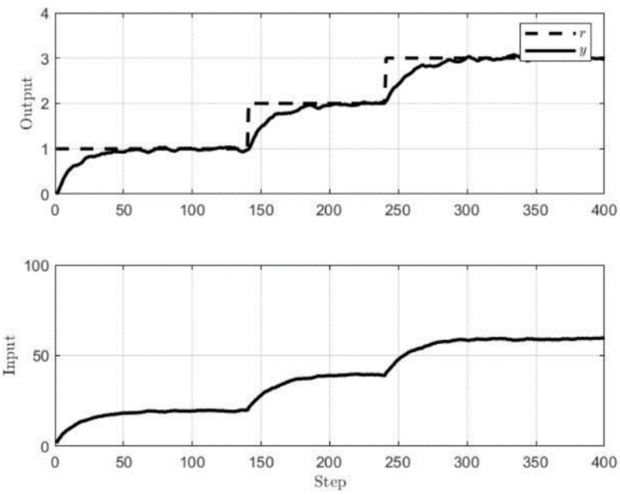


Fig. 7. Control result by using manually adjusted λ.

From the result, using manually adjusted λ can not get desirable control performance. Then Fig.8 shows the control result by using the newly proposed method, where the design parameters of Nelder-Mead method are set as shown in Table 1.

Table 1. Parameters used in Nelder-Mead method.

S_{max}	S_{min}	β	γ	η
(1.0,1.0)	(0.0,0.0)	0.1	1.1	0.1

Using the NM method, $\lambda = 0.05$ is the most suitable parameter. Comparing with the two kinds of control results, it is clear that the newly proposed control scheme works well.

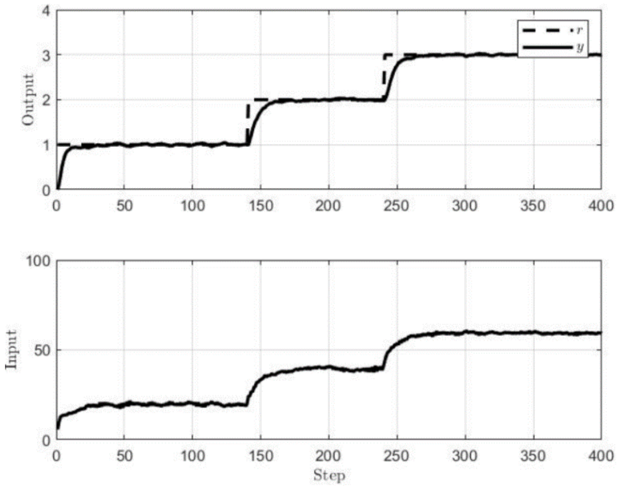


Fig. 8.the control result by using calculated λ.

6. Conclusions

In this paper, a design scheme of database-driven GMV controller has been proposed, which can obtain GMV control parameters by applying the NM method based on PID controller. The features of the newly proposed control scheme are summarized as follows:

- PID parameters are adjusted based on the GMV control system.
- The user-specified parameter included in GMV control system, is calculated by NM method for liner system.
- The estimation of closed-loop response is introduced in database-driven approach.

The effectiveness of the proposed scheme has been evaluated on the simulation example.

7. References

1. Michael A. J, Mohamman H. M. PID Control. *Springer-Verlag*, 2005.
2. Antonio V. Practical PID control. *Springer-Verlag*, 2006.
3. Harris, T.J. Assessment of closed loop performance. *Can. J. Chem. Eng.* 1989, pp.856-861.
4. S. Wakitani, S. Hanata, T. Yamamoto. Design and Application of a Direct GMV-Based PID Controller. *Proceedings of the 2012 International Conference on Advanced Mechatronic Systems*, 2012.
5. Astrom K. J; Hagglund T. The future of PID control. *Control Engineering Practice* 2001, pp.1163-1175.
6. Nelder J. A, Mead R. A Simplex Method for Function Minimization. *Computer J.* 1965, pp.100-313.
7. T. Sakatoku, K. Yubai, D. Yashiro, S. Komada. Estimation of Closed-Loop Response and Controller Retuning. *IEEJapan*, 2020.

Design of a Database-Driven GPC for Nonlinear Systems

Zhe Guan
KOBELCO Construction Machinery Dream-Driven Co-Creation Research Center, Hiroshima University
Higashi-Hiroshima, 739-8527, Japan
E-mail: guanzhe@hiroshima-u.ac.jp,

Toru Yamamoto
Graduate School of Advanced Science and Engineering, Hiroshima University
Higashi-Hiroshima, 739-8527, Japan
E-mail: yama@hiroshima-u.ac.jp

Abstract

This paper presents a Generalized Predictive Control (GPC) design scheme based on database-driven approach for nonlinear systems. In industrial systems, lots of controlled plants with unknown time-delay and strong nonlinearity, are difficult to be handled in terms of control performance. The databased-driven approach has been attracted attention to tackle non-linearity issue without system information. On the other hand, GPC is considered in cases where the time-delay is unknown. The proposed scheme is validated through a numerical simulation.

Keywords: GPC, Database-Driven approach, Nonlinear Systems

1. Introduction

Industrial process control is a large and diverse field, and its broad range of applications require a wide range of controllers including model predictive control (MPC) and a variety of nonlinear controllers in order to maintain system performance. Among them, the Generalized Predictive Control (GPC) algorithm[1] can adapt to process time-delay and model order based on the long-range prediction. As a result, it is shown to be particularly effective for the self-tuning control of industrial processes[2], in which are widely applied. However, regarding to the GPC, the excellent control behavior can be achieved, provided that the number of parameters in the model are required. It is important to include a reasonable representation of disturbances and other unknown dynamics, such that the correct controller can be designed. In other words, the GPC is model-based technique, so good performance needs a high-quality process model. Unfortunately, as it is well known, modeling for an industrial process using first principles or identification is a difficult task. Further, even the mathematical model of the industrial process is established, it only approximates the process but not involves entire process dynamics. Moreover, it is time- and cost-consuming to build the relative accurate complicate model of industrial process. Therefore, some challenges have been addressed in the model-based technique.

With the rapid development of the computer technology, industrial processes have undergone significant changes. It is possible to store and process huge amounts of process data with the help of computers. In recent years, it is a trend to design the controller by directly using online or offline Input/Output (I/O) data obtained from the controlled system or knowledge, instead of the explicit mathematical model of the controlled process[3]. Several methods[4,5] have been proposed by considering the above idea. Moreover, the data-oriented GPC controller[6] has been proposed, where the initial control parameters are estimated from the I/O data, and then the corresponding polynomials are calculated. However, two issues are still not addressed. The aforementioned method is not able to deal with some systems with nonlinearity. The other issue is that the tuning performs in off-line manner.

The database-driven approach[7] is one of control strategies applied in systems with non-linearity since it can tune the control parameters rapidly at each equilibrium point. Furthermore, the database-driven approach performs the tuning in on-line manner in order to deal with the operation condition along the time, such that the performance deterioration is able to be avoided. Successful implementations of the data-driven approach have been presented in the technical literatures[8,9,10].

Based on the above observation, this paper newly presents a design scheme that the controller is designed with GPC algorithm based on database-driven approach. The proposed scheme inherits the advantages of GPC algorithm in terms of handling delay. Moreover, the control parameters are computed on-line in a local bounded neighbor data which is extracted from the database.

2. Problem Statement

Consider the following nonlinear system:

$$y(t) = f(\phi(t-1)) + \xi(t) \tag{1}$$

where $y(t)$, $f(\cdot)$ and $\xi(t)$ denote the system output, the nonlinear function and the white Gaussian noise with zeros mean and variance σ^2, respectively. Additionally, $\phi(t-1)$ is called information vector, which is defined as follows:

$$\begin{aligned}\phi(t-1) :=&[y(t-1), \cdots, y(t-n_y),\\ &u(t-d-1), \cdots, u(t-d-n_u)],\end{aligned} \tag{2}$$

where $u(t)$ and d represent the control input and system delay time. Also, n_y and n_u are the order of the system output and control input, respectively.

When considering regulation about a particular operation point, it is reasonable to assume that the nonlinear system can be described as a locally-linearized model by the following equation:

$$A(z^{-1})y(t) = z^{-(d+1)}B(z^{-1})u(t) + \frac{\xi(t)}{\Delta}, \tag{3}$$

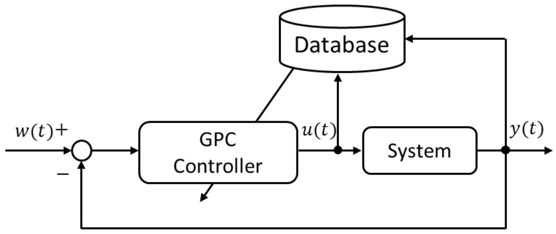


Fig. 1. Block diagram of the proposed scheme.

where z^{-1} represents the backward shift operator which implies $z^{-1}y(t) = y(t-1)$, and Δ denotes the differential operator defined by $\Delta := 1 - z^{-1}$. $A(z^{-1})$ and $B(z^{-1})$ are the polynomials which are formulated as follows.

$$A(z^{-1}) = 1 + a_1 z^{-1} + \cdots + a_{n_y} z^{-n_y}, \tag{4}$$

$$B(z^{-1}) = b_0 + b_1 z^{-1} + \cdots + a_{n_u} z^{-n_u}. \tag{5}$$

Based on the above statement, the objective of this study is to design a GPC controller that generates $u(t)$ to controlling the system Eq. (1) based on database-driven approach such that the output can track the reference value. As a result, the block diagram of the proposed scheme is depicted in Fig. 1.

3. Proposed Scheme Design

3.1. *Generalized Predictive Control*

The GPC law is derived based on the minimization of the criterion in the following equation:

$$\begin{aligned} J \;=\; E\Bigg[&\sum_{j=1}^{N}\{y(t+d+j) - w(t+d+j)\}^2 \\ &+\sum_{j=1}^{N}\lambda(j)\{\Delta u(t+j-1)\}^2\Bigg], \end{aligned} \tag{6}$$

where $E(\cdot)$ is the expectation value and $[1, \mathrm{N}]$ denotes the range of control horizon and prediction horizon. $w(t)$ is the reference value and $\lambda(j)$ is a user-specified control weighting sequence. The control law based on Eq. (6) is derived as:

$$\begin{aligned}&\sum_{j=1}^{N} p_j F_j(z^{-1})y(t+d) + \left\{1 + z^{-1}\sum_{j=1}^{N} p_j S_j(z^{-1})\right\}\Delta u(t) \\ &-\sum_{j=1}^{N} p_j w(t+d+j) = 0,\end{aligned} \tag{7}$$

with $F_j(z^{-1})$ and $G_j(z^{-1})$ are derived by the following Diophantine equations:

$$1 = \Delta A(z^{-1})E_j(z^{-1}) + z^{-j}F_j(z^{-1})$$
$$E_j(z^{-1})B(z^{-1}) = G_j(z^{-1}) + z^{-1}S_j(z^{-1}) \tag{8}$$

where

$$E_j(z^{-1}) = 1 + e_1 z^{-1} + \cdots + e_{j-1}z^{-(j-1)}$$
$$F_j(z^{-1}) = f_0^j + f_1^j z^{-1} + \cdots + f_{n_y}^j z^{-n_y}$$
$$G_j(z^{-1}) = g_0 + g_1 z^{-1} + \cdots + g_{j-1}z^{-(j-1)}$$
$$S_j(z^{-1}) = s_0^j + s_1^j z^{-1} + \cdots + s_{n_u-1}^j z^{-(n_u-1)} \tag{9}$$

Moreover, p_j is formulated by

$$[p_1, p_2, \cdots, p_N] := [1, 0, \cdots, 0][\boldsymbol{G}^T\boldsymbol{G} + \Lambda]^{-1}\boldsymbol{G}^T \tag{10}$$

with the matrix $\boldsymbol{G}$ is composed by

$$\boldsymbol{G} := \begin{bmatrix} g_0 & 0 & & \\ g_1 & g_0 & 0 & \\ \vdots & \vdots & \ddots & \\ g_{N-1} & g_{N-2} & \cdots & g_0 \end{bmatrix} \tag{11}$$

3.2. *Derivation of output by control parameters*

Considering the Diophantine equations Eq. (8) and substitute to system Eq. (3), the system output can be obtained one-step ahead as:

$$\hat{y}(t+1) = F_1 y(t) + G_1 \Delta u(t-d) \tag{12}$$

The above equation is written in the form of information vector as:

$$\hat{y}(t+1) = \phi^T(t)\theta(t) \tag{13}$$

with F_1 and G_1 are the components in $\theta(t)$.
Once the control parameters F_1 and $G_1$are identified, and then by recursive formula, the control signal is able to be derived. Please refer to the Ref. 6.

3.3. *Database-driven approach*

The detail of this approach will not be given because of pages limitation. Please refer to the Ref. 7. We will give core part in terms of control parameters update. As mentioned in Eq. (13), the control parameters will be updated based on database by adopting the steepest descent method.

$$\theta^{new}(t) = \theta^{old}(t) - \eta\frac{\partial J(t+1)}{\partial \theta(t)} \tag{14}$$

with η learning rate, and $J(t+1)$ is defined by:

$$J(t+1) := \frac{1}{2}\varepsilon(t+1)^2 \tag{15}$$

(16)

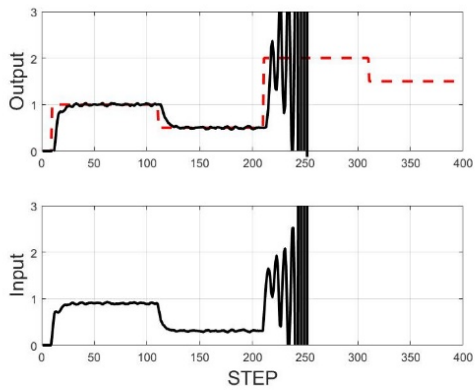


Fig. 2. Control results using the conventional PID control.

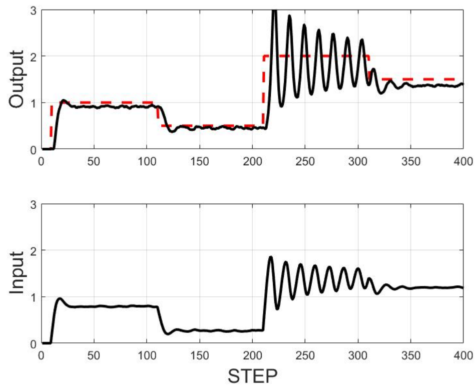


Fig. 3. Control results using the data-oriented GPC.

4. Numerical Simulation

In this section, we have conducted the simulation and illustrate the effectiveness of the proposed scheme. The Hammerstein model[11] is considered and is given in the following form:

$$y(t) = 0.6y(t-1) - 0.1y(t-2)$$
$$+1.2x(t-d-1) - 0.1x(t-d-2) + \xi(t)$$
$$x(t) = u(t) - u^2(t) + 0.5u^3(t) \tag{17}$$

where $\xi(t)$ is the Gaussian noise with zero mean and variance 0.01^2, and delay time d is set as 2 in this study. The reference signals are given by:

$$w(t) = \begin{cases} 1.0(0 \le t < 100) \\ 0.5(100 \le t < 200) \\ 2(200 \le t < 300) \\ 1.5(300 \le t < 400) \end{cases} \tag{18}$$

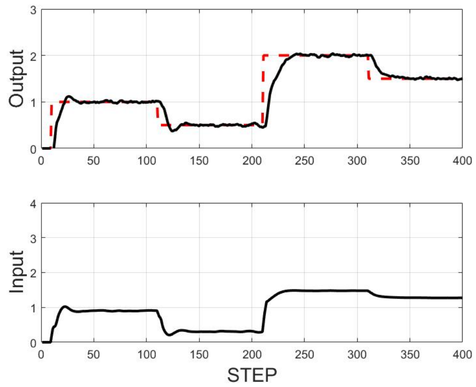


Fig. 4. Control results using the proposed scheme.

Additionally, the user-specified parameters involved in the proposed scheme are determined as shown in Table 1.

Table 1. The user-specified parameters.

The orders of information vector	$n_y = 2$ $n_u = 5$
Numbers of Neighbors	$k = 5$
Weighting factor	$\lambda = 15$
Initial data-base *N(0)*	20
Learning rate	100

The proposed scheme cannot work unless the initial database is generated. Consequently, the conventional Proportional-Integral-Derivative (PID) control tuned by the Chien-Hrones-Reswick method[12] is employed. The related parameters are shown as:

$$K_P = 0.486, K_I = 0.227, K_D = 0.122.$$

The control results are depicted in Fig. 2, where the output is become unstable after 200 [STEP] because of nonlinearity and delay time. These I/O data form the basis of the initial database. Fig. 3 shows the control results where the control parameters F_1 and G_1 in GPC algorithm are identified from the I/O data and are identical in the control process. The tracking property is poor, and oscillation occurs after 200 [STEP]. Finally, the proposed scheme is employed to the system, and Fig. 4 shows that the proposed scheme outperforms the ones in Fig. 3. It should be noted that the control parameters F_1 and G_1 are updated based on database along the time.

5. Conclusions

In this paper, the design of a data-driven GPC control for nonlinear systems has been presented. First, according to the proposed scheme, the model information is not required in designing the controller. Then, the database-driven approach is considered to deal with systems with strong non-linearity and provides a way for the online utilization of the proposed scheme. Furthermore, the proposed scheme inherits the advantage of GPC algorithm such that good performance can be achieved regardless of time-delay. Finally, the effectiveness has been verified on a numerical simulation. The experimental evaluation will be in our future work.

References

1. D. W. Clarke, C. Mohtadi, and P S. Tuffs, Generalized Predictive Control – Part I. The basis algorithm, *Automatica*, **23**(2), 1987, pp. 137-148.
2. D. W. Clarke, Application of generalized predictive control to industrial processes, *IEEE Control Systems Magazine.* **8**(2), 1988, pp. 49-55.
3. Z. S. Hou and S. S. Xiong, On Model-Free Adaptive Control and its Stability Analysis, *IEEE Transactions on Automatic Control*, **64**(11), 2019, pp. 4555-4569.
4. M. C. Campi, A. Lecchini, and S. M. Savaresi, Virtual reference feedback turning: a direct method for the design of feedback controllers, *Automatica*, **38**(8), 2002, pp. 1337-1346.
5. H. Hjalmarsson, M. Gevers, S. Gunnarsson, and O. Lequin, Iterative feedback tuning: theory and applications, *IEEE Control Systems Magazine,* **18**(4), 1998, pp. 26-41.
6. Z. Guan, S. Wakitani, and T. Yamamoto, Design and Experimental Evaluation of a Data-Oriented Generalized Predictive PID Controller, *Journal of Robotics and Mechatronics*, **28**(5), 2016, pp. 722-729.
7. T. Yamamoto, K. Takao, and T. Yamada, Design of a Data-Driven PID control, *IEEE Transactions on Control Systems Technology,* **17**(1), 2009, pp. 29-39.
8. K. Takao, T. Yamamoto, and T. Hinamoto, A new GPC-based PID controller using memory-based identification, *Proceedings of the 11th IFAC International Workshop on Adaptation and Learning in Control and Signal Processing,* 2013, pp. 587-592.
9. S. Wakitani, M. C. Deng, and T. Yamamoto, Design of a data-driven controller for a spiral heat exchanger, *Proceedings of 11th IFAC Symposium on Dynamics and Control of Process Systems Including Biosystems,* **49**(7), 2016, pp. 342-346.
10. Z. Guan, I. Mizumoto, and T. Yamamoto, Design of PCA-based Data-Driven Adaptive Output Feedback Control for Nonlinear Systems, *IEEE 28th International Symposium on Industrial Electronics*, 2019, pp. 497-502.
11. Z. Q. L, On identification of the controlled plants described by the Hammerstein system, *IEEE Transactions on Automatic Control*, **39**(2), 1994, pp. 569-573.
12. K. L. Chien, J. A. Hrones, and J. B. Reswick, On the automatic control of generalized passive systems, *Transaction ASME*, **74**, 1972, pp. 175-185.

Design of a Data-Driven Controller using Open-Loop Data

Yasuteru Nishiya
Graduate School of Advanced Science and Engineering, Hiroshima University
1-4-1, Kagamiyama, Higashihiroshima city, Hiroshima, Japan

Takuya Kinoshita
Graduate School of Advanced Science and Engineering, Hirosihma University,
1-4-1, Kagamiyama, Higashihiroshima city, Hiroshima, Japan

Toru Yamamoto
Graduate School of Advanced Science and Engineering, Hirosihma University,
1-4-1, Kagamiyama, Higashihiroshima city, Hiroshima, Japan
E-mail: {nishiya-yasuteru, kinoshita-takuya, yama}@hiroshima-u.ac.jp
http://www.hiroshima-u.ac.jp

Abstract

In recent years, data-driven control has been proposed and extended to a non-linear system by using database. At this time, various data are required to obtain good control performance but the cost is required. In this paper, a new scheme that enables various data generation and control system design from a set of open-loop data is proposed. Besides, the filter is designed to keep the value of the reference signal constant. The effectiveness of the proposed scheme is numerically verified.

Keywords: data-driven control, PID controller, response prediction, offline, reference signal, filter

1. Introduction

Recently, data-driven control schemes[1,2,3] have been proposed, and it can calculate control parameters directly from a set of input/output data without system modeling. In addition, a Database-Driven control methods (DD)[4,5] have been proposed for nonlinear systems using databases. In a DD control system, various data need to be stored in a database to obtain good control performance, but collecting enough data requires multiple experiments, and it takes time and human cost.

In order to avoid the costs associated with data acquisition, DD control using an Estimated Response Iterative Tuning (DD-ERIT)[6] has been proposed. It generates estimated input/output data from a closed-loop data set. However, in the ERIT method, a two-degree-of-freedom control system must be used to obtain the estimated data. Moreover, the estimation accuracy is degraded when constructing a one-degree-of-freedom control system. Therefore, in this paper, a new control scheme that can generate I/O data offline using a set of open-loop data. The proposed scheme is a simple method that requires only the replacement of blocks on the block diagram to generate data. It is also an effective method for the unknown high order systems. Since some prediction I/O data can be obtained offline with experimental data, the time and human costs for experiments can be reduced. Furthermore, the feature of the proposed scheme are as follows:

(i) The frequency response of the system can be estimated.
(ii) It can be extended to control system design.

2. Basic idea of the proposed scheme

Fig. 1. Block diagram of open-loop system.

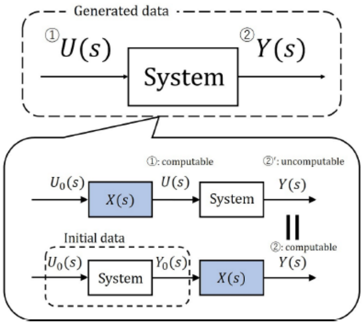


Fig. 2. Outline of the proposed scheme.

In the proposed scheme, open-loop data $U_0(s)$ and $Y_0(s)$ are obtained in Fig. 1. Next, consider applying an arbitrary input signal $U(s)$ shown in Fig. 3. The experiment is required to obtain $Y(s)$ by applying an arbitrary $U(s)$ when the system is unknown. Therefore, in the proposed scheme, the following $U(s)$ is given by using the initial data $U_0(s)$:

$$U(s) = X(s)U_0(s), \tag{1}$$

where $X(s)$ denotes the transfer function given by the user to calculate $U(s)$ using $U_0(s)$. Here, the output $Y(s)$ in Fig. 2 is equivalent even if the front and back of the blocks are swapped because of the properties of the linear block diagram. Therefore, the output $Y(s)$ is calculated by following equation:

$$Y(s) = X(s)Y_0(s). \tag{2}$$

From equation (1) and (2), the output $Y(s)$ can be calculated offline by using the initial open-loop data $U_0(s)$, $Y_0(s)$. Also, in Fig. 2, various input/output data can be generated offline by changing $X(s)$. The following sections describe the control system design using an open-loop data.

3. Design of a data-driven PID controller using a set of open-loop I/O data

Fig. 3 and Fig. 4 show an overview of the controller design using the proposed scheme. First, a set of data is generated using the proposed scheme according to Fig. 3 and then design the controller $C(s)$ to minimize $\epsilon_1(s)$ shown in Fig. 4 based on the generated data. The details

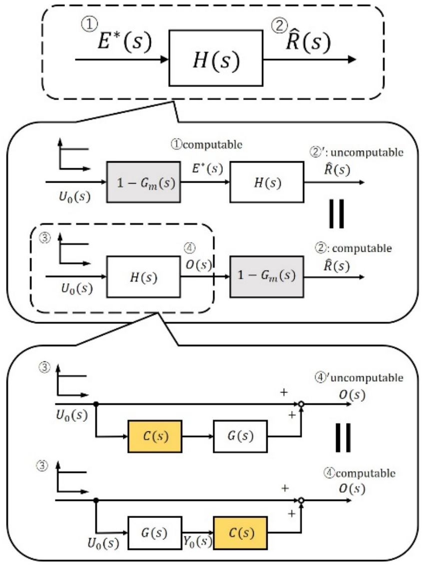


Fig. 3. Generate signal $E^*(s)$ and $\hat{R}(s)$ based on the proposed scheme.

of Fig. 3 and Fig. 4 will be described later. First, the output of the reference model $Y_m(s)$ is given by

$$\begin{aligned} Y_m(s) &= G_m(s)R(s) \\ &= \frac{G(s)C^*(s)}{1+G(s)C^*(s)}R(s), \end{aligned} \tag{3}$$

where $G_m(s)$ denotes the reference model, $G(s)$ denotes the controlled system, $C^*(s)$ shows the optimal controller and $R(s)$ represents the reference signal. In this case, the desired error $E^*(s)$ is defined by

$$\begin{aligned} E^*(s) &= R(s) - Y_m(s) \\ &= \big(1 - G_m(s)\big)R(s). \end{aligned} \tag{4}$$

The error $E(s)$ is expressed by

$$E(s) = \frac{1}{1+G(s)C(s)}R(s). \tag{5}$$

From equation (5), $R(s)$ is expressed by

$$R(s) = \big(1 + G(s)C(s)\big)E(s). \tag{6}$$

Here, $\hat{R}(s)$ which achieve the desired error $E^*(s)$ is introduced by following equation:

$$\begin{aligned} \hat{R}(s) &\coloneqq \big(1 + G(s)C(s)\big)E^*(s) \\ &= H(s)E^*(s) \end{aligned} \tag{7}$$

$$H(s) \coloneqq 1 + G(s)C(s). \tag{8}$$

$\hat{R}(s)$ is a reference signal in a closed-loop system, but it is not computable because the parameters of the transfer function $G(s)$ in equation (7) are unknown. Here, $\hat{R}(s)$ is generated based on the proposed scheme, and its overview is shown in Fig. 3. Assuming that the reference signal $R(s)$ and the initial input $U_0(s)$ are equivalent step signals, the $R(s)$ in equation (4) can be replaced by

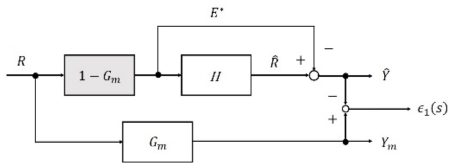


Fig. 4. Block diagram of the minimization problem.

$U_0(s)$. $E^*(s)$ in equation (4) can be computed offline because $U_0(s)$ has already been obtained as initial input data. However, from equation (7), it is impossible to compute $R(s)$ offline using the desired error $E^*(s)$ because the controlled system $G(s)$ is unknown. Therefore, in the proposed scheme, $\hat{R}(s)$ can be computed offline by replacing the order of the blocks, as shown in Fig. 3. Note that, the signal $O(s)$ in Fig. 3 can also be calculated according to the proposed scheme. From the above, $E^*(s)$ and $\hat{R}(s)$ can be computed offline for any $C(s)$ using the initial data $U_0(s)$ and $Y_0(s)$.

However, the original goal of control is to design a controller $C(s)$ in which output $Y(s)$ follows to the reference output $Y_m(s)$. The output $\hat{Y}(s)$, where the reference signal is $\hat{R}(s)$ and the error is $E^*(s)$ in a closed-loop system, can be expressed by the following equation using (4) and (7):

$$\hat{Y}(s) = \hat{R}(s) - E^*(s). \quad (9)$$

Fig. 4 shows the relationship between $\hat{Y}(s)$ and $Y_m(s)$ using the step signal. At this time, let $Y(s)$ and $Y_m(s)$ be Laplace inverse transformed signals $\hat{y}(t)$ and $y_m(t)$, respectively, and define the following equation as an evaluation function using the sum of the squares of the errors:

$$J_1 = \int \left(y_m(t) - \hat{y}(t)\right)^2 dt. \quad (10)$$

The controller in a closed-loop system is designed by optimizing the control parameters using equation (10). However, the input response is unknown when designing a control system focusing on only the output signal. In this case, the system will be overloaded and cause failure if the input response is excessive. Therefore, the following criterion J_2 with a weight coefficient λ is introduced to consider input signal:

$$J_2 = \int \left(\left(y_m(t) - \hat{y}(t)\right)^2 + \lambda\left(u(t) - u(t-1)\right)^2\right) dt, \quad (11)$$

where $\hat{u}(t)$ is an estimate of the input and the Laplace-transformed value $\hat{U}(s)$ can be calculated as follow equation:

$$\hat{U}(s) = C(s)E^*(s). \quad (12)$$

4. Simulation Examples

4.1. *Controlled system*

The following controlled system is discussed:

$$G(s) = \frac{s+6}{s^3 + 6s^2 + 11s + 6}. \quad (13)$$

The step signal used in the initial experiments $u_0(t)$ is equal to 1.0. The reference output $y_m(t)$ is given by

$$y_m(t) = \frac{z^{-1}T(1)}{T(z^{-1})} r(t). \quad (14)$$

Here, let $T(z^{-1})$ be the denominator of the discrete time is defined as follows[7]:

$$T(z^{-1}) = 1 + t_1 z^{-1} + t_2 z^{-2} \quad (15)$$

$$\begin{cases} t_1 = -2\exp\left(-\dfrac{\rho}{2\mu}\right)\cos\left(\dfrac{\sqrt{4\mu - 1}}{2\mu}\rho\right) \\ t_2 = \exp\left(-\dfrac{\rho}{\mu}\right) \\ \rho = \dfrac{T_s}{\sigma} \\ \mu = 0.25(1-\delta) + 0.51\delta \end{cases}, \quad (16)$$

where T_S denotes sampling time. σ and δ show the parameters related to the rise-time and the damping characteristic, respectively. Here, $T(z^{-1})$ was set as follows:

$$T(z^{-1}) = 1 - 1.213z^{-1} + 0.368z^{-2}, \quad (17)$$

where $T(z^{-1})$ was designed by setting σ as 0.5 and δ as 0.0. Here, T_S is equal to 0.001. In addition, the PID controller $C(s)$ is designed as follows:

$$C(s) = K_P + \frac{1}{s}K_I + sK_D. \quad (18)$$

4.2. *Design of data-driven PID controller*

First, the step response for equation (13) as a set of initial data $u_0(t)$ and $y_0(t)$ is obtained. Next, Fig. 5 shows the control results of applying the proposed scheme to minimize the evaluation function of equation (10)(9) using the initial data. In Fig. 5, $y(t)$ and $u(t)$ denote the output and input for a closed-loop system with a reference signal $\hat{r}(t)$. $\hat{y}(t)$ and $\hat{u}(t)$ represent the prediction output and input. Also, $y_m(t)$ denote the output of the reference model. In this case, the control parameters were tuned as follows:

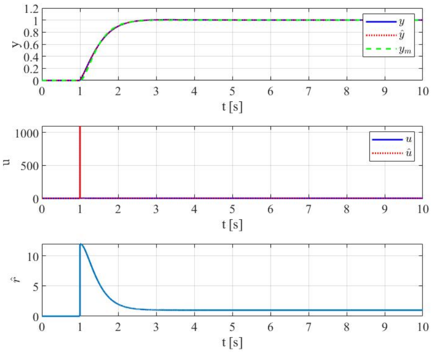


Fig. 5. Control results by the proposed scheme that minimizing equation (10).

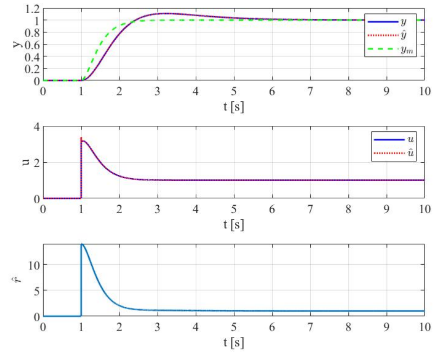


Fig. 6. Control results by the proposed scheme that minimizing equation (11).

$$K_P = 0.261, K_I = 0.167, K_D = 0.091. \quad (19)$$

Fig. 5 shows that the accuracy of $\hat{y}(t)$ and $\hat{u}(t)$ are high. However, the input $u(t)$ changes abruptly and the actuator is heavily loaded since evaluation function in equation (9) is considered only $y(t)$.

Next, the controller is designed by minimizing the evaluation function of equation (11). Fig. 6 shows the control results. In this case, the design the desired characteristic polynomial included in the reference model was set as equation (17), and the weight coefficient λ is set as 0.016. In addition, the control parameters were tuned as follows:

$$K_P = 0.223, K_I = 0.143, K_D = 1.85 \times 10^{-5}. \quad (20)$$

Comparing Fig. 6 with Fig. 5, it can be confirmed that the input changes gently and the maximum value becomes smaller by changing the evaluation function. This indicates that the load on the actuator becomes smaller. In addition, Fig. 6 shows the prediction I/O data $\hat{y}(t)$ and $\hat{u}(t)$ are consistent with the actual data $y(t)$ and $u(t)$, respectively. The prediction accuracy is well estimated without degradation.

5. Conclusion

In this paper, a new scheme that multiple data can be generated offline from a set of open-loop I/O data has been proposed. According to the proposed scheme, controller can be designed by generating data offline from a set of open-loop I/O data. The effectiveness of the proposed scheme has been verified by a numerical simulation. It is possible to design the controller to avoid overloads in actual machines by setting the evaluation functions appropriately. The future work for this study is planned to design a controller for nonlinear systems and verify the effectiveness of the proposed scheme by experiment.

References

1. S. Soma, O. Kaneko and T.Fujii, "A New Approach to Parameter Tuning of Controllers by Using One-Shot Experimental Data: A Proposal of Fictitious Reference Iterative Tuning (in Japanese)", *Trans. of the Institute of systems, Control and Information Engineers*, **17**(2) (2004) pp. 528–536.
2. M.C. Campi, A. Lecchini and S.M. Savaresi, "Virtual Reference Feedback Tuning: A Direct Method for the Design of Feedback Controllers", *Automatica*, **38**(8) 2002, pp. 1337-1346.
3. T. Nakamura, O. Kaneko, "Estimated Response Iterative Tuning of Feed-Forward Controller in the Two-Degree of Freedom Control System (in Japanese)", *The society of Instrument and Control Engineers*, 2017
4. T. Yamamoto, K. Takao and T. Yamada, "Design of a Data-Driven PID Controller", *IEEE transactions on control systems technology*, **17**(1), 2009, pp. 23-29.
5. S. Wakitani, T. Yamamoto and B. Gopaluni, "Design and Application of a Database-Driven PID Controller with Data-Driven Updating Algorithm", *Industrial & Engineering Chemistry Research,* **58(**26), 2019, pp. 11419-11429.
6. R. Okada, T, Kinoshita and T. Yamamoto, "Design of a Database-Driven Controller Using Estimated Data", *Proceedings of the Society of Instrument and Control Engineers Annual Conference 2019*, WeB02.3, Hiroshima.
7. T. Yamamoto, S. L. Shah, "Design and Experimental Evaluation of a Multivariable Self-Tuning PID Controller", *IEEE Proc. of Control Theory and Applications,* **151**, 2004, pp. 645-652.

Improved Estimation of Sway-Angle for Overhead Crane based on Phase Difference of Acoustic Signals in Frequency Domain

Hanako Ogawa, Takeshi Yamada, and Masayoshi Nakamoto

Hiroshima University, Kagamiyama 1-4-1
Higashi Hiroshima, 739-8527, Japan

E-mail:oagwa-hanako@hiroshima-u.ac.jp, msy@hiroshima-u.ac.jp

Abstract

During a crane operation, swaying of the payload is an issue that should be controlled. For anti-sway control, it is essential to detect the sway angle. In this paper, we propose a method for obtaining the phase difference of the acoustic signals within the frequency domain by using the discrete Fourier transform. Based on the phase difference, we compute the time of arrival and convert it into the sway-angle of deflection. In experiment, we show the effectiveness of the proposed method by comparing with the sway angle observed from camera.

Keywords: Overhead crane, sway angle, two microphones, time difference of the arrival (TDOA)

1. Introduction

Cranes are widely used in industries such as factories, warehouses, and construction sites, and various types of cranes available according to their intended use. Among them, overhead cranes are frequently applied. An overhead crane carries a load using a cart on rails provided in the ceiling of the building. During the operation of the crane, there is a need for anti-swaying of the load. Anti-swaying is currently conducted manually and requires skilled workers. However, crane operators are aging, and the number of skilled operators is decreasing. Therefore, there is a growing need for computerized anti-sway control of the lifting load to support unskilled personnel, promote efficiency, and improve the safety of crane operations. With this background, research on anti-sway control is being conducted.

To conduct anti-sway control using a computer, it is essential to detect the sway angle, and an image sensor[1] or encoder[2]-[4] is mainly used. To estimate the sway angle using an image sensor, it is necessary to install the sensor at the location where the sway angle can be captured correctly. In additionFuthermore, a lengthy processing time is required to process the video signal obtained from the sensor. Direct measurements of the sway angle are difficult to achieve using an encoder when the rope of the load is deflected.

Another method for measuring the sway angle of a crane was also proposed, which uses the time difference of the arrival (TDOA) of sound waves [5]. Although with this method the sway angle is calculated from TDOA of the sound waves, the accuracy depends on the sampling frequency. In addition, with such methods, the difference in the arrival time is calculated based on the correlation; however, the computational burden is high because multiple signals are applied to find the maximum value. In addition, the experimental results also show large errors, and thus there is still room for improvement. Therefore, we propose a method for determining the phase of the acoustic signal using da discrete Fourier transformtr (DFT) and calculating the difference in the arrival time and the sway angle from the phase. We also present a frequency selection method for a sound source generation that improves the accuracy of the DFT phase calculation. There are multiple candidate values when determining the difference in the arrival time. When the frequency of the acoustic signal is low, it is more susceptible to environmental noise, and at higher frequencies, even small noise can affect the accuracy of the signal because the difference in the interval between the candidate values of the arrival time narrow despite the higher accuracy. Such aspects make it difficult to choose an appropriate difference in the arrival time. To avoid this problem, we also propose using signal processing for the correct tracking of the differences in the arrival time.

In experiments on measuring the crane sway angle, a comparison was made between the sway angle estimated by the proposed method and that measured by the cameras. We also confirmed that the difference in the arrival time can be correctly tracked through signal processing.

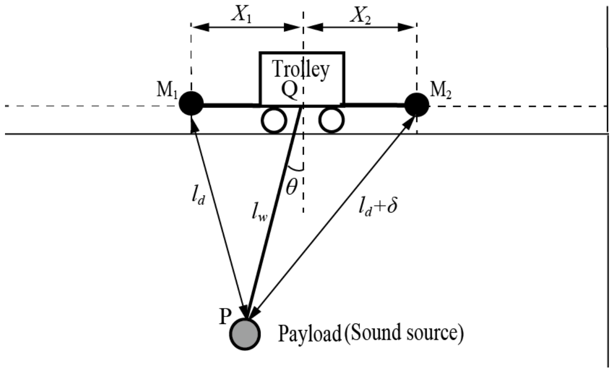


Fig. 1 Overhead crane system and its parameters

2. Acoustic signal processing for measurement of sway angle

In [5], two microphones are used for the contactless sensing of the angle, given the assumption that the payload (or hook) includes sound sources, as shown in Fig. 1. Using the two microphones, we estimated TDOA between the acoustic signals of the sound source in the payload.

2.1 Estimation of TDOA

Fig. 1 shows an overhead crane system in which the trolley has two microphones labeled M_1 and M_2. The sound source is embedded in the payload (P in Fig.1). We assume the sway angle is θ , and the distance from the trolley to the sound source is l_w. Using the microphones, we first observe the sound signal originating from the sound source. It should be noted that the sensitivity to the sound signal from behind the microphones is extremely low because the microphones have directivity. Now, let $s_1(t)$ be the sound signal observed in M_1, and $s_2(t)$ be the sound signal observed in M_2.

Assume that the distance between M_1 and the trolley (Q in Fig. 1) is X_1 and the distance between M_2 and the trolley is X_2.

Hence, the distance between M_1 and the sound source is $l_d+\delta$, and the distance between M_2 and the sound source is l_d. Assume that X_1, X_2, and are known in Fig. 1. It should be noted that δ is the difference in distance between PM_1 and PM_2. Now, consider estimating δ using the TDOA. Let τ_k be the number of samples for the time difference between the two signals.

In [5], τ_k can be computed using the correlation of two signals. The TDOA between M_1 and M_2 is $\tau_k\, T_s$, where T_s is the sampling period. It follows that δ can be computed as follows:

$$\delta = \tau_k^* T_s v_a \qquad (1)$$

where v_a is the velocity of sound.

The problem here is that the resolution of δ depends on T_s because $\tau_k\, T_s$ are discrete values of interval T_s. Hence, we present a new approach to obtain δ that is independent of T_s. Let $\lambda_1(\tilde{k})$ and $\lambda_2(\tilde{k})$ be phases of $s_1(t)$ and $s_2(t)$, respectively. Using a discrete Fourier transform (DFT), $\lambda_1(\tilde{k})$ and $\lambda_2(\tilde{k})$ can be obtained by the following:

$$S_1(\tilde{k}) = \sum_{n=0}^{N-1} s_1(n) e^{-j\frac{2\pi}{N}\tilde{k}n} = A_1(\tilde{k}) e^{j\lambda_1(\tilde{k})}, \qquad (2)$$

$$S_2(\tilde{k}) = \sum_{n=0}^{N-1} s_2(n) e^{-j\frac{2\pi}{N}\tilde{k}n} = A_2(\tilde{k}) e^{j\lambda_2(\tilde{k}),} \qquad (3)$$

where $\tilde{k}$ is a positive integer that corresponds to the frequency of the sound source, N is the number of data, and $A_1(\tilde{k})$ and $A_2(\tilde{k})$ are the amplitudes of the frequency spectra.

Define the difference between $\lambda_1(\tilde{k})$ and $\lambda_2(\tilde{k})$ as

$$\Delta\lambda = \lambda_1(\tilde{k}) - \lambda_2(\tilde{k}). \qquad (4)$$

Using (4), the TDOA can be computed through the following:

$$\Delta t = \frac{\Delta\lambda}{2\pi f_{\tilde{k}}}, \qquad (5)$$

where $f_{\tilde{k}}$ is the frequency of the sound source.

Consequently, the following velocity of sound v_a, δ can be obtained :

$$\delta = v_a \Delta t. \qquad (6)$$

To generate the acoustic signal of the sound source, the discrete frequency of the DFT is computed as

$$0, \frac{f_s}{N}, \frac{2f_s}{N}, \dots, \frac{(N-1)f_s}{N}, \tag{7}$$

Here f_s is the sampling frequency $f_s = 1/T_s$. Now, it is assumed that N and $f_s$are given in the specifications of the hardware. In this study, the frequencies corresponding to the discrete frequencies, as indicated in (7), are used. Hence, the frequency (or frequencies) of the sound are selected such that

$$f_{\tilde{k}} = \frac{\tilde{k} f_s}{N}, \tag{8}$$

where $1 \le \tilde{k} \le N-1$. That is, to achieve a high performance of (6), the frequency (or frequencies) of the sound source is decided by N and f_s of the sound receiving device.

2.2 Sway angle estimation Consider the transformation of δ to θ by the following procedure: Using the cosine theorem for $\triangle \mathrm{M_1PQ}$ and $\triangle \mathrm{M_2PQ}$, we have the following:

$$\begin{cases} l_d = X_1^2 + l_w^2 - 2X_1 l_w \cos\left(\frac{\pi}{2} - \theta\right) \\ (l_d + \delta)^2 = X_2^2 + l_w^2 - 2X_2 l_w \cos\left(\frac{\pi}{2} + \theta\right). \end{cases} \tag{9}$$

Now, letthe following:

$$f(\theta) = \delta + \sqrt{X_1^2 + l_w^2 - 2X_1 l_w \sin\theta} - \sqrt{X_2^2 + l_w^2 + 2X_2 l_w \sin\theta}. \tag{10}$$

Then, θ can be obtained by computing the solution of the following equation:

$$f(\theta) = 0, \quad \theta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right). \tag{11}$$

However, it is difficult to solve (11) explicitly because 1it is a non-linear equation. Hence, we introduce an algorithm to solve (11) based on Newton's method.

Differentiating $f(\theta)$ by θ, we have the following:

$$\frac{df(\theta)}{d\theta} = \{X_1^2 + l_w^2 - 2X_1 l_w \sin\theta\}^{-\frac{1}{2}}\{-X_1 l_w \cos\theta\} - \{X_2^2 + l_w^2 + 2X_2 l_w \sin\theta\}^{-\frac{1}{2}}\{X_2 l_w \cos\theta\}. \tag{12}$$

Let θ_0 be an initial value of the sway angle, and θ_n be the sway angle at the n-th step. It follows from (12) that Newton's method can be used to solve (11). Thus, we obtain the sway angle by repeating the following procedure:

$$\theta_{n+1} = \theta_n - \frac{f(\theta_n)}{\left\{\frac{df(\theta_n)}{d\theta}\right\}}. \tag{13}$$

The terminal condition is $|\theta_{n+1} - \theta_n| < \varepsilon$ wwhere ε is a small positive number.

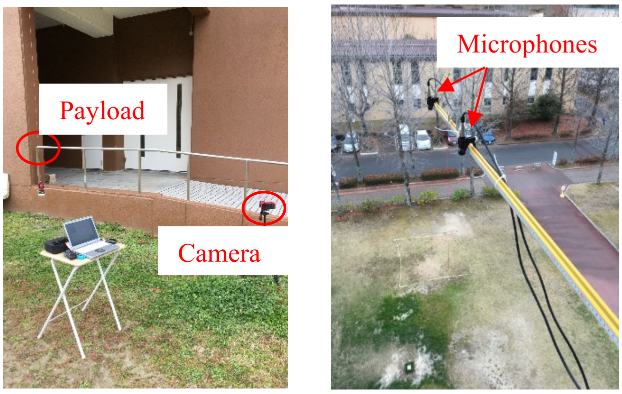


Fig. 2 Experimental devices (microphones, payload, camera)

3. Experiment

Fig. 2 shows the experimental equipment. In this measurement, a sine wave was used as the sound source in the payload. Here, the parameter of the terminal condition is set as $\varepsilon = 0.0001$. The parameters are $l_w = 15.11$ [m], $N = 10^2$, $f_s = 44100$ [Hz] , and $X_1 = X_2 = 30$ [cm]. In addition, video clips were captured, and the sway angles were then calculated from the image processing and compared. The frequency spectra of the observed signals are shown in Fig. 3. In this experiment, the frequencies of $f_k = 5038.77$ [Hz] and $f_k = 8010.35$ [Hz] were used.

The payload (sound source) was individually vibrated. The estimation of θ for $f_k = 5038.77$ [Hz] is shown in Fig. 4 where an offset of stationary state is considered. Similar to Fig. 4, the estimation of θ for $f_k = 8010.35$ [Hz] is shown in Fig. 5. The blue line is the

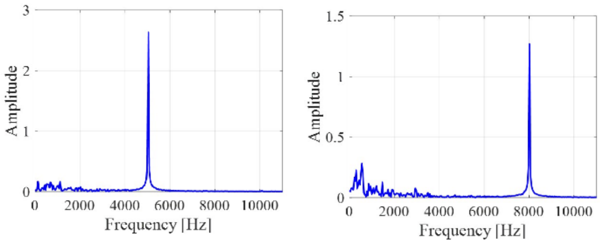


Fig. 3 Frequency spectra of observed signals

estimated sway angle from the sound (proposed method), and the red line is the sway angle from the image processing.

Fig. 5 shows that the sway angle of the proposed method agrees well with the sway angle of the image processing, and it can be confirmed that the quality point tracking can be achieved even under the influence of disturbances and noises on the sound source from a blower. Comparing Figs. 4 and 5, we can see that the result in Fig. 5 has a smaller error. As the reason for this, the frequency bands of the blower noise do not overlap with the frequency bands of the sound source. That is, the deflection angle can be accurately estimated, which confirms the effectiveness of the proposed method.

4. Conclusion

We considered a soft sensor to measure the sway angle for the anti-sway control of an overhead crane. We proposed an approach to estimate the angle based on acoustic signals to achieve contactless sensing of the angle. The method uses the phase difference of acoustic signals measured by two microphones. We presented an algorithm to obtain the angle from the phase difference based on Newton's method. During the experiment, we compared the proposed method with measurements taken from photographs. As a result, it was confirmed that the proposed method can dynamically follow the sway angle of the payload . Furthermore, we confirmed that the proposed method can correctly track the sway angle.

5. References

1. C.-Y. Chang and H. W. Lie: Real-Time Visual Tracking and Measurement to Control Fast Dynamics of Overhead Cranes, *IEEE Trans. Ind. Electron.*, vol.59, no.3, pp.1640-1649, March 2012.
2. B. Lu, Y. Fang, N. Sun and X. Wang: Antiswing Control of Offshore Boom Cranes With Ship Roll Disturbances, *IEEE Trans. Control Systems Technology*, vol.26, no.2, pp.740-747, Mar. 2018.
3. N. Sun, T. Yang, H. Chen and Y. Fang: Dynamic Feedback Antiswing Control of Shipboard Cranes Without Velocity Measurement: Theory and Hardware Experiments, *IEEE Trans. Ind. Inform.*, vol.15, no.5, pp.2879-2891, May 2019.
4. N. Sun, Y. Fu, T. Yang, J. Zhang, Y. Fang and X. Xin: Nonlinear Motion Control of Complicated Dual Rotary Crane Systems Without Velocity Feedback: Design, Analysis, and Hardware Experiments, *IEEE Trans. Automation Science and Engineering*, vol.17, no.2, pp.1017-1029, April 2020.
5. M.Matsunaga, M.Nakamoto and T.Yamamoto: A sound-based measurement of sway angle for anti-sway control of over head crane, Journal of Robotics, Networking and Artificial Life, Vol.4, No.4, pp.322-355, March 2018.

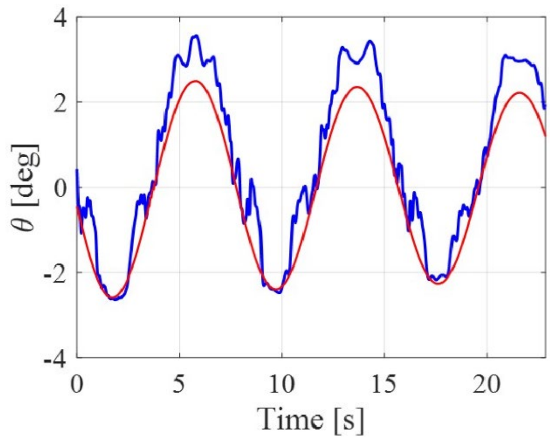


Fig. 4 Results of estimating θ for $f_k = 5038.77$ [Hz]

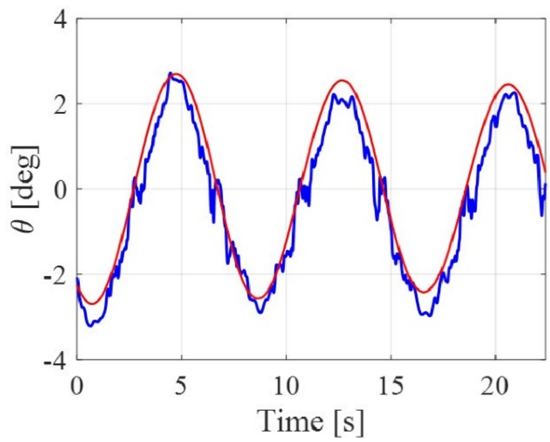


Fig. 5 Result of estimating θ for $f_k = 8010.35$ [Hz]

* Study on an Optimal Design Method for Control Systems based on Bayesian Optimization

Koichi Hirota†
Graduate School of Advanced Science and Engineering, Hiroshima University, 1-4-1, Kagamiyama Higashi-Hiroshima city, Hiroshima, Japan

Shin Wakitani,
Graduate School of Advanced Science and Engineering, Hiroshima University, 1-4-1, Kagamiyama Higashi-Hiroshima city, Hiroshima, 739-8527 Japan
E-mail: wakitani@hiroshima-u.ac.jp

Toru Yamamoto,
Graduate School of Advanced Science and Engineering,Hiroshima University, 1-4-1, Kagamiyama Higashi-Hiroshima city, Hiroshima, 739-8527 Japan
E-mail: yama@hiroshima-u.ac.jp

Abstract

When designing a complex product system, must be considered the optimal function distribution between the controller and the plant. This makes it possible to design a plant that is easy to control for fixed PID control, and the performance of the entire system can achieve the desired performance. Here, the evaluation function of the system is calculated from the output error and the input difference value. Then, the optimum design parameters are calculated by using Bayesian optimization.

Keywords: Optimal Allocation of Function, Model-Based Development, Bayesian Optimization, Evaluation Function,.

1. Introduction

Most product systems consist of a combination of multiple subsystems. Appropriate setting and design of operating goals (functional goals) of these subsystems are important to achieve and maintain the target performance of the system. In addition, the performance required by users for products is becoming more diverse and complex year by year. It is necessary to efficiently design a combination of subsystems to satisfy these required performances in a short period of time and at a low cost. In recent years, model-based development (MBD) has been attracting attention in the industrial world, and its introduction is being active. In the MBD process, entire subsystems are modeled as simulation models and are connected to each other on the simulation environment (such as MATLAB/Simulink). Moreover, a user can evaluate the subsystem's performance and easily improve their design without any prototyping. Meanwhile, the actuators of many subsystems are becoming electrified. Therefore, a feedback control system including controllers represented by by-wire technology is used in various situations. However, in many situations of control system design, plant design is completed in advance. In the conditions of this plant characteristic, the structure and parameters of the controller are often adjusted to satisfy the desired control performance. In particular, when the control target includes high-order components, strong non-linearity, and large dead time, it is difficult to achieve high control performance with a low-order controller such as a PID

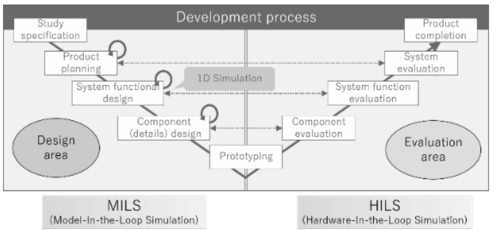


Fig. 1. V-shaped development process.

controller. On the other hand, if a high-order controller is introduced, the adjustment of control parameters becomes complicated, and the uncertainty of the system may impair the stability of the control loop. Therefore, in subsystem design, a plant and a controller are required to be optimized at the same time for the functional goals determined upstream design process. Based on the background, this study considers the simultaneous optimum design of the controller and the plant.

2. Simultaneous Optimal Design of Controller and Plant in MBD

2.1. V-shaped development process

Figure 1 shows the product development process, V-shaped development process, in MBD. The V-shaped development process (V-model) is divided into the model-in-the-loop simulation (MILS) which is a design area using a model and the hardware-in-the-loop simulation (HILS) which is a verification area that uses hardware and a model simultaneously. Since this research is mainly related to the design method, only MILS is be explained. Firstly, the necessary functions for a designed product system and specifications will be examined based on the results of market research. Next, the product system is decomposed into several subsystems, and the functions that these subsystems should satisfy are examined using 1D simulation technology. Furthermore, the parts (detailed design), that compose a subsystem is designed so that it operates in the same way as the functional model determined in the upstream design process. In this process, a detailed design including the shape of parts is performed using 3D simulation technology.

2.2. Optimal allocation of plant and controller functions

When designing a feedback system composed of a plant and a controller, the controller is usually designed from the state in which the plant has already been designed. However, when designing a control system, there are input restrictions and non-linearities in the designed plant, and these characteristics appear as constraints when designing the controller. Therefore, it is often the case that the desired control performance cannot be obtained with low-order controllers such as PID controllers that are usually introduced in systems. However, if the spread of MBD makes it possible to allocate the optimum functions of the plant and controller in advance, the performance of the entire system will be closer to the desired performance by designing a plant that is easy to control for fixed PID controller. Therefore, this study, considers the optimum allocation method of plant and controller functions for the system function in the design phase. In the subsystem, to reach a given functional goal is the most important. Where, it is assumed that the functional goal can be described as the dynamic input/output relationship, that is the reference model[1]. The error between the output of the subsystem corresponding to a certain input scenario and the output of the reference model is evaluated as follows.

$$I_e = \frac{1}{N}\sum_{k=1}^{N}\left[1 - \exp\left\{-\frac{\left(y_r(k) - y(k)\right)^2}{2\sigma_e^2}\right\}\right]. \qquad (1)$$

However, N represents the number of data acquired in the simulation evaluation interval. In Eq. (1), the Gaussian kernel is used for the output error. As a result, the value of Eq. (1) is normalized between 0 to 1 and the permissible ratio to the error can be adjusted by σ_e^2. Moreover, in the controller design, it is desirable to impose some restrictions on the controller output from the viewpoint of actuator protection and noise robustness. Here, the following evaluation function is designed for the input difference value $\Delta u(t) = u(t) - u(t-1)$.

$$I_e = \frac{1}{N}\sum_{k=1}^{N}\left[1 - \exp\left\{-\frac{(\Delta u(k))^2}{2\sigma_{\Delta u}^2}\right\}\right]. \qquad (2)$$

By using the Gaussian kernel as in Eq. (2), the evaluation value for the input difference value is normalized, and a certain permissible ratio is given as $\sigma_{\Delta \mathrm{u}}$. For Eqs. (1) and (2), the evaluation function for optimizing the entire subsystem is defined by the following equation.

$$I = \lambda I_e + (1-\lambda) I_{\Delta u}. \qquad (3)$$

However, $\lambda (0 \le \lambda \le 1)$ is an adjustment parameter. Based on the above equation, if the design parameter vector composed of control parameters and plant parameters is $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_n,)^T$, the optimization problem is formulated by the following equation.

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \left(I(\boldsymbol{\theta})\right). \qquad (4)$$

The next chapter explains the design parameter optimization method based on Bayesian optimization.

3. Bayesian Optimization

Bayesian optimization[2,3] can be divided into the regression process based on Gaussian process regression using obtained input/output data and a process for determining the next search point by calculating an acquisition function.

3.1. Gaussian process regression

In the Gaussian process regression, the mean and variance of the system output are estimated. First, this calculation method is explained. Let the input x and the output y are expressed by Eqs. (5) and (6).

$$\boldsymbol{x} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_N)^T . \quad (5)$$

$$\boldsymbol{y} = (I_1, I_2, \cdots, I_N)^T . \quad (6)$$

Where, N is the number of data.

The feature vector of $\boldsymbol{x}$ is given as Eq. (7).

$$\boldsymbol{\phi} = (\phi_0(\boldsymbol{x}), \phi_1(\boldsymbol{x}), \cdots, \phi_H(\boldsymbol{x}))^T . \quad (7)$$

If the weights vector is described as $\boldsymbol{w} = (w_0, w_1, \cdots, w_H)^T$, the linear regression model becomes Eq. (8).

$$\begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_N \end{pmatrix} = \begin{pmatrix} \phi_0(\boldsymbol{x}_1) & \phi_1(\boldsymbol{x}_1) & \cdots & \phi_H(\boldsymbol{x}_1) \\ \phi_0(\boldsymbol{x}_2) & \phi_1(\boldsymbol{x}_2) & \cdots & \phi_H(\boldsymbol{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\boldsymbol{x}_N) & \phi_1(\boldsymbol{x}_N) & \cdots & \phi_H(\boldsymbol{x}_N) \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_H \end{pmatrix} . \quad (8)$$

$$\hat{\boldsymbol{y}} = \Phi \boldsymbol{w} . \quad (9)$$

However, $\hat{\boldsymbol{y}}$ is the predicted value of the output y and is shown by Eq. (10).

$$\hat{\boldsymbol{y}} = (\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_N)^T . \quad (10)$$

In addition, Φ is called a design matrix and becomes Eq. (11).

$$\Phi = \begin{pmatrix} \phi_0(\boldsymbol{x}_1) & \phi_1(\boldsymbol{x}_1) & \cdots & \phi_H(\boldsymbol{x}_1) \\ \phi_0(\boldsymbol{x}_2) & \phi_1(\boldsymbol{x}_2) & \cdots & \phi_H(\boldsymbol{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\boldsymbol{x}_N) & \phi_1(\boldsymbol{x}_N) & \cdots & \phi_H(\boldsymbol{x}_N) \end{pmatrix} . \quad (11)$$

Where, by assuming $\boldsymbol{y} = \hat{\boldsymbol{y}}$, Eq. (9) can be obtained.

$$\boldsymbol{y} = \Phi \boldsymbol{w} . \quad (12)$$

Here, if a weights vector $\boldsymbol{w}$ is generated from the Gaussian distribution $\boldsymbol{w} \sim N(\boldsymbol{0}, \lambda^2 \boldsymbol{I})$, $\boldsymbol{y}$ also follows a Gaussian distribution, then the mean E [$\boldsymbol{y}$] and the covariance matrix $\boldsymbol{K}$ becomes as follows.

$$E[\boldsymbol{y}] = E[\Phi \boldsymbol{w}] = \Phi E[\boldsymbol{w}] = 0 . \quad (13)$$

$$\begin{aligned} \boldsymbol{K} &= E[\boldsymbol{y}\boldsymbol{y}^T] - E[\boldsymbol{y}]E[\boldsymbol{y}] \\ &= E[(\Phi \boldsymbol{w})(\Phi \boldsymbol{w})^T] \\ &= \Phi E[\boldsymbol{w}\boldsymbol{w}^T]\Phi^T \\ &= \lambda^2 \Phi \Phi^T . \end{aligned} \quad (14)$$

Therefore, the distribution of $\boldsymbol{y}$ is a multivariate Gaussian distribution shown in Eq. (12).

$$\boldsymbol{y} \sim N(\boldsymbol{0}, \lambda^2 \Phi \Phi^T). \quad (15)$$

The function that finds the value of the $\boldsymbol{K}$ element K_{ij} is called a kernel function. Kernel functions represent the similarity of inputs. There are several types, and the Gaussian kernel is widely used. The Gaussian kernel is shown in Eq. (13).

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \theta_1 \exp\left(- \frac{\left\| \boldsymbol{x}_i - \boldsymbol{x}_j \right\|^2}{\theta_2} \right) . \quad (16)$$

$\|\cdot\|$ indicates the Euclidean norm of the vector. In addition, θ_1 and θ_2 are parameters that determine the properties of the kernel function. Here, consider finding the prediction point y^* for an arbitrary vector x^*. First, using the output data vector in Eq. (6) and the predicted output y^*, a new vector $\boldsymbol{y}^{new}$ is defined as follows.

$$\boldsymbol{y}^{new} = (\boldsymbol{y}^T, y^*)^T . \quad (17)$$

If the covariance matrix $\boldsymbol{K}^{new}$ is calculated from $\boldsymbol{y}^{new}$, $\boldsymbol{x}$, and $\boldsymbol{x}^*$, then the following relationship is hold.

$$\boldsymbol{y}^{new} \sim N(\boldsymbol{0}, \boldsymbol{K}^{new}). \quad (18)$$

Here, the covariance matrix $\boldsymbol{K}^{new}$ is expressed by the following equation.

$$\boldsymbol{K}^{new} = \left(\begin{array}{c|c} \boldsymbol{K} & \boldsymbol{k}_* \\ \hline \boldsymbol{k}_*^T & k_{**} \end{array} \right) . \quad (19)$$

$$\boldsymbol{k}_* = \left(k(\boldsymbol{x}^*, \boldsymbol{x}_1), k(\boldsymbol{x}^*, \boldsymbol{x}_2), \cdots, k(\boldsymbol{x}^*, \boldsymbol{x}_N) \right)^T . \quad (20)$$

$$k_{**} = k(\boldsymbol{x}^*, \boldsymbol{x}^*). \quad (21)$$

From Eq. (18), the mean and variance of $\boldsymbol{y}^*$ can be calculated by Eq. (23).

$$y^* \sim N(\mu, \sigma^2). \quad (22)$$

$$\sim N(\boldsymbol{k}_*^T \boldsymbol{K}^{-1} \boldsymbol{y}, k_{**} - \boldsymbol{k}_*^T \boldsymbol{K} \boldsymbol{k}_*). \quad (23)$$

From the above results, the mean μ and variance σ^2 of the unknown $\boldsymbol{y}^*$ corresponding to arbitrary point $\boldsymbol{x}^*$ can be derived.

3.2. Acquisition function

The acquisition function is an equation calculated using the mean μ and standard deviation σ obtained by Gaussian process regression. This function is used to determine the next point to search. There are several types of acquisition functions, in this study, the following Lower Confidence Bound (LCB) is introduced.

$$a_{LCB} = \mu - \sqrt{\frac{\log N}{N}} \sigma . \quad (24)$$

The LCB searches for a place with a large standard deviation, that is, an unknown region, as the number of data decreases. Also, as the number of data increases, the search proceeds based on the data. From the above, it is possible to search for the optimum value of the function obtained by Bayesian optimization globally and efficiently.

4. Numerical Example

The effectiveness of the proposed design method is evaluated by numerical example. Here, consider a control system consisting of a controller, actuator, and plant as shown in Fig. 2.

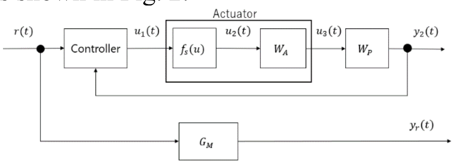


Fig. 2 Feedback control system

Here, $f_s\,(u)$ and W_A are the characteristics of the actuator, and W_P is the dynamic characteristics of the plants that make up the subsystem. $f_s\,(u)$ is a function that expresses the input saturation of the actuator and is given by the following equation.

$$u_2(t) = \begin{cases} u_{min}, & (u_1(t) < u_{min}) \\ u_1(t), & (u_{min} \le u_1(t) < u_{max}) \\ u_{max}, & (u_1(t) \ge u_{max}) \end{cases}. \quad (25)$$

The transfer function of plant 1 (W_A) is shown in the following equation.

$$W_A(s) = \frac{K_A}{1 + T_A s}. \quad (26)$$

Here, T_A is the time constant of the actuator, and K_A is the maximum output of the actuator. The transfer function of plant 2 (W_P) is shown in the following equation.

$$W_P = \frac{K\omega^2}{s^2 + 2\zeta\omega s + \omega^2}. \quad (27)$$

Where,

$$K = \frac{1}{k_P}, \zeta = \frac{D_P}{2\sqrt{k_P M_P}}, \omega = \sqrt{\frac{k_P}{M_P}}. \quad (28)$$

However, it is assumed that the system parameters M_P, k_P , and D_p are all positive real numbers. In this simulation, K_A, k_P, and D_P are the design parameters, and the other parameters are the fixed values shown in Table 1. In the controller design, the PID gains are determined based on the pole assignment controller design only using the model of plant 2. Here, the reference model expressing the desired function of the subsystem is given by the following equation.

$$G_m = \frac{1}{\left(1+\frac{\sigma}{3}s\right)^3}. \quad (29)$$

The parameter $\sigma = 0.5$ for the rise time was set in this case. The evaluation function uses Eq. (3), and Table 2 shows the parameters used in the function design.

Table 1. Plant parameters

parameters	value
u_{min}	0
u_{max}	1
T_A	0.01
M_P	10

Table 2. Evaluation function parameters

parameters	value
λ	0.5
σ_e	0.1
$\sigma_{\Delta u}$	0.01

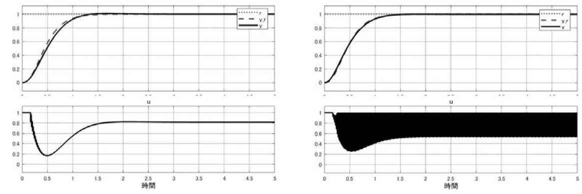


Fig. 3. Control result using parameters (left-hand side: λ = 0.5, right-hand side: λ = 1.0).

As a result of Bayesian optimization under the condition λ = 0.5, the optimum values became $K_A = 61,\ k_P = 50,$ $D_P = 0.1$. At this time, the control result is shown in Fig. 3 (left-hand side). From the control result, it can be seen that the output y generally follows the reference output y_r. If λ = 0.5 in the evaluation function, the optimum values are $K_A = 102,\ k_P = 80,\ D_P = 35$ when Bayesian optimization is performed. At this time, the result of the simulation is shown in Fig. 3 (right-hand side). Compared with the result obtained by λ = 0.5, it seems that the output error is smaller. However, since it can be seen that the input fluctuates drastically, it can conclude that a system in which the fluctuation of the input is suppressed can be designed by adding $I_{\Delta u}$ to the evaluation function.

5. Conclusion

This study proposes an optimal design method of a subsystem composed of a plant and a controller by using Bayesian Optimization. This paper defines a criterion for optimizing the plant parameters so that the plant is designed suitable for a controller. The effectiveness of the proposed design scheme is verified by a simulation example.

6. References

1. Takeshi Shigemasa, Yasuo Takagi, Yoshinori Ichikawa and Toshiyuki Kitamura, A Practical Reference Model for Control System Design, *Transactions of the Society of Instrument and Control Engineers*, Vol. 19, Issue. 7, 1983, pp.592-594.
2. Daichi Mochihashi and Ooba Shigeyuki, Gaussian Process and Machine Learning ~ Machine Learning Professional Series, Kodansya, Tokyo,Japan, 2019, pp. 57-95.
3. C. M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, 2012

Unconstrained Measurement of Heart Rate Considering Harmonics of a Respiratory Signal by Flexible Tactile Sensor Sheet

Kazuya Matsuo
Department of Mechanical and Control Engineering, Kyushu Institute of Technology, 1-1, Sensui, Tobata, Kitakyushu, Fukuoka, 804-8550, JAPAN

Toshiharu Mukai
Meijo University, JAPAN

Shijie Guo
Hebei University of Technology, CHINA

E-mail: matsuo@cntl.kyutech.ac.jp

Abstract

This study describes a heart rate measurement method by considering the harmonics of a respiratory signal. Respiration and heart rates were measured by applying frequency analysis to the time-series data of body pressure. The harmonics of a respiratory signal serve as noises in heart rate measurement. Therefore, heart rate measurement was improved by eliminating the effects of the harmonics. The results of experiments demonstrated that our proposed method enhances the precision of heart rate measurement.

Keywords: Harmonics of a respiratory signal, Heart rate measurement, Unconstrained measurement, Sleeping state.

1. Introduction

In Japan, the percentage of the elderly, aged 65 or older, against the total population will increase to 29% in 2020, 32% in 2030, and surpass 36% in 2040 [1], [2]. The population of the elderly was estimated to be 2,924,000 in 2010 and is projected to be 3,619,000 in 2020 [3]. It is also estimated that the number of individuals requiring care will increase. However, the increase in the number of caregivers may not keep pace with this rise. Therefore, Japan may face an acute lack of caregivers in the immediate future, and the burden on nursing staff is expected to increase. The need for the introduction of a nursing-care apparatus has thus been heightened, and the watching of the perception sickness has been decided by the national government as one of the priority fields.

If the condition of the patient could be assessed using sensors, it could greatly contribute to the reduction of burden on the nursing staff. Conditions of the nursed person to be distinguished include whether they are on the bed and whether their sleeping posture is supine or prone. Addition- ally, measurements of the sleeping state, such as the depth of sleep, would be useful to monitor the health of the nursed person.

Polysomnography (PSG) [4] is a well-known method of measuring the sleeping state. PSG can also diagnose sleep apnea syndrome and periodic limb movement disorder. To recognize the sleeping state, PSG requires biophysiological information, such as brain activity (electroencephalograms), heart activity (electrocardiograms), eye movements, muscle activity (electromyograms) of the chin, and respiratory air- flow. The monitored body functions are used to score each period of sleep according to one of the six sleep stages based on the R&K manual [5]; these include awake, REM sleep, and four stages of non-REM sleep. As a medical specialist makes the final diagnosis, PSG is capable of accurately determining the sleep stage. However, attaching many sensors to the body to obtain numerous biological signals is necessary. Thus, PSG hinders natural sleeping, accompanied by a feeling of restrain. Further, there is a need for the judgment of an inspection technician to distinguish the sleeping state from the biological signal, and thus, automation of this task is difficult. Therefore, a technique that can conveniently estimate the sleeping state is required. The sleeping state can be estimated from biological information such as the respiration rate, heart rate, body motion, and lying posture [6]–[13], which can be measured relatively easily. For example, there is a strong correlation between heart rate and sleep state [6]. Many unconstrained and non- invasive sleep estimation methods have been reported [7]-[11]. Watanabe *et al.* developed a noncontact algorithm for the sleeping state estimation using heartbeat and body- movement signals [7]. Kambayashi *et al.* estimated the sleeping state cycle using body-movement density, measured from a passive infrared type motion sensor [8]. Okada *et al.* proposed a non-restrictive and non-invasive sleeping state evaluation technique from body movements obtained from video images [9]. Deguchi *et al.* evaluated heart rate variability and respiratory variability [10]. Harper *et al.* evaluated the potential of the classification of sleep states using cardiorespiratory measures [11]. This paper proposes a method to measure sleep posture [12], respiration, and heart rate [13] using a flexible tactile sensor sheet.

Conventional sleep estimation methods are often basedon the assumption that respiration and heart rates can be accurately measured. However, it is not always possible to accurately measure biological information. In fact, in the method used in [13], the percentage of successful heart rate measurement in overnight experiments was only 40%.

There are several possible reasons for the failure of ac- curate heart rate measurement. For example, if the pressure information corresponding to the type of sleeping posture is used, and if the identification of the sleeping posture fails, accurate measurements of the respiration and heart rates also fail. Additionally, there is a problem that the harmonics of the respiratory signal behave like noise signals in the heart rate measurement.

As the heartbeat signal is weaker than the respiratory signal, the intensity of the frequency spectrum may be smaller than the harmonics of the respiratory signal. In such a case, the heart rate measurement will fail. In this paper, we propose two methods to improve the measurement accuracy of the heart rate by eliminating the effects of the harmonics of the respiratory signal.

2. Heart Rate Measurement Method Considering Harmonics of a Respiratory Signal

2.1. Measurement method of respiratory and heart rates from body pressure

A method for measuring respiration and heart rates from body pressure information using a flexible planar tactile sensor has been described (Fig. 1) [13]. The body pressure of a person lying on a bed is measured using a tactile sensor sheet. It is observed that the respiration and heart rates remain nearly constant during rest. The frequency bands of respiratory and heart rates do not overlap. Frequency analysis is applied to the time-series data of body pressure, measured by sensor cells in the appropriate position, such as on the chest. A fast Fourier transform (FFT [14], has been used for frequency analysis. Finally, the respiratory and heart rates were acquired by extracting the appropriate frequency band. The frequency bands of the extracted respiratory and heart rate data were determined by preliminary experiments in the following manner: (a) The frequency band of

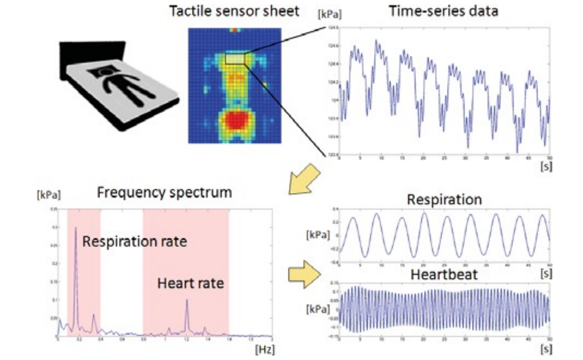


Fig. 1. Measurement method of respiration and heart rates using a flexible planar tactile sensor.

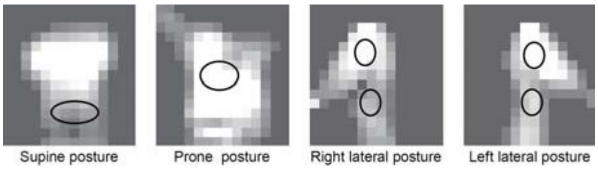


Fig. 2. Appropriate positions for measuring heart rate. The oval shapes show the appropriate positions [12].

respiration is 0.1 to 0.4 Hz (6 to 24 times/min). (b) The frequency band of the heart rate is 0.8 to 1.6 Hz (48 to 96 times/min). When obtaining the frequencies of the respiration and heart rates, their dominant frequencies can be found.

Measuring body pressure at appropriate positions such as the chest to estimate the heart rate is necessary. Fig. 2 shows the appropriate positions for heart rate measurement determined by Mukai [12].

2.2. Heart rate measurement by eliminating respiratory harmonics

In the conventional method [13], the harmonics of the respiratory signal act as noise signals in the measurement of the heart rate. When the heartbeat signal is weak, the intensity of the frequency spectrum may be smaller than the harmonics of the respiratory signal. Hence, the conventional method sometimes fails to measure the heart rate. In this paper, we propose two methods to improve the measurement accuracy of the heart rate by eliminating the harmonics of the respiratory signal. A sample measurement response of the proposed methods is shown in Fig. 3.

The specifications of the proposed methods to eliminate the harmonics of the respiration are described below.

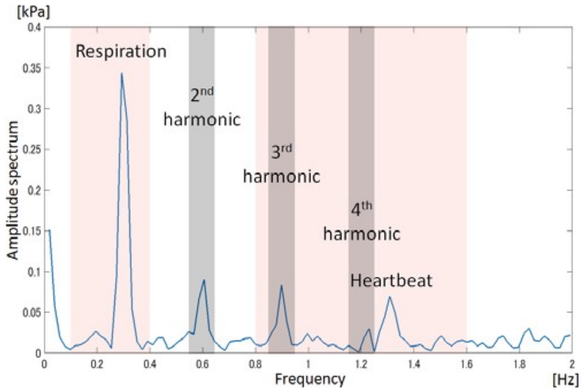


Fig. 3. Measurement of heart rate by eliminating respiratory harmonics.

Proposed Method 1: The frequencies of the second, third, and fourth harmonics are calculated from the frequency of respiration, obtained using the conventional method. The spectral intensity is then set in the ranges of zero to 0.05 Hz around the harmonics. The frequency of the heart rate is obtained by finding the dominant frequency in this modified frequency spectrum. The eliminated ranges are determined on the basis of the results of the preliminary experiments.

Proposed Method 2: In the case that a harmonic of the respiratory signal and the frequency of the heart rate are close to each other, proposed method 1 also eliminates the heart rate signal. Thus, some ratio of the intensity of the fundamental respiratory signal is subtracted from those ranges instead of setting the intensity in the ranges to zero. Specifically, 20% of the respiratory signal is subtracted from the range of the second harmonic and 10% of the respiratory signal is subtracted from the ranges of the third and the fourth harmonics. These ratios (20%, 10%) are determined using preliminary experiments.

These two methods are compared with the conventional method to investigate the validity of our proposed methods.

3. Flexible Tactile Sensor Sheet

To conduct experiments according to the proposed methods, a sensor that can measure the body pressure of a person lying on a bed at a high resolution is required. To measure the small pressure fluctuation of the heartbeat signal, a high- pressure resolution sensor is required. In addition, to identify the type of lying posture and to measure the body pressure at appropriate positions, a high spatial resolution sensor is required. A Smart Rubber (SR) sensor [15] (Sumitomo Riko, Ltd.), which is a flexible and stretchable planar tactile sensor, made of rubber-based materials, including the wiring, was used. Hence, it was sufficiently soft and elastic and was able to measure body pressure without causing the subject to feel any discomfort. The body pressure could be measured non-invasively, and without restraint.

Fig. 4 shows an outline of the structure of the SR sensor comprising two electrode layers made of conductive rubber sheets and a thin dielectric layer made of urethane. The electrode layers have many long narrow belt-like printed electrodes. The dielectric layer is sandwiched between the two electrode layers. The belt-like electrodes in the two electrode layers are orthogonal to each other, and an electric capacitance sensor cell is formed at every intersection of the electrodes. When one layer has *m* electrodes, and the other layer has *n* electrodes, m x n sensor cells are formed as a whole. When the widths of the electrodes were narrowed, high spatial resolution measurement was realized. In contrast, when the widths were widened, a high-pressure resolution measurement was realized.

When a vertical force is exerted on the SR sensor, the dielectric layer is deformed to be thinner. The electrostatic capacitance of the sensor cell increases because the distance between the two electrode layers becomes shorter. The pressure distribution on the SR sensor was calculated by measuring the electrostatic capacitances of all cells. Many pressure sensors having the same mechanism have been proposed. For instance, XSENSOR is commercially available [16]. However, this sensor is expensive and is less stretchable owing to its metal electrodes. It is, thus, not suitable for use in medical and nursing care fields. However, the SR sensor is suitable for these fields. As this sensor is made of rubber printing, it is affordable and stretchable.

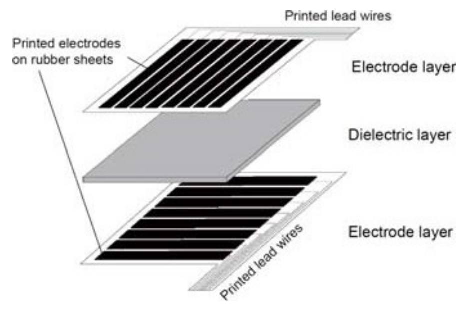


Fig. 4. Outline of the SR sensor

structure.

Table 1 summarizes the specifications of the SR sensor used in this study. The size of a sensor cell enables sufficient spatial resolution to distinguish sleep postures. In contrast, when the sensor cell is smaller, the pressure resolution deteriorates. This sensor can measure respiration without any problems because of large pressure fluctuations. However, the heart rate cannot be measured from the noise because the pressure fluctuation is small. Therefore, to measure the heart rate, increasing the pressure resolution by suppressing the noise is critical. The signal-to-noise ratio can be improved by sampling many times, compared to the signal period at high frequency and by obtaining the average of these signals. This technique is called oversampling. In this study, the noise is suppressed by taking an average of 46 samples that were collected.

Table 1. Specifications of the SR Sensor

Number of sensor cells	800	
	Length	Width
	32	25
Size of the whole sensor	882 [mm]	686 [mm]
Size of a sensor cell	14 [mm]	14 [mm]
Size of a gap between cells	14 [mm]	14 [mm]
Thickness of the sensor	3.5 [mm]	
Measuring frequency	20 [Hz]	

It is possible to measure the heart rate by oversampling. However, it takes a considerable amount of time to oversample all the cells of the sensor; thus, the sampling rate required for measuring the respiration and heart rates cannot be obtained. Therefore, the high-pressure resolution measurement using oversampling should be limited to certain cells. The high-pressure-resolution cells are referred to as the precision cells. In this study, the number of precision cells is four, owing to the limitation of the computational power of the microcomputer mounted on the sensor. As the number of precision cells is limited, they have to be appropriately positioned to measure the heart rate.

4. Experiments

4.1. Experimental setup

To evaluate the effectiveness of the proposed methods, tests were conducted on subjects positioned on the SR sensor and laid on a bed. Eight males aged 21 to 35 years, were selected for this study. This study has been approved by the Research Ethics Committee of the Kyushu Institute of Technology, and an informed consent has been acquired from all the subjects.

The methods were applied to three types of sleeping postures: supine position, prone position, and lateral position. For every posture of each subject, ten measurements were taken. In each measurement, four precision cells were positioned at suitable locations. One hundred and twenty data points per subject were collected. For the FFT, 1,024-point data during 51.2 s were used (51.2 s x 20 Hz = 1,024).

To calculate errors in the frequencies of the heart rate, the true values of the heart rate were collected using a biological information sensor (biosignalsplux) [17]. The subjects wore the sensor to enable measurements of the true values of the biological information. The sensor measures the subject's heart rate in an electrocardiogram. This sensor also measures respiration by measuring the movement of the chest.

4.2. Measurement results of heart r ate

Fig. 5 shows the measurement frequency errors of heart rate. The horizontal axis shows the label corresponding to every subject, and the right end indicates the average of all the subjects. The two proposed methods (described in Section II-B) and the conventional method [13] were applied. The vertical axis shows the averages of the absolute values of the errors evaluated by the three methods in the subjects. The error bars show the standard deviations of these values. The average errors of all subjects decreased in the following order: proposed method 2 (0.144 Hz), proposed method 1 (0.146 Hz), and the conventional method (0.149 Hz). The calculation of the paired t-test shows that the difference between the proposed method 2 and the conventional method is statistically significant at the significance level, $p = 0.05$ (significance probability: $p = 0.011$, t-ratio: $t = 2.29$). In contrast, the calculation of the paired t-test shows that the difference between the proposed method 1 and the conventional method is not statistically significant at the significance level $p = 0.05$ (significance probability: $p = 0.107$, t-ratio: $t = 1.25$). The proposed method 2 significantly improves the measurement of the heart rate.

However, the errors of subject-B and subject-G reduce in the following order: the conventional method, proposed method 2, and proposed method 1, potentially because the harmonics of the respiratory signal and the frequency of the heart rate of these two subjects may be close to each other. In this case, the proposed method 1 eliminates not only the harmonics of the respiratory signal, but also the heart rate signal. Furthermore, even with the proposed method 2, the peak of the heart rate signal is lower, thus increasing the possibility of measurement failure.

Table 2 summarizes the number of data points where the harmonics of the respiratory signal and frequency of the heart rate are close to each other for each subject. Data, where the absolute value of the frequency difference is 0.05 Hz or less, are considered to be close. Of the 120 data points per subject, the number of data points where the second, third, and fourth harmonics of the respiratory signal and frequency of the heart rate are close to each other is shown in the columns titled 2nd, 3rd, and 4th in Table 2. Subject-B and subject-G have more data where the harmonics of the respiratory signal and frequency of the heart rate are closer to each other than the other subjects. This is believed to be the reason for the measurement error of heart rate, which was greater with the proposed method than with the conventional method for these two subjects.

Table 2. Date points where the absolute value of the difference between the harmonics of the respiratory signal and the frequency of the heart rate is 0.05 Hz or less.

Subjects (A to H)	Harmonics			Number of data points
	2nd	3rd	4th	
A	0	0	0	0
B	0	4	40	44
C	0	0	8	8
D	0	0	0	0
E	0	0	8	8
F	0	0	0	0
G	0	12	28	40
H	0	0	8	8

5. Conclusion

Two methods were developed to improve the measurement of heart rate by eliminating the effects of the harmonics of the respiratory signal. The experiments showed that the error in heart rate measurement could be reduced by using the proposed method 2 than the conventional method [13]. Our proposed method has enhanced the precision of the heart rate measurement. This will lead to strengthening of measurements of the sleeping state.

For the subjects in whom harmonics of the respiratory signal and frequency of the heart rate were close to each other, the accuracy of measurement of heart rate decreased when the proposed methods were used. The proposed method cannot enhance the precision of the heart rate measurement for such subjects. Identifying such subjects is a future challenge. Additionally, if a parameter for the percentage of the respiratory signal that should be subtracted from the harmonic band in the proposed method 2 can be determined appropriately, this would enable the heart rate to be measured at the same level as the conventional method, even for such subjects. In future research, a method to determine the optimal parameter for each subject will be developed. The method will be able to avoid the deterioration of the accuracy of heart rate measurement.

6. References

[1] Statistics Bureau, Ministry of Internal Affairs and Communications, Japan. (2011, Oct.). The 2010 Population Census of Japan. Public announcement on 26 Oct. 2011. [Online]. Available: <http://www.stat.go.jp/english/data/kokusei/index.htm>, (accessed on 3rd Dec. 2020).

[2] National Institute of Population and Social Security Research. (2012, Mar.). Population Projections for Japan (January 2012) - 2011 to 2060-. Public announcement on 30 Mar. 2012. [Online]. Available: <http://www.ipss.go.jp/site-ad/index english/ esuikei/gh2401e.asp>, (accessed on 3rd Dec. 2020).

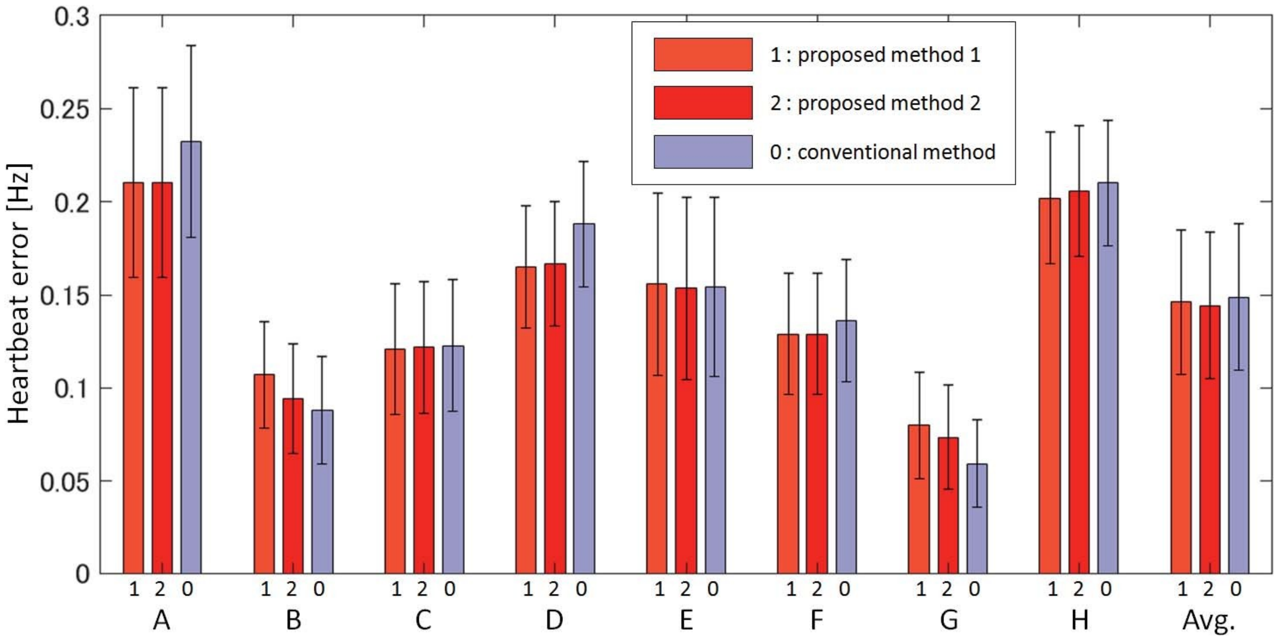


Fig. 7. Measurement frequency errors of heart rate of the two proposed methods and the conventional method. The proposed method 1 eliminates the harmonics of a respiratory signal. The proposed method 2 reduces the effects of the harmonics.

[3] Ministry of Health, Labour and Welfare, Japan. (2011, June). General Welfare and Labour. Public announcement on 2 June 2011. [Online]. Available: <http://www.mhlw.go.jp/english/policy/ other/social-security/index.html>, (accessed on 3rd Dec. 2020).

[4] C. A. Kushida, M. R. Littner, T. Morgenthaler, C. A. Alessi, D. Bailey, J. Coleman Jr, L. Friedman, M. Hirshkowitz, S. Kapen, M. Kramer, T. Lee-Chiong, D. L. Loube, J. Owens, J. P. Pancer, and M. Wise, "Practice parameters for the indications for polysomnography and related procedures: An update for 2005," *Journal of Sleep*, vol. 28, no. 4, pp. 499–521, 2005.

[5] A. Rechtschaffen and A. Kales, *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Bethesda, MD. : U.S. Dept. of Health, Education, and Welfare, Public Health Services-National Institutes of Health, National Institute of Neurological Diseases and Blindness, Neurological Information Network, 1968.

[6] T. Watanabe and K. Watanabe, "Estimation of the Sleep Stages by the Non-restrictive Air Mattress Sensor: Relation between the Change in the Heart Rate and Sleep Stages," *Trans. of the Society of Instrument and Control Engineers*, vol. 37, no. 9, pp. 821–828, 2001.

[7] Y. Kurihara and K. Watanabe, "Sleep-Stage Decision Algorithm by Using Heartbeat and Body-Movement Signals," *IEEE Trans. on Systems, Man, and Cybernetics, Part A*, vol. 42, no. 6, pp. 1450–1459, 2012.

[8] Y. Kambayashi and H. Hagiwara, "Estimating Sleep Cycle Using Body Movement Density," in *Proc. 2012 5th Int'l Conf. on Biomedical Engineering and Informatics*, pp. 1029–1033, 2012.

[9] S. Okada, Y. Ohno, Goyahan, K. Kato-Nishimura, I. Mohri, and M. Taniike, "Examination of Non-restrictive and Non-invasive Sleep Evaluation Technique for Children Using Difference Images," in *Proc. 30th Annual Int'l Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 2483–2487, 2008.

[10] A. Deguchi and H. Hagiwara, "Evaluation of Heart Rate Variability and Respiratory Variability during Sleep Using a Lorenz Plot," *Journal of Sleep and Biological Rhythms*, vol. 9, no. 4, p. 366, 2011.

[11] R. M. Harper, V. L. Schechtman, and K. A. Kluge, "Machine Classification of Infant Sleep State Using Cardiorespiratory Measures," *Electroencephalography and Clinical Neurophysiology*, vol. 67, no. 4, pp. 379–387, 1987.

[12] T. Mukai, K. Matsuo, Y. Kato, A. Shimizu, and S. Guo, "Lying Posture Detection for Unconstrained Measurement of Respiration and Heartbeat of a Person on a Bed," in *Proc. 11th IASTED Int'l Conf. on Biomedical Engineering*, pp. 46–53, DOI: 10.2316/P.2014.818-014, 2014.

[13] K. Matsuo, T. Mukai, Y. Kato, A. Shimizu, and S. Guo, "Measurement of Respiration and Heartbeat Using a Flexible Tactile Sensor Sheet on a Bed," in *Proc. 11th IASTED Int'l Conf. on Biomedical Engineering*, pp. 59–64, DOI: 10.2316/P.2014.818-031, 2014.

[14] J. W. Cooley and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series," *Mathematics of Computation*, vol. 19, no. 90, pp. 297–301, 1965.

[15] S. Guo, Y. Kato, H. Ito, and T. Mukai, "Development of Rubber-Based Flexible Sensor Sheet for Care-Related Apparatus," *Sumitomo Electric Industries Technical Review*, no. 75, pp. 125–131, 2012.

[16] R. Cork, "XSENSOR Technology: a Pressure Imaging Overview," *Sensor Review*, vol. 27, no. 1, pp. 24–28, 2007.

[17] PLUX Wireless Biosignals S. A., Biosignalsplux. [Online]. Avail- able: <https://plux.info/12-biosignalsplux>, (accessed on 3rd Dec. 2020).

Image Registration Method for Chest MDCT Images Based on 2-D Finite Element Method

Takuji Ogimoto, Tohru Kamiya
Department of Mechanical and Control Engineering, Kyushu Institute of Technology, 1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka, 804-8550, Japan

Takatoshi Aoki
University of Occupational and Environmental Health, 1-1, Iseigaoka, Yahatanishi, Kitakyusyu-shi 807-8555, Japan

Abstract

Multi detector-row computed tomography (MDCT) device has been used for early detection of lung cancer. However, there is concern that an increase in the burden on doctors will be caused by improvements in CT performance. Therefore, developing a computer aided diagnosis (CAD) reduces the burden on doctors. A temporal subtraction (TS) technique is one of the CAD systems and it is the subtraction operation of the current image and the previous one of the same patients that emphasizes temporal changes. In the TS image, when the position of the current image and the previous image are misaligned, subtraction artifacts are remained. In this paper, we propose a registration method based on the 2-D finite element method (FEM). In our proposed method, to improve the high computed cost that is the biggest problem of the FEM, we introduce the 2D FEM registration. We applied this method to 31 series MDCT image sets which was obtained previous and current from the same subjects and evaluated. As a result, full width at half maximum (FWHM) of 28.0, artifact to lung volume ratio of 5.77% and computational time of 140 sec were obtained.

Keywords: Computer Aided Diagnosis System, Temporal Subtraction Technique, Non-Rigid Image Registration, Finite Element Method.

1. Introduction

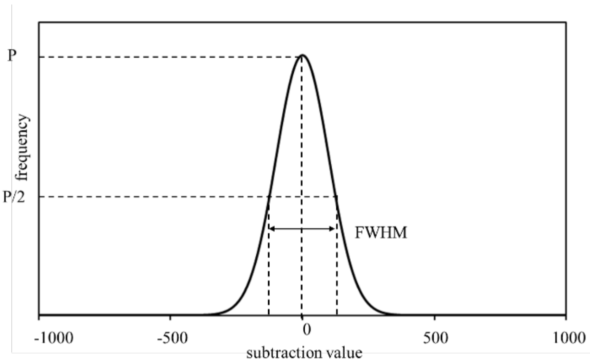
In recent years, multi detector-row computed tomography (MDCT) of the chest device has been used for the early detection of lung cancer. Chest MDCT images can detect lung cancer at an earlier and more curable stage than routine chest radiography. However, while MDCT images can be obtained in a short period of time and have a high resolution, there are problems that the frequency of CT scans and the number of images increase. To solve these problems, computer aided diagnosis (CAD) has been attracting attention. The results of the analysis by the CAD system are provided to the radiologists as a "second opinion" [1]. By using the CAD system, it is expected to shorten the time required to make a diagnosis and eliminate the variability in diagnosis by doctors. In this paper, we propose a registration method for the generation of temporal subtraction (TS) images from CT images of the chest, which is a CAD system. A TS technique is the subtraction operation of the current image and the previous one of the same patients and emphasizes temporal changes [2]. The TS technique can emphasize interval changes such as new lesion shadows and shape changes in the current image by removing the normal structures (blood vessels, bones, muscles etc.) that appear in both images that do not change much over time. Such a technique has been shown to be effective in detecting nodular shadows by physicians in a clinical evaluation experiment [3].

In general, subtraction artifact in TS images is a normal structure that remains in the temporal subtraction image without being erased by the differencing process. Also, artifacts can be caused by misalignment between current and past images. Misalignment is caused by various factors such as the photographic equipment and the photographic environment. Therefore, we can obtain a subtraction image in which the normal structure is removed and only the abnormal structure is emphasized by deforming the previous image to the current image and computed the difference between the two images. Such an image registration technique is necessary to generate the TS images.

In this paper, we propose a registration method for reducing the subtraction artifacts on TS image by using the finite element method (FEM). In a previous study [4], to reduce the computational cost (processing time and memory) required for the FEM registration use other simple alignment methods [5] for generating a temporal difference image. After that a local matching technique was performed on to the high-density regions (i.e., abnormal regions or subtraction artifacts) on the TS images. In this paper, to further reduce the computational cost compared to the previous studies, we applied 2-D FEM to each slice instead of 3-D FEM to reduce the computational cost. This can be expected to reduce the processing time and to generate highly accurate TS images. In our experiments, we apply the proposed method to chest MDCT images and verify its effectiveness in terms of both positioning accuracy and calculation time.

2. Proposed Method

Our proposed method consists of three major phases: create an initial temporal subtraction image, mesh of triangle elements, 2-D FEM registration for final subtraction image. Flow is shown in Fig 1. In the first phase, extract the regions to be aligned using the temporal subtraction technique. In the second phase, create the mesh for FEM for the extracted regions. In the final phase, 2-D FEM registration is performed.



2.1. *Initial temporal subtraction image*

Itai et al. [5] proposed a method based on voxel matching technique for creating a temporal subtraction image. In this study, to reduce the computational time, we made an initial subtraction image based on voxel matching technique. In general, when we implement the FEM technique on the image for registration, it required computational time. To avoid this problem, we performed the FEM technique on the initial subtraction areas.

2.2. *Creating the mesh of triangle elements*

We set the region of interest (ROI) around the detected regions on initial temporal subtraction image that meets the conditions. The conditions are the subtraction value is 200 over and volume is over $50mm^3$. We installed node point at equal intervals and created tetrahedron elements by using Delaunay triangulation [6].

2.3. *Creating the final subtraction image*

Ferrant et al. [7] proposed a registration method using FEM. They expressed potential energy E as:

$$E = \int_{\Omega} \sigma^T \cdot \epsilon d\Omega - \int_{\Omega} [I_1(x+\delta(x)) - I_2(x)]^2 d\Omega. \quad (1)$$

where Ω is the body on which one is working, σ is the stress vector, ϵ is the strain vector, I_1 is deformation image, δ is displacement vector, and I_2 is current image, respectively.

From principle of minimum potential energy, they sought minimum of this function by solving $\frac{\partial E}{\partial \delta} = 0$. Eq. (1) then becomes:

$(\boldsymbol{K}+\boldsymbol{G})\boldsymbol{\delta}=\boldsymbol{P}$. (2)

Matrix $\boldsymbol{K}$, $\boldsymbol{G}$, and $\boldsymbol{P}$ are defined as follows: $\boldsymbol{K}=\int_{\Omega}\boldsymbol{B}^{T}\boldsymbol{D}^{T}\boldsymbol{B}\,d\Omega$, $G=\int_{\Omega}\boldsymbol{N}^{T}\nabla\boldsymbol{I}_1^{T}\nabla\boldsymbol{I}_1\boldsymbol{N}d\Omega$, and $\boldsymbol{P}=\int_{\Omega}(I_1-I_2)\nabla\boldsymbol{I}_1\boldsymbol{N}d\Omega$; where $\boldsymbol{B}$ is strain displacement function, $\boldsymbol{D}$ is D matrix, N is shape function. By solving Eq. (2), the displacements of all nodes are determined and the displacements inside the elements are determined. Then we obtained warped images. We performed FEM registration on a slice by slice to reduce the computational cost.

3. Result

In this paper, we evaluated the performance of the image registration method using the 2-D FEM. Two experiments were conducted. Experiment 1 confirmed the effectiveness of the method using synthetic data. In Experiment 2, previous and current chest CT of the same subject were set as a set and registration processing was performed.

3.1. *Quantitative evaluation on synthetic data*

We confirmed the effectiveness of the method using synthetic data. The previous image is rotated by 5 degrees to become the current image, and the previous image is warped and registered to create the temporal subtraction image.

As the evaluation method, Full Width at Half Maximum (FWHM) and subtraction artifact to lung volume ration were used. FWHM represents the spread of the histogram and smaller values of FWHM indicate fewer artifacts. The offset width is 71. An example of a concentration histogram is shown in Fig. 2.

In the experiment, we applied this method to 31 cases. As a result of the experiment, FWHM: 12.1, artifact to lung volume ratio: 2.70%, cost time: 22sec were obtained. Table 1 shows the result.

3.2. *Quantitative evaluation on real data*

The previous and current chest CT of the same subject was set as one pair and the registration process was performed. We applied the proposed method to all 31 pairs and found the result. From Table 2, FWHM: 28.0, artifact to lung volume ratio: 5.77%, cost time: 140 were obtained.

4. Discussion and conclusion

In this study, we proposed a 2-D FEM registration. The result of the 2-D FEM registration method was obtained with FWHM: 28.0, artifact to lung volume ratio: 5.77%, cost time: 140sec. Table 2 shows the result of the previous method [4] and the proposed method. From the Table 2, the value of FWHM and artifact to lung volume ratio for the previous method and the proposed method is almost the same, and the proposed method reduces computational time. Therefore, the proposed method improved cost time without decreasing registration accuracy.

The previous method performed 3D-FEM registration and has high registration accuracy. So, we need to develop a more accurate registration method.

In this paper, we developed the registration method for chest MDCT images based on 2-D FEM. Our proposed method has three phases which are temporal subtraction technique, creation of the mesh of triangle elements, and registration based on 2-D FEM. We evaluated our method by FWHM and artifact to lung volume ratio, and cost time and obtained. Furthermore, the result

	FWHM	Artifact to lung volume ratio [%]	Cost time [sec]
Previous method [4]	28.1	5.70	228
Proposed method	28.0	5.77	140

Table 2 Experimental 2 result

FWHM:28.0, artifact to lung volume ratio: 5.77%, and cost time: 140 sec was obtained.

(a)previous image(b)rotated image (c)output image

Fig. 3 Experimental result: rotated image

(a)previous image (b)current image (c)output image

Fig.4 Experimental result

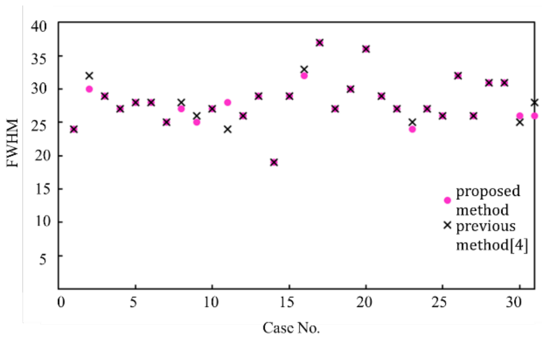


Fig. 5 Experimental 2 FWHM result

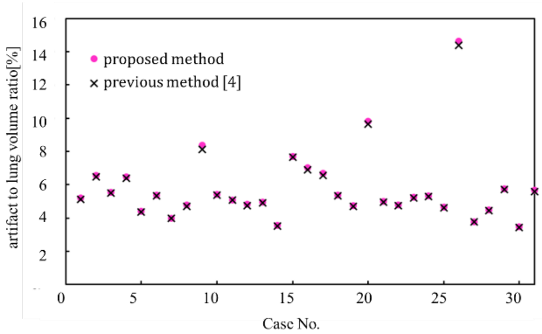


Fig. 6 Experimental 2 artifact to lung volume ratio result

Table 1 Experimental 1 results

	FWHM	Artifact to lung volume ratio [%]	Cost time [sec]
Previous method [4]	12.1	2.68	45
Proposed method	12.1	2.70	25

References

1. K. Doi, Current status and future potential of computer-aided diagnosis in medical imaging, *British Journal of Radiology*, Vol. 78, pp. S3-S19, 2005.
2. 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.
3. T. Aoki et al., Temporal Subtraction Method for Lung Nodule Detection on Successive Thoracic CT Soft-Copy Images, *Radiology*, Vol. 27, No. 1, pp. 255-261, 2014.
4. S. Yamada et al., Three-dimensional non-rigid registration of thoracic CT image based on finite element method, Joint International Conference on Soft Computing and Intelligent Systems (SCIS) and International Symposium on Advanced Intelligent Systems (ISIS), pp. 1-6, 2014.
5. Y. Itai et al., Development of A Voxel Matching Technique For Substantial Reduction of Subtraction Artifacts in Temporal Subtraction Images Obtained from Thoracic MDCT, *Journal of Digital Imaging*, Vol. 23, No. 1, pp. 31-38, 2010.
6. L.P. Chew, Constrained Delaunay triangulations, *Algorithmica*, Vol. 4, No. 1, pp. 97-108, 1989.
7. M. Ferrant et al., 3D Image Matching Using a Finite Element Based Elastic Deformation Model, Lecture Notes in Computer Science, Vol. 1679, pp. 202-209, 1999.

Detection of Abnormal Shadows in Low-dose CT Images Using CNN

Hiromu Ikeda, Tohru Kamiya
Department of Mechanical and Control Engineering, Kyushu Institute of Technology, 1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka, 804-8550, Japan

Takatoshi Aoki
University of Occupational and Environmental Health, 1-1, Iseigaoka, Yahatanishi, Kitakyusyu-shi 807-8555, Japan

Abstract

CT imaging is a very effective way to detect abnormal shadows in the lungs. Considering the physical burden on the patient, it should be taken at a low dose. However, there are currently few studies using low-dose CT as compared to normal-dose CT. Therefore, it is necessary to develop a system to support the diagnosis of low-dose CT. To detect lung nodules and GGO from a low-dose CT image, we propose image processing techniques based on CNN. Our approach consists of two main elements. First, candidate abnormal shadow areas are extracted from low-dose CT images of the chest by threshold processing and filtering. Second, we implement a CNN model to classify candidate region on images. To improve identification accuracy in CNN, we devised a method to add spatial information to the input image. In this paper, the proposed method is applied to 13 abnormal shadow images and 100 normal tissue images, and true positive ate of 84.6%, false negative rate of 15.4% were achieved respectively.

Keywords: Low-dose CT, Computer aided diagnosis, Dual-ring filter, Convolutional neural network, input image processing

1. Introduction

According to the world health organization (WHO), cancer is the second leading cause of death worldwide, and is responsible for an estimated 9.6 million people in 2018. Among them, lung cancer accounting 2.09 million cases, and is the large number of deaths cause. It is ranked the most common causes of cancer death [1]. Early detection is important factor to reduce mortality from cancer. If the cases are detected and treated in early stage, the mortality rate from cancer may decrease. When the cancer is detected in early stage, it has high survival rate, low morbidity, low treatment costs and high potential for effective treatment. Also, detecting the cancer in early time and avoiding delays in treatment may make a huge difference in the lives of cancer patients.

CT imaging is often used for early detection of the disease. However, there are some problems in CT imaging, such as an increase in the burden of the reader due to the huge number of images, and a difference in diagnostic results due to differences in the experience of physicians. To avoid this problem, there are some approaches in medical imaging fields. The CAD system digitizes the medical data, quantifies and analyzes the medical data using a computer. As a result, physician can use the results of the CAD as a "second opinion". In addition, the low-dose CT used in this paper, and is less burdensome on the patient than the one taken at a normal dose, however the image quality is poor.

In recent years, deep learning has attracted a lot of attention in the field of artificial intelligence. Especially in the field of image recognition fields, convolutional neural networks (CNNs) have become the most important technology to classified a target from an image. The various CNN architectures are also becoming increasingly important in the medical imaging field. In this paper, we describe the development of a CAD system for the detection of abnormal shadows from low-dose CT images. In this approach, candidate abnormal shadow areas are extracted from CT images and the regions are inputted into the CNN to classify normal or not.

2. Method

In this system, we extract candidate abnormal shadow areas from CT images based on CNN. The details are shown in following. Figure 1 shows an overview of the CAD system which is built in this paper. Figure 2 shows the flow of the initial candidate region extraction.

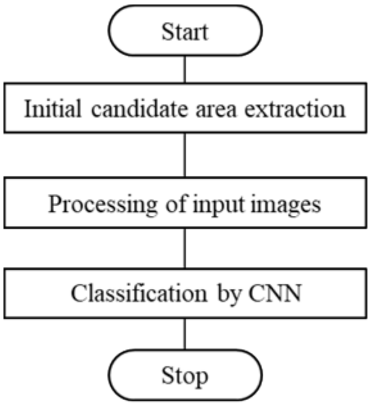


Fig. 1. The flow of CAD system

2.1. *Extraction of initial candidate area*

Lung field areas are extracted from CT images. In this paper, the areas were extracted by binarization, labeling, and thresholding techniques. Next, emphasis was applied to make it easier to detect abnormal shadows. The details are shown in follows.

A. *Segmentation of lung area*

To segment region of interest (ROI), we perform the Gaussian filter for smoothing, and binarization based on Otsu's method [2]. Next, the labeling process is performed, and the largest area detected as lung area.

B. *Enhance abnormal shadows based on double-ring filter*

Most of the abnormal shadows in CT images hascircular shape. In addition, because of the use of low-dose CT, the effect of noise should be considered. From these points of view, we use a double-ring filter [3] for emphasis processing, which can enhance the difference in pixel values between the center and the surrounding area to reduce the effect of noise. The output value of the double-ring filter is the average value of the inner pixel value minus the average value of the outer pixel value. In this paper, the 7×7 pixels of a 9×9 filter are considered.

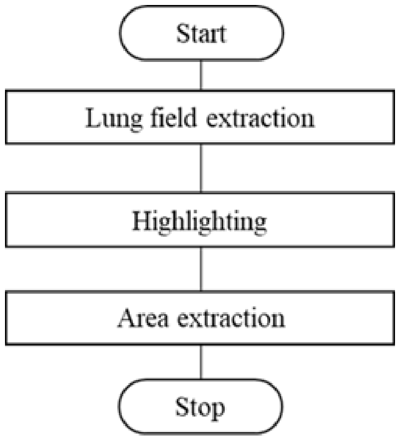


Fig. 2. initial candidate area extraction

2.2. *Input images to CNN*

To make it easier to identify the abnormal shadows, we added processing to the images input to the CNN. Specifically, the image before one slice of the candidate region was made into R components, the image of the candidate region was made into G components, and the image after one slice of the candidate region was made into B components on the CNN model. This allows us to obtain a two-dimensional image with spatial information. Figure 4 shows an example of the input image.

2.3. *CNN*

In this paper, we perform transfer learning based on the VGG-16 [4] network structure learned using ImageNet. In addition, we frozen weights up to convolutional layer 10 and retrained after that layers with the generated images. Also, we add a batch normalization [5] behind all convolutional layers to suppress the influence of bias in the data. To adjust the weights of the training, and to suppress over-learning and to obtain global features, we change all the coupled layers to global average pooling [6] and reconstruct the model as shown in Fig.5.

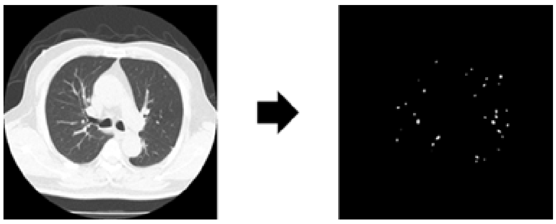


Fig.3 Extraction of candidate areas

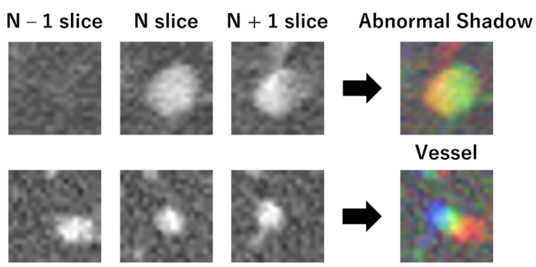


Fig.4 Input images to the CNN

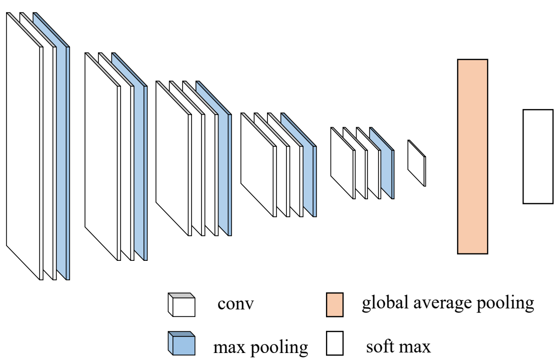


Fig.5 CNN

3. Results

3.1. *Detail of Dataset*

In this paper, we prepare 13 classes of data sets. One class consists of test and training data. The test data contains 1 abnormal shade and 6 normal tissues. The training data contains 156 abnormal shades and 936 of normal tissue. The training data contains augmented data. Data augment includes vertical and horizontal inversion, mirror image inversion, contrast adjustment and gamma correction.

3.2. *Evaluation metrics*

The performance is evaluated using leave-on-out cross-validation. Classification accuracy is evaluated by the following equations: TPR(True Positive Rate) %, and FPR(False Positive Rate) %, respectively. Here a,b,c,d are shown in Table.1.

$$TPR = \frac{a}{a+b} \tag{1}$$

$$FPR = \frac{c}{c+d} \tag{2}$$

3.3. *Results*

In this paper, we use the CNN model described in 2.3. We discuss the case where the inter-slice information is reflected in the input image (proposed method) and the case where it is not reflected in the input image as described in 2.2. Table 2 shows classification results. In Table 2, the case where the inter-slice information is reflected in the input image is shown as the proposed method, and the case where it is not reflected is shown as denoted as baseline. As a comparison, we also performed with a method of classifying the SVM [7] with features extracted using CNN, which is effective when the number of images in the dataset is small. The proposed method achieved a TPR of 84.6%, FPR of 15.4%. In contrast, when the input image does not reflect the inter-slice information, the TPR of 53.8%, FPR 15.4%. The SVM is used as another classifier, the TPR of 53.8%, FPR 1.3% was obtained.

4. *Discussion*

We used a CNN model to discriminate the presence or absence of lung cancer in low-dose CT images. The TPR of the proposed method was 30.8% higher than that of the baseline. This suggests that reflecting inter-slice information in the input image may be useful. In some cases, abnormal shading could not be detected. The reason may be that what could not be detected was very faint. This can be solved by using an image with enhanced shading in the input image. In addition, we would like to improve the structure of CNN in the future.

Table.1 Confusion matrix

		Prediction Class	
		positive	negative
Actual Class	positive	a	b
	negative	c	d

Table.2 Experimental results

	TPR%	FPR%
Proposed method	84.6	15.4
Baseline	53.8	15.4
SVM	53.8	1.3

5.*Conclusion*

We developed a CNN-based method for detecting abnormal shadows from low-dose CT images. We achieved a TPR of 84.6% and FPR of 15.4%. In future, we need to develop a shading enhancement method for the input images and improve the structure of the CNN to further improve the classification accuracy.

References

1. World Health Organization, Cancer, https://www.who.int/en/news-room/fact-sheets/detail/Cancer
2. Otsu, "A threshold selection method from gray-level histograms", IEEE Transactions on Systems, Man and Cybernetics, Vol.9, No. 1, pp.62-66, 1979.
3. Y. Hatanaka, et al., "Automated detection algorithm for arteriolar narrowing on fundus images," Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 286–289, 2006.
4. K. Simonyan, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv: 1409. 1556, 2014.
5. Sergey Ioffe Christian Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", arXiv:1502.03167, 2015.
6. Min Lin et al., Network In Network " , Cornell University Library, arXiv: 1312.4400, 2014.
7. B. E. Boser et al., "A training algorithm for optimal margin classifiers", Proc. 5th ACM Workshop on Computational Learning Theory, pp.144-152, 1992.

EEG Signal Extraction Method Based on HHT and CSP

Lei Gong, Chengzhi Liu

Tianjin University of Science and Technology, China;

*E-mail: * 1442064639@qq.com*

www.tust.edu.cn

Abstract

In this paper, the Hilbert Huang transform is used to extract the marginal spectrum of the four-task motor imagery of EEG signals. Firstly, a spatial filter constructed by common spatial pattern is used to extract the feature of signal marginal spectrum. The MI EEG Data2a provided by The University of GRAZ is verified by the experiment, and the results show that compared with the traditional single CSP feature extraction method, the effect of MSCSP feature extraction was significantly improved.

Keywords: motor imagery (MI), Hilbert Huang transform (HHT), common spatial pattern (CSP), marginal spectrum (MS)

1. Introduction

Electroencephalogram (EEG) is to record the spontaneous and rhythmic activity of brain cells through electrodes. EEG based brain computer interface (BCI) system does not depend on muscle nerve activity, but on EEG activity recorded at scalp [1]. As a new interdisciplinary subject, BCI technology has attracted the attention and research of many international and domestic researchers, and has shown broad application prospects in the fields of rehabilitation medicine engineering, auxiliary control and military [2]. With the rapid development of science and technology, although the quality of BCI technology has been improved rapidly, there is still a certain distance between BCI technology and practical application, among which the search for excellent feature extraction method is one of the keys.

The brain computer interface based on MI is a commonly used experimental paradigm of spontaneous EEG signals. The basic principle of MI is that when people imagine a certain limb of the body moving, according to the previous motor memory, the motor cortex of the brain will have corresponding potential changes [3]. Imagining hands, feet and tongue is a typical four-task in MI brain computer interface. EEG signal processing and analysis is the core part of the whole BCI system. Feature extraction is mainly through the analysis of EEG signal in time domain, frequency domain and spatial domain to find the features which are difficult to be found in the original EEG signal and can best reflect the thinking activity [4]. The existing feature extraction methods mainly include power spectral density (PSD), independent component analysis (ICA), discrete wavelet transform (DWT), Hilbert Huang (HHT) and common space pattern (CSP)[5,6,7,8].

In this paper, the marginal spectrum (MS) of four-task motor imagery of EEG is extracted by Hilbert Huang transform (HHT). For the first time, the common spatial pattern (CSP) is used to construct a spatial filter to extract the marginal spectral features (MSCSP) of the signals. Combining with the classification method of linear discriminant analysis (LDA), the classification accuracy of EEG signals is improved, which creates conditions for the practical application of BCI system.

2. Experimental Data

The data set selected in this paper is the BCI competition IV dataset2a motor imagery data set in the fourth brain computer interface competition in 2008, provided by Graz University of science and technology, Austria [9]. The data set was collected from 9 subjects. Each subject performed four-task motor imagery tasks: left hand, right hand, double feet, and tongue. Each subject was given two experiments. The EEG data collected for the first time is used as training set, and the task tag is known. The EEG data collected in the second time is used as the test set, and the task tag is unknown. In each data set, the subjects conducted 6 rounds of experiments, and 48 groups of data were collected in each round. Therefore, a total of 288 groups of experimental data were collected for each subject [10].

The whole process of the experiment is shown in Fig.1.The first stage is the preparation stage, i.e., when t = 0s, a cross symbol appears on the screen and a sound prompts the subject to prepare to start the experiment; the second stage is the motor imagery task prompt stage, that is, when t = 2s, a task prompt arrow appears on the screen. At this time, the subjects were reminded to start to prepare for the motor imagery task; the third stage was the motor imagery stage, that is, when t = 3s, the motor imagery task was started, lasting for 6s until the cross symbol on the screen disappeared. The fourth stage is the rest stage. After a short interval, the next round of experiment is carried out.

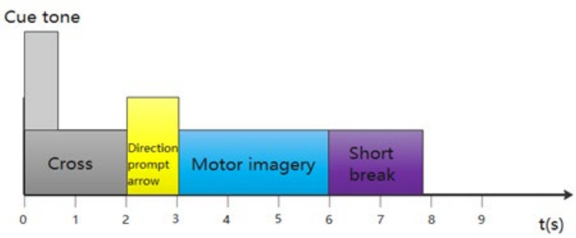


Fig.1. Experimental flowchart

3. Hilbert Huang Transform

HHT includes two parts: Empirical Mode Decomposition (EMD) and Hilbert transform[11,12]. It retains the advantages of time-frequency local transform in discrete wavelet transform and overcomes the problem of no adaptive basis function. This method of adaptive EEG decomposition does not necessarily require automatic signal processing.

After HHT transformation, the three-dimensional spectrum of time amplitude frequency is obtained. As shown in Fig.2, the three-dimensional spectrum of an EEG signal after HHT transformation can obtain the time-frequency information of EEG signal more visually.

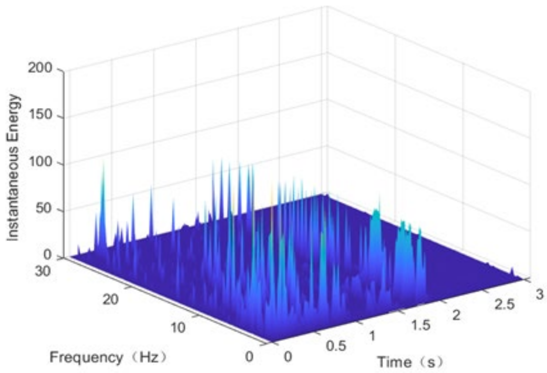


Fig.2. Three-dimensional spectrum of EEG signal after HHT transformation

4. Improved CSP Algorithm Based on Marginal Spectrum

The traditional CSP algorithm is operated directly in the time domain signal. It is difficult to train a good adaptability projection matrix on account of the complexity of EEG signal time-domain waveform change. When the brain carries out motor imagery, obvious ERD and ERS phenomena will appear in the C3, C4 and CZ regions of the cerebral cortex [13]. This phenomenon will cause obvious changes in the amplitude of the rhythm (8-13hz) and β rhythm (14-25hz) of the signals in this region. This frequency phenomenon has more obvious differences than the time-domain signals. Therefore, this paper proposes a method combining CSP algorithm with signal marginal spectrum (MS) algorithm (MSCSP). The algorithm combines the good binary classification effect of CSP algorithm and the good time-frequency characteristics of the marginal spectrum, so the MSCSP feature matrix has better adaptability.

4.1. Marginal spectrum feature extraction

In order to obtain the marginal spectrum, EMD decomposition is used to divide the complex signal into simple IMF single component signal, so that the concept of instantaneous frequency can be applied. Then the instantaneous frequency is analyzed by Hilbert transform, and the Hilbert spectrum is integrated with time to obtain the Hilbert marginal spectrum[14,15]. Therefore, MS is defined as follows:

$$h(w) = \int_0^T H(w,t)dt \quad (1)$$

Where T is the length of the signal. four-task of motor imagery data, including left hand, right hand, double feet, and tongue, are selected from the standard data set, and the marginal spectra of C3, C4 and CZ channel data of these four tasks are extracted by HHT transformation, as shown in Fig.3.

It is obvious from the Fig.3 that when the subjects are imagining the movement of their left and right hands, feet and tongue, obvious ERD and ERS phenomena appear in C3, C4 and CZ channels, and the amplitude of the edge spectrum changes with the change of frequency in the whole frequency band. Although it is similar to the spectrum, its meaning is different from that of the frequency spectrum, which represents the value of each frequency amplitude in the whole data set. It can accurately reflect the actual frequency component of the signal. The amplitude of a point in the spectrum proves that there is a trigonometric function of the frequency in the whole signal. The larger the amplitude, the higher the matching degree between the trigonometric function of the frequency and the original signal. By comparing the difference between the marginal spectrum and Fourier transform, it can be found that the marginal spectrum can not only process the stationary signal, but also the non-stationary signal. We can sensitively find the energy transformation of a certain frequency in the signal. So the marginal spectrum can reflect the actual frequency component of the signal more accurately.

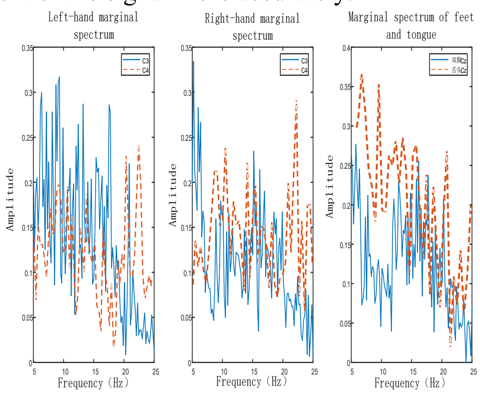


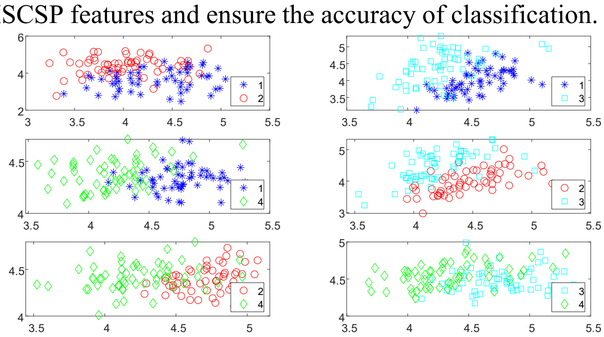
Fig.3. Marginal Spectra of Four Kinds of Motion Imagination

4.2. Improved algorithm of CSP

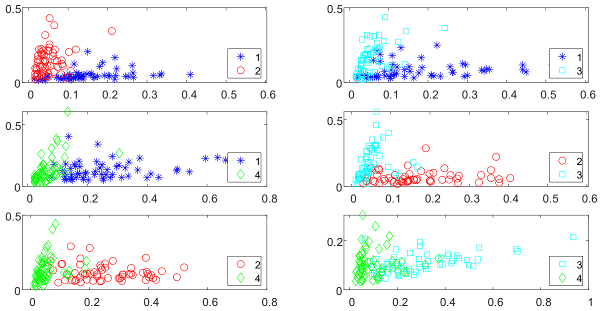
In this paper, the "one-to-one" CSP feature extraction method is used to process four kinds of task motor imagery EEG signals. In this paper, four-task of motor imagery signals are divided into two groups (left hand 1vs right hand 2, left hand 1vs feet 3, left hand 1vs tongue 4, right hand 2vs feet 3, right hand 2vs tongue 4, feet 3vs tongue 4). Six sets of CSP features are obtained by constructing six spatial filters through CSP, which are used as input of classifiers and the final output category.

After preprocessing the motor imagery data of the training set 4S, the marginal spectrum of each channel is obtained by HHT transformation. The CSP projection matrix and training set features are obtained by CSP operation of the marginal spectrum of each channel. For the test set data, after the same preprocessing, the marginal spectrum of each channel is obtained by HHT transformation, and the test set characteristics can be obtained by multiplying it with the CSP projection matrix.

The MSCSP characteristics of EEG signals before and after marginal spectrum processing are shown in Fig.4, in which Fig.4(a) is the six groups of CSP feature maps without obtaining the marginal spectrum by HHT transformation, and Fig.4 (b) is the six groups of MSCSP feature maps with marginal spectrum obtained by HHT transformation. From the feature map, it can be found that the CSP features of left hand and right hand, left hand and tongue, right hand and tongue, and tongue and feet before transformation have small distance and feature overlap seriously, so they can not be well distinguished. After the marginal spectrum is obtained by HHT transform, the interval between the two kinds of MSCSP features increases and the feature overlap decreases, which can distinguish the two kinds of signals well. Therefore, the marginal spectrum EEG signal processing method based on HHT transform can improve the discrimination of MSCSP features and ensure the accuracy of classification.



(a)



(b)

Fig.4. (a) CSP feature map before marginal spectrum processing, (b) MSCSP feature map before marginal spectrum processing

5. Analysis of Experimental Results

After preprocessing the BCI competition IV dataset 2a data set, the EEG signals were divided into four groups according to the motor imagery tasks of left hand, right hand, double feet and tongue. Each group contained 72 single experiments. In each experiment, the EEG signal components during motor imagery, i.e., the data from 3s to 6s, were intercepted, and the power spectral density (PSD), discrete wavelet coefficients (DWT), common space pattern features (CSP) and the improved algorithm MSCSP were extracted respectively. SVM classifier was used for feature classification. The final classification results are shown in Table 1.

Table 1. Comparison of classification accuracy of features extracted by each algorithm

Subjects	Feature extraction method (accuracy%)			
	PSD	DWT	CSP	MSCSP
A01T	44.83	51.72	71.43	67.86
A02T	43.38	48.28	53.57	57.14
A03T	41.38	68.97	76.79	83.93
A04T	20.69	41.38	50.89	58.93
A05T	43.10	20.69	33.57	41.07
A06T	25.86	31.03	42.14	55.36
A07T	46.55	51.72	68.21	67.86
A08T	48.83	55.17	80.18	82.14
A09T	44.83	48.28	63.75	66.07
average	39.93±9.77	46.36±14.01	60.05±15.98	64.48±12.61

From the final classification results in Table 1, under the same preprocessing conditions, it can be found that the method of extracting PSD of signal correlation frequency band has the lowest classification effect. Secondly, DWT is used to extract low-frequency coefficients of signals, which has no good classification effect. In time domain, the effect of CSP feature extraction is much better than the first two methods. However, compared with the improved MSCSP algorithm, the final classification result is still relatively low. Therefore, compared with the other three algorithms, the improved MSCSP algorithm can extract EEG features and has great advantages.

6. Conclusion

In the BCI system based on motor imagery, effective feature extraction and classification is the key to the system. This paper improves the CSP algorithm, uses HHT to extract the marginal spectrum of EEG signals, and selects the marginal spectrum as the input to extract the features of CSP. Based on the data provided by GRAZ University, the paper analyzes the feature extraction effect of the improved MSCSP algorithm, and finds it is better than the traditional CSP algorithm. The feature extraction method in this paper is in line, it is with the current development direction of BCI, and provides an experimental basis for application of BCI system.

Acknowledgements

References

1. Blankertz B， Tomioka R， Lemm S， et al． Optimizing spatial filters for robust EEG single-trial analysis． IEEE Signal Processing Magazine， 2007， 25(1)：41-56．
2. Emanuel N ， Felix B ， Harald A ， et al． Regularized Linear Discriminant Analysis of EEG Features in Dementia Patients． Frontiers in Aging Neuroence， 2016， 8：273-．
3. Haas， L F． Hans Berger (1873-1941)， Richard Caton (1842-1926)， and electroencephalography． J Neurol Neurosurg Psychiatry， 2003， 74(1)：9．
4. Vidal Toward direct brain-computer communication． Annual Re- view of Biophysics and Bioengineering， 1973， 2(1)：157．
5. Tiwari N， Edla D R， Dodia S， et al． Brain computer interface： A comprehensive survey． Biologically Inspired Cognitive Architectures，2018，26：118-129．
6. Dong E Z， Li C H， Li L T， et al． Classification of multi-class motor imagery with a novel hierarchical SVM algorithm for brain-computer interfaces． Medical & Biological Engineering & Computing， 2017， 55(10)：1809-1818．
7. Jeunet C， Lotte F， Batail J M， et al． Using recent BCI literature to deepen our understanding of clinical neurofeedback： A short review． Neuroscience，2018，378： 225-233．
8. Naeem M， Brunner C， Leeb R， et al． Seperability of four-class motor imagery data using independent

components analysis． Journal of neural engineering，2006， 3(3)： 208．
9. Naeem M， Brunner C， Leeb R， et al． Seperability of four-class motor imagery data using independent components analysis． Journal of neural engineering，2006， 3(3)： 208．
10. Pfurtscheller G ， Neuper C ， Guger C ， et al． Current trends in Graz Brain-Computer Interface (BCI) research． IEEE Trans Rehabil Eng， 2000， 8(2)：216-219．
11. Dong E ， Zhou K ， Tong J ， et al． A novel hybrid kernel function relevance vector machine for multi-task motor imagery EEG classification． Biomedical Signal Processing and Control， 2020， 60：101991．
12. Bauschke H H ， Combettes P L ． Convex Analysis and Monotone Operator Theory in Hilbert Spaces． 2011，10．1007/978-1-4419-9467-7．
13. Liu C ， Zhao H B ， Li C S ， et al． Classification of ECoG motor imagery tasks based on CSP and SVM 2010 3rd International Conference on Biomedical Engineering and Informatics． IEEE， 2010．
14. Yeo M V M ， Li X ， Shen K ， et al． Can SVM be used for automatic EEG detection of drowsiness during car driving． Safety Science， 2009， 47(1)：115-124．
15. Vladimir Cherkassky and Yunqian Ma． Practical selection of SVM parameters and noise estimation for SVM regression． Neural Networks， 2004．

Verification of CG character operation by brain wave discrimination

Kenji Sakoma, Kodai Miyamoto, Taketo Kamasaka, Makoto Sakamoto*, and Amane Takei
Faculty of Engineering, University of Miyazaki,
Miyazaki-City, Miyazaki, Japan

Tsutomu Ito
National Institute of Technology, Ube College
Ube-City, Yamaguchi, Japan

Takao Ito
Graduate School of Engineering, Hiroshima University
Higashi-Hiroshima, Hiroshima, Japan

E-mail: hm16011@student.miyazaki-u.ac.jp, hm15037@student.miyazaki-u.ac.jp, hm16043@student.miyazaki-u.ac.jp, fruits2000jp@yahoo.co.jp
**Corresponding Author*

Abstract

Recently VR technology is expected to develop in various fields. In this research, we verified whether the VR space can be operated just by thinking in the head by using a device that is cheaper than the device used in the existing research. We used the fast Fourier transform and the support vector machine as the methods of EEG analysis. In the future, we will try to improve it further by changing the method of EEG analysis and other conditions.

Keywords: EEG, BMI, FFT, SVM, VR

1. Introduction

Since 2016, which was called the "first year of VR (Virtual Reality)", VR technology has been researched and developed all over the world and has made remarkable progress. Various companies have begun to develop and sell VR equipment, and although it is a little expensive, even ordinary people can easily purchase it. VR technology has become widespread, as PCs and other devices that can use VR devices have become affordable even for individuals. As a result, VR technology is expected to develop in various fields such as medical care, education, business, and entertainment. However, there are some major challenges in the development of VR technology.

We paid attention to the troublesomeness of wearing in this. When using VR equipment, it is often necessary to set up cables and sensors in advance. Also, when using a VR device, it is necessary to have a certain amount of space around the user. This is to ensure the safety of the surroundings and the user himself / herself because the controller is always used when operating / moving in the world inside VR. Therefore, there is a problem that the place where the VR device can be used is limited. It tends to be more expensive than models that require cables, but recently, integrated (standalone) models that do not require cables have also been released. Therefore, the problem of setting up the cable or the like can be solved. However, the spatial problem when using VR equipment has not been solved.

Therefore, in order to try to solve the spatial problem, the operation by brain waves --- BMI (Brain-Machine Interface) is used. Since the frequency band of human brain waves, such as "α wave" and "β wave", fluctuates depending on the activity state, the purpose is to use them for controlling CG characters and robots by reading them with a device and discriminating them. It is a study. I think that it may be possible to solve the spatial problem when using VR equipment by using this method[1-4].

2. Research Method

2.1. *VR*

VR (Virtual reality) is a technology and system that creates an environment that is not the actual thing or the real thing but has the same essence as a function by stimulating the senses including the five senses of the user. Translated as "artificial reality" or "virtual reality" in Japanese. In the olden days, novels, paintings, plays, television, etc. also had more or less VR functions.

Several problems have been identified with VR technology, and these are problems that must be resolved for future development.

1. Troublesome installation problem
2. Spec problem
3. Social and market issues
4. Problems at the production site
5. Health issues

The above-mentioned "installation troublesome problem" includes a safe space problem that must be ensured in order to use the VR product that occurs when the VR product is used. This problem means that you can't play anywhere when using a VR product.

This study focused on this problem.

2.2. *Used equipment*

- Mind Wave Mobile

Fig. 1. Mind Wave Mobile.

Equipment sold by NeuroSky (see Figure 1). In this study, we will use this device to measure brain waves.

2.3. *Development environment*

OS	Windows 10
Programming language	C#
Measuring equipment	Mind Wave Mobile
Software	Visual Studio 2019 Unity 2019.2.18f

Table 1 . Development environment.

2.4. *Library*

- libStreamSDK

  Used to receive brainwave data from Mind Wave Mobile in Visual Studio 2019.
- MathNet.Numerics
- Accord
- Accord.MachinerLearning
- Accord.Math
- Accord.Statistics

  Used to process data received from Mind Wave Mobile in Visual Studio 2019.

2.5. *Brainwave measurement*

Use NeuroSky's Mind Wave Mobile[6] brain wave sensor to measure brain waves. In addition, in order to receive brain wave data in Visual Studio 2019, we used a library called libStreamSDK included in the development tool provided by NeuroSky.

2.6. *EEG processing method*

2.6.1. *FFT (Fast Fourier Transform)*

It is a type of DFT (Discrete Fourier Transform) and is an algorithm devised so that DFT can be executed at high speed on a computer (see Equation 1). When using this analysis method, the MathNet.Numerics library was used[7].

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i\frac{2\pi}{N}} \quad (1)$$

2.6.2. *EEG feature discrimination method*

For EEG characterization, SVM (Support Vector Machine), which is generally used for EEG characterization, was used. SVM is one of the pattern recognition models that uses supervised learning (see

Equation 2). SVM has high discrimination performance for untrained data. In this study, a Gaussian kernel (see Equation 3, 4) was used as the kernel function. In addition, 5 libraries of MathNet.Numerics, Accord, Accord.MachinerLearning, Accord.Math, and Accord.Statistics were used to use SVM[5].

$$y_{(x)} = \mathrm{sgn}\left\{\sum_{n=SV} w_n K(x_n, x) + b\right\} \quad (2)$$

$$= \begin{cases} 0 & x \in classA \\ 1 & x \in classB \end{cases} \quad (3)$$

$$K(x_1, x_2) = \exp(-\|x_1 - x_2\|^2 / 2\sigma^2) \quad (4)$$

2.6.3. Visual Studio 2019

Visual Studio 2019 processes the brain wave data read by Mind Wave Mobile and performs feature detection. At this time, FFT is used to detect the feature amount of the electroencephalogram data. Then, using SVM, the processed brain wave data is discriminated so that 0 is output when the brain wave is relaxed and 1 is output when the brain wave is conscious of moving.

2.6.4. Unity 2019.2.18f

We created CG characters and programs that act according to the characteristics of the read brain waves. The brain wave data was processed by Visual Studio 2019, the result of discrimination by SVM was received, and the CG character was created to act according to the discrimination result. The CG character uses Unity-Chan !, which is provided free of charge in Unity's Asset Store (see Fig. 2) .

Fig. 2. The CG character which used in Unity.

3. Experimental Method

Have the subject sit in a chair, wear Mind Wave Mobile on his head, and then start measuring brain waves. During the experiment, the subjects are made aware of relaxing and moving, and evaluate how the CG character in Unity moves accordingly. In this study, there are only two types of CG character movements: stop when you are relaxed and keep moving forward while you are thinking of moving your body.

4. Evaluation Experiment

In this experiment, three university students will be the subjects. An evaluation experiment was conducted and a questionnaire was conducted. The evaluation contents were "good points", "bad points", and "others" in a free description format.

5. Evaluation Result

The results of the questionnaire are as follows. In "Good points", "It is fresh and interesting to move just by thinking" was mentioned.

"Bad points" are "I was worried because there was a little time from thinking until the operation was reflected in the CG character", "Mind Wave Mobile is uncomfortable to wear", "The action that the CG character is thinking There were times when it did the opposite of what it was. "

"Others" included "Can I do anything other than move forward?" "Can I use running properly when walking?"

"Other" were that "is it not compatible with other movements?" and "is there no other CG character?",etc.

6. Consideration

From the "good points" of the evaluation results, it was found that the control of the CG character by brain waves, which is the purpose of this study, can obtain a certain result in distinguishing only two patterns of forward and stop.

From the "bad points", it was found that there was a problem with the accuracy and speed of EEG discrimination, and it was found that it was necessary to improve the EEG feature detection method and discrimination method, and to improve the performance of the processing device. It was also found that there was a problem with the installation of Mind Wave Mobile, but it is thought that the problem with this is that the size of the product used this time cannot be adjusted. From the above "bad points", it was found that there are points that should be further improved in this study.

From the opinion of "Other", it was found that actions other than advancing and stopping are necessary in this research.

7. Conclusion

In this research, we tried to solve the space problem around the user, among the "troublesomeness of wearing" which has become an issue in the development of VR technology. The brain wave data was measured by Mind Wave Mobile, the brain wave data from Mind Wave Mobile was received and processed by Visual Studio 2019, and the characteristics of the brain wave data processed by SVM were determined. We sent the discrimination result to Unity and verified the method in the field called BMI, which manipulates the characters in Unity based on the discrimination result.

From the results of the evaluation experiment, it was found that although the operation of the CG character by brain waves can obtain certain results, the accuracy is not sufficient and there are many points to be improved. Future tasks include improving the accuracy of brain wave feature detection and brain wave feature discrimination, increasing character movements by discriminating the ratio of brain wave frequency bands, and improving the uncomfortable wearing of Mind Wave Mobile. It is thought that there is, and we plan to work on it in the future.

If this research progresses, it is expected that it will play a major role in the medical field. By using this technology for people with physical disabilities who are physically handicapped or who are using artificial limbs, it will be possible to operate the body only with brain waves, which will be much more convenient than it is now. In addition, the field of e-sports is currently developing in the world, and VR technology is attracting attention in that field as well, so it is expected that if this research progresses, it will be able to play an active role in the field of e-sports.

References

1. Takahir Ishizu, Kenji Sakoma, Makoto Sakamoto, “Fundamental Study on Control of CG Characters by Electroencephalography (EEG) analysis”, MEMORIES OF FACULTY OF ENGINEERING UNIVERSITY OF MIYAZAKI, NO.48, 159-162, 2019.
2. “2017 Advanced Content Technologies, etc. Distribution Promotion Business Report, Part I, guidelines for using content production technology such as VR”, 2018 [Online].https://www.vipo.or.jp/u/I-1_SenshinContents_Guideline.pdf.
3. Tomoki Hidaka, Yuki Shiraishi, Yukiko Kawai, Jiro Okud, “Information control system by electroencephalogram using a portable terminal”, Information Processing Society of Japan, 2014.
4. Uchida Yu, Kikuro Fujimura, Yoshihiro Maegaki, “A Study on Numerical Processing of EEG Data for EEG Analysis”, Japan Society for Fuzzy Theory and Intelligent Informatics, 2014.
5. Takio Kurita, “Introduction to Support Vector Machine” [Online].https://home.hiroshima-u.ac.jp/tkurita/lecture/svm.pdf.
6. “Mind Wave Mobile headset by NeuroSky”, [Online]. https://www.neurosky.jp/mindwave-mobile2/
7. Jun Sakamoto, Yasuhiro Mori, Takayoshi Sekioka, “Probability Analysis Method By Discreate Fast Fourier Transform”, J.Struct, Constr, Eng.AIJ, No.472, 39−45, Jun, 1995.

Approach to tourism support by aerial video using CG

Kenji Sakoma, Noritake Seto, Kodai Miyamoto, Taketo Kamasaka, Makoto Sakamoto*, Amane Takei
Faculty of Engineering, University of Miyazaki,
Miyazaki-City, Miyazaki, Japan

Tsutomu Ito
National Institute of Technology, Ube College
Ube-City, Yamaguchi, Japan

Takao Ito
Graduate School of Engineering, Hiroshima University
Higashi-Hiroshima, Hiroshima, Japan

E-mail: hm16011@student.miyazaki-u.ac.jp, hm15037@student.miyazaki-u.ac.jp, hm16043@student.miyazaki-u.ac.jp, fruits2000jp@yahoo.co.jp
**Corresponding Author*

Abstract

Recently, as a first step in the tourism business, we will consider whether it is possible to increase the number of tourists by incorporating IT technologies such as CG in Introductory video. In this paper, we created a CG image of a hawk flying in the sky combined with a real Takachiho video taken with a drone. As a result, various problems have been identified, and I hope to improve them in future research to produce even better ones.

Keywords: EEG, BMI, FFT, SVM, VR

1. Introduction

Sightseeing is for people to refresh themselves and for their hobbies, so it heals the mind in a different environment than usual. However, from the perspective of the local people, it is a big commerce that can be expected to revitalize the area and increase the number of residents.

In recent years, rural areas have been declining due to urban concentration of work and declining birthrate and aging population. Meanwhile, the number of local governments that are trying to regain the vitality of the past by focusing on the tourism industry is increasing year by year. As a first step in the tourism industry, consider whether it is possible to increase the number of tourists by incorporating IT technologies such as CG into introductory videos such as PV and commercials.

This research is part of a joint development with Takachiho Town, and the beautiful nature of Takachiho is created by making the images taken by the drone look like the mountain hawk-eagle inhabiting Takachiho is swirling or plunging in the sky. We conducted basic research to create a promotional video to introduce people in the area.

2. Benefits of tourism[4)]

When many tourists come to visit due to the development of the tourism industry, economic activities in various fields such as accommodation, transportation, eating and drinking, and travel industry become active, and the economic ripple effect is high. For example, by utilizing the nature and historic sites that originally existed in the area, even a small scale can be established as an industry. Therefore, at least some funds can be developed. Also, if you can attract tourists from abroad, you can earn foreign currency. In particular, in Takachiho Town, the mythical nature is the centerpiece of tourism, so even a small amount of assets can demand high productivity.

Initiatives of Takachiho Town[5)].

- A tour of mythical nature and shrines (see Fig. 2).
- PR of traditional Takachiho Kagura based on myth (see Fig. 1).
- Sale of gourmet foods and souvenirs that make use of the rich land of Takachiho Town.

Fig. 1. Actual Kagura.

Fig. 2. Amanoyasukawara, Kunimigaoka, Takachiho Gorge.

3. CG

In a narrow sense, it refers to images and videos generated from scratch through drawing, editing, and processing on a computer, and in a broader sense, it is a composite of images created by a computer with the original photographs, drawings, and videos. Including those processed and processed in a manner significantly different from the original[6)]

In recent years, AR (Augmented Reality) and VR (Virtual Reality) have made remarkable progress. However, CG is playing an active role not only in entertainment, but also in invisible parts such as model production of architectural design and simulation in the event of a disaster.

4. CG Tool

In this research, Blender is used for CG creation, and Unity is used for CG operation control.

4.1. Blender[8)]

Objects are created from a single cube using the extrusion and split functions (see Fig. 3). In order to smoothly reproduce the flapping of the wings, which is the greatest feature of birds, a total of 12 bones, 6 on each side from the base to the tip of the wings, are added to play the role of joints. Bones are normal ones and B-bones that are easier to express curves.

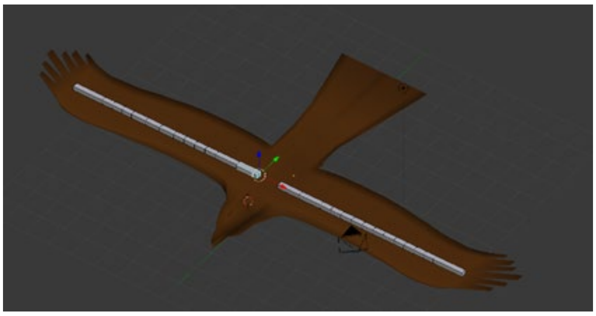


Fig. 3. The created hawk.

4.2. Unity[1-3)]

· animation

In order to express the flapping and dive of the wings, the animation function is used to assign numerical values to the main body of the hawk-eagle and the bones attached to the wings. The animation function applies keyframe animation, and by changing the interval of each numerical value, smooth movement is reproduced.

· Turning

The movement of the object turning is processed only by the script without using the animation function. In this research, since the object is moved according to the video actually shot, the rotation is only rotated around the z-axis.

The turning is limited to 30 degrees on each side, and it is set so that it does not turn even if the key input is continued beyond that. However, as an important process here, the turning angle is also different between the value on the console and the value to be acquired, so conditions are set so that the value is the same.

・ Image composition

An ICO ball is used to make the image change when the hawk turns.

This is achieved by installing a hawk object and a main camera inside the ICO sphere (see Fig. 4) so that the sphere rotates when the arrow keys are input.

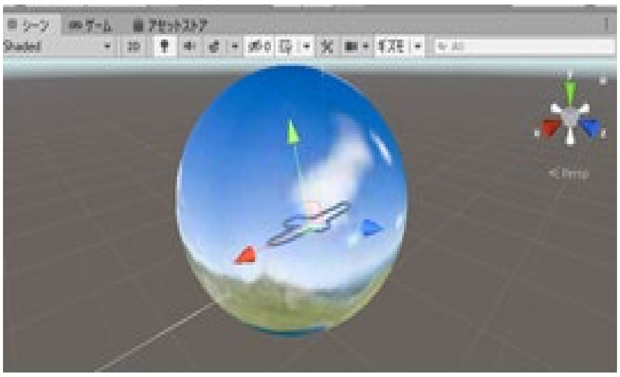


Fig. 4. Composition of 360-degree image on a sphere

・ ICO sphere

Fig. 5. 3D sphere model.

In this research, since it is necessary to paste the texture without distortion as much as possible, ICO spheres are used in the 3D sphere model (see Fig. 5), and ICO spheres are treated as distributed[9].

・ glTF

Blender can export in glTF format, which allows it to be imported into Unity with almost the same shading as Blender (see Fig. 6).

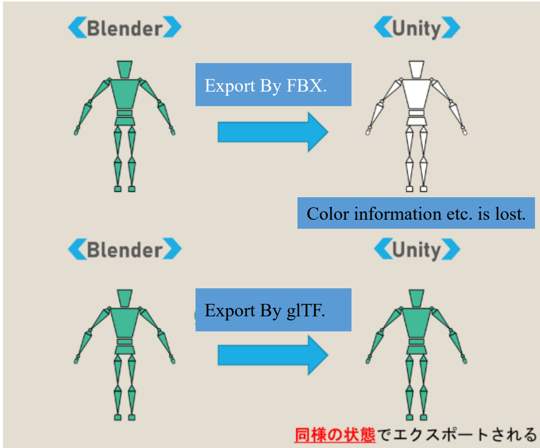


Fig. 6. glTF

Since it is necessary to paste the image on the inner surface of the sphere, download and use the UniGLTF plug-in[10].

・ Screen transition due to a sudden descent of a hawk

The screen transition is processed by the script, and the screen transition is performed when the space key is input.

In this state, the screen transition will be performed before the animation of the sudden descent of the mountain hawk eagle is performed, so when the space key is input, the elapsed time will be observed and the video of the tourist introduction will be played when the elapsed time reaches the end of the animation. Transition to the scene.

・ Sound / Video

Free materials are used for sounds, hawk-eagle calls, aerial images, and sightseeing introduction videos [7].

5. Consideration / Future issues

In this research, we used Blender and Unity to create a CG animation that promotes Takachiho by synthesizing the CG of a mountain hawk eagle flying in the sky with the actual image of Takachiho taken aerial with a drone. This time, we succeeded in pasting the image on the sphere, but there is a problem that the resolution of the free and usable image is too low and it becomes too rough at the time of compositing.

In addition, although the specification was originally to combine images, the moving image file must be in the Equirectangular format in order to combine images from a 360-degree viewpoint, so in this research, images are combined.

Since it is possible to transition to only one scene in the screen transition, I think that it will be possible to transition to multiple scenes by acquiring the rotation angle of the sphere and assigning it to the scene corresponding to a specific angle.

References

1. Mikito Yoshito, “Unity5 3D/2D Introduction to game development practice”, Soshimu, 2015.
2. “Ministry of Internal Affairs and Communications Information and Communications White Paper Internet penetration (2015 version)”, [Online]. http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h27/html/nc372110.html.
3. “Unity DOCMENT/Unity Manual”, [Online]. https://docs.unity3d.com/ja/current/Manual/index.html.
4. “Tourism Wikipedia”, [Online]. https://ja.wikipedia.org/wiki/%E8%A6%B3%E5%85%89%E6%A5%AD.
5. “Takachiho Kankokyokai Official site”, [Online]. http://takachiho-kanko.info/.
6. “IT Yogo jiten e-Words CG”, [Online]. http://e-words.jp/w/CG.html.
7. “On-Jin ~Onjin~”, [Online].http://on-jin.com/sound/kan.php?kate=%E9%A2%A8.
8. “Blender.jp”, [Online]. https://blender.jp/.
9. “360-dgree eizo no gosei”, [Online]. https://shinrinmusic.com/unity-360-image/.
10. “gITF”, [Online]. https://xr-hub.com/archives/20285.

Method for detecting eye misalignment based on movement near the center of the pupil

Noriyuki Uchida[1,2]
[1] Kyushu University of Health and Welfare
1714-1 Yoshino-machi, Nobeoka-City, Miyazaki, Japan

Kayoko Takatuka[2], **Hisaaki Yamaba**[2], **Masayuki Mukunoki**[2], **Naonobu Okazaki**[2]
[2] University of Miyazaki
1-1 Gakuen Kibanadai-nishi, Miyazaki-City, Miyazaki, japan
E-mail: n-uchida@phoenix.ac.jp

Abstract

Only ophthalmologists and orthoptists are eligible to diagnose misalignment of the eyes amid persistent understaffing. We automated the traditional eye misalignment evaluation (Cover-test) in our earlier studies and developed a technique to distinguish the misaligned eyes from its behavior. This method, however, had a problem with the vertical eye movements being affected by the eyelids and eyelashes, which prevented thorough detection. Therefore, by observing only the movement near the center of the pupil, where it is less likely to be disturbed by the eyelids and eyelashes, we arranged an alternative method to recognize anomalies. Based on this process, we achieved an inspection system and carried out a performance experiment. We confirmed that we could observe not only horizontal motion, but also that of the vertical eye.

Keywords: eye position examination, cover test, strabismus/ heterophoria, compact camera

1. Introduction

As misalignment of the eyes, such as strabismus and heterophoria, causes defective binocular vision and eyestrain, early detection and training are required. Strabismus in children, in particular, is a risk factor that causes amblyopia. One of the research subjects of pediatric ophthalmology is the diagnosis of strabismus at an early age and the start of treatment.[1] Only qualified personnel such as ophthalmologists and orthoptists can diagnose eye misalignment, despite the ongoing shortage of skilled personnel. Therefore, it is significant to develop a method to verify the alignment of the eye without trained staff.[1,2]

There are two types of eye position examination: qualitative tests and quantitative tests. Specifically, the qualitative test investigates abnormalities, while the quantitative test measures the magnitude of the ocular deviation. While there are research reports on methods for measuring the amount of ocular deviation,[2] this approach only works for exotropia, as it is easy to assess the eye location during the measurement. We also explored the idea of using the current eye-tracking system as an alternative tool. With this method, we considered it necessary to plan various parameters for the application site and light to identify the slightest eye movements that suggest strabismus and heterophoria. After the recent development of medical image diagnostic systems, we explored the possibility of creating a machine learning-based detection system based on continuous pictures of eye movements during the examination. We found it arduous to collect study data because the ocular inspection has never been a procedure for image diagnosis before.

As a result, our research developed a basic inspection system capable screening strabismus and heterophoria

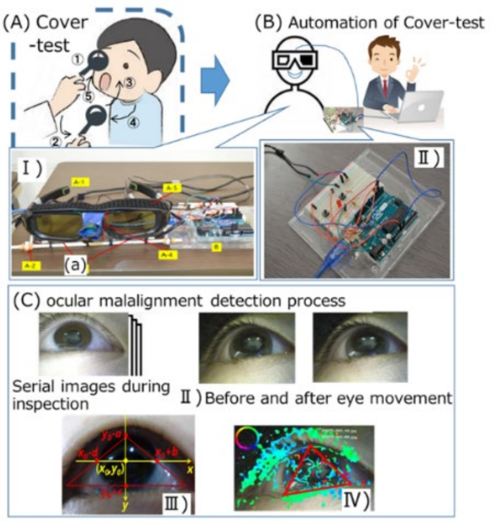


Fig. 1 based on the eye movement

screening without restrictions. We automated the inspection process of the traditional qualitative inspection form, Cover-Test (Fig.1-(A)) (Fig.1-(B)-I, II),[1] and established and systematized a procedure for the detection of eye movement abnormalities. (Fig.1−(C)−I to IV). Cover-test is an inspection method to decide tropia or heterophoria by the subject focusing on an object. We determine it tropia if the exposed eye moves when the other eye is covered. If the covered eye moves as it is exposed, we judge it heterophoria.

The conventional approach is to cover-uncover the eyes with an occluder and evaluate abnormality by visual inspection (Fig.1-(A)). In the previous analysis, we automated the cover test as follows. With the method we originally developed, we extracted the area for anomaly detection from the moving image around the eye (Fig.1-(C)-III). Then, using optical flow technology,[3] we developed a series of processes that detect eye movement (Fig. 1-C-IV) and assess abnormality.[4,5] During the test, we took the moving picture around the eye with an ultra-compact camera (Fig.1-B-I-(a)) mounted within the left and right lenses. Eye movements displaying signs of eye misalignment occur in approximately 7 to 8 seconds immediately after being covered or exposed. Therefore, we extracted a series of photos taken at a time interval of about 0.2 seconds during that span. However, due to the influence of eyelids and eyelashes, we were unable to observe the vertical movements of the eye.

Vertical strabismus/heterophoria does not have as many cases compared with horizontal strabismus. Early-stage diagnosis is crucial, however, as it can be a sign of severe disorders such as cerebral palsy and heart failure. Also, the number of cases has increased over the past few years due to an aging population.[6] Therefore, there is considerable significance in developing a system for detecting strabismus and heterophoria in the vertical direction

2. Research purpose

The goal of this study was to develop an alternative method for detecting eye misalignment to distinguish not only horizontal but also vertical eye movements. Based on the issues of the previous study, we devised a technique to diagnose disorders by monitoring only the actions near the center of the pupil, where eyelids and eyelashes are less likely to affect the pupil. Section 3 explains the proposed method, Section 4 introduces the performance evaluation experiments of the proposed method, and Section 5 summarizes them.

3. Proposal method

We use an ultra-compact infrared camera (referred to as an "infrared camera") to take pictures of the center of the pupil and extract its contour. At the beginning of the test, we refer to the location of the pupil center as the "reference point"(described later in 3.3). Focusing on the dispersion of the distance from the reference point, we find that we can detect strabismus and heterophoria in both directions, both horizontally and vertically, and abnormality.

The following explains the three processes in order.

(1) the extraction of the pupil area (Fig.2-I, II), (2) the measurement of the pupil center coordinates (Fig.2-III), and (3) the determination of the abnormality.

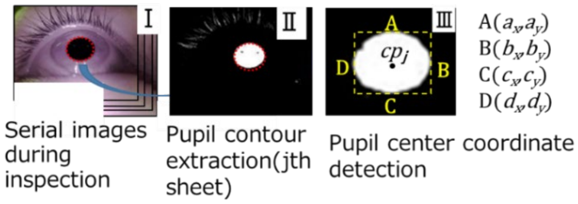


Fig. 2. process of pupil center detection

3.1. *Extraction of the pupil area*

First, we tried to remove the pupil contour with the existing ellipse fitting method.[7] But we could not extract stable enough to detect the slightest twitch of the eye showing signs of strabismus and oblique location. We, therefore, use the color identification function based on the RGB color model given in OpenCV as another tool. We found that we can extract the pupil contour by setting the RGB values from (0,0,0) to (15,15,15).

3.2. *Calculation of pupil center coordinates*

Fig.2-Ⅲ shows the coordinates of the extracted pupil area. The y coordinate, where the position coordinate is maximum, is A (a_x, a_y), and the minimum coordinate is C(c_x, c_y). The position coordinate where the x coordinate is maximum is B (b_x, b_y), and the minimum is D (d_x, d_y). We calculate the pupil central coordinate cp as follows.

$$cp_j = \left(d_x + \frac{b_x - d_x}{2},\quad a_y + \frac{c_y + a_y}{2}\right) \quad (1)$$

(j: Image number)

3.3. *Anomaly Evaluation*

In the previous analysis, we used the real automatic Cover-Test apparatus (Fig.1-(B)) to assess the abnormality of constant eye movements. We found it possible to evaluate anomaly by checking the eye movements seen continuously in the same direction, in the photographs taken at 0.2-second intervals for around 7 to 8 seconds immediately after the cover/uncover: specifically, the 4th to 10th pictures. Using the pupil center location at the start of the test as a reference point, we found that the distance difference between this reference point and the pupil center obtained at each time point during the test estimate the degree of abnormality.
Our initial assumption was that the degree of the abnormality would correspond with the absolute distance of eye movement.
However, the findings of the study did not show any correlation. The gap between the eyes and the location of the infrared camera differs between subjects. From the above, we set the measurement parameter based on the variance of the distance from the reference point of the pupil center and carried out an abnormality judgment based on these evaluation criteria.

Setting evaluation criteria

We achieve a series of transitions of the pupil center coordinates $cp_j = (x(cp_j),\ y(cp_j))$ from a series of image data (referred to as a test image) taken from around the eye. ($1 \leqq j \leqq n$ and j is the serial number sequence assigned to the test image.) Next, we set the central coordinate cp_1 taken as the reference point from the first sheet of a series of test images. The distance from the reference point of the center coordinates cp_j $(2 \leq j \leq n)$ on the second and subsequent sheets is measured in horizontal and vertical directions.We obtain the average horizontal direction a_x and the average vertical direction a_y, and the variances: $s^2{}_x$, $s^2{}_y$.

$$s^2{}_x = \frac{1}{n}\sum_{n=1}^{n}\left(\left(|x(cp_j) - x(cp_1)|\right) - a_x\right)^2 \quad (2)$$

$$a_x = \frac{1}{n}\sum_{j=1}^{n}\left(|x(cp_j) - x(cp_1)|\right)$$

$$s^2{}_y = \frac{1}{n}\sum_{n=1}^{n}\left(\left(|y(cp_j) - y(cp_1)|\right) - a_y\right)^2 \quad (3)$$

$$a_y = \frac{1}{n}\sum_{j=1}^{n}\left(|y(cp_j) - y(cp_1)|\right)$$

Through the above method, we have achieved the variances of all the collaborators to produce the study data and the correct judgment results of each individual with the Maddox rod: the precision inspection device. We set the criterion by comparing each variance with the Maddox rod test results. Table 1 indicates the evaluation criteria. Forty-four collaborators contributed to the development of this criterion; the anomalies were 13 in the right position and 36 in the presumed horizontal strabismus or oblique position. The result was: 15 individuals had an angle of deviation in strabismus of 1Δ (prism), 15 had 2Δ, three had 3Δ, and three 4Δ. To generate pseudo-perspective data on minimal cases of study data for the development of a criterion for vertical strabismus, we followed the specified procedure under the guidance of an expert with the aid of 8 of the 13 collaborators listed above.
Apart from the main contents of the proposed detection method that we mentioned, it was necessary to differentiate between the blinking movement and to exclude the corresponding system from the assessment objective to automate the proposed judgment process completely. And specify the frame when the shutter opens and closes. We also automated these processes and modified the automation of the entire detection process.

Table.1. evaluation criteria of abnormal

vertical direction		horizontal direction	
deviation angle	mean of variance	deviation angle	mean of variance
<1Δ	1.43	<1Δ	2.10
1 Δ	1.90	1 Δ	2.56
2 Δ	4.02	2 Δ	5.63
3 Δ	9.66	3 Δ	11.60
4 Δ	16.73	4 Δ	21.10

4. Verification experiment

4.1. *Experiment 1*

First, we performed an anomaly judgment using this method for all 20 subjects, compared the judgment results with those of the Maddox rod test, and assessed the system performance by the coincidence rate. We evaluated the outcomes in the following three stages.

Deviation angles of strabismus larger than 1Δ and smaller than 3Δ were abnormal, those larger than 3Δ were highly abnormal, while those less than 1Δ were normal.
Table 2-(1) displays the results of the analysis. The notation form is that V and H provide the direction of eye movement, and the numerical value after V and H represents the angle of strabismus. For example, V0 of V0H6 indicates normality in the vertical direction, and H6 shows that there is a horizontal strabismus/heterophoria of about 6Δ. H0-1 means that the horizontal anomalies are less than 1Δ. Table 2-(1) shows a case, subject B, in which the resulting judgment differs from the correct result of the Maddox rod test resulting in a concordance rate of 95 percent (19/20).

4.2. *Experiment 2*

As mentioned at the beginning, this study aimed to develop a method to detect misalignment in horizontal and vertical eye movements. However, of the 20 subjects, only one case was abnormal in the vertical direction, which did not verify whether we achieved the purpose of this study. Therefore, using pseudo strabismus data, we performed further experimental verification.
With the cooperation with 15 {A, C, D, E, E, F, G, H, I, K, L, N, O, S, T} of the 20 subjects whose system judgment result in Experiment 1 was Normal, abnormality degree smaller than 1Δ, we used 20 data (A' to T') randomly selected from the pseudo-strabismus data - 8 x 15 data: pseudo-strabismus data for 15 subjects, four types (1Δ, 2Δ, 3Δ, 4Δ) each in the vertical and horizontal directions - as test data. We performed an abnormality judgment and compared it with the True Value expected when creating pseudo data.
Table 2-(2) shows the judgment results. Four {D', J', M', T'} out of 20 cases differed from the expected correct judgment by the system, and the concordance figure was 75 percent (15/20).

4.3. *Discussion*

Experiment 1
As shown in Table 2, the test results of subject B from our system were highly abnormal in both the vertical and horizontal directions. The accurate findings of the Maddox rod test were consistent with the results of the horizontal direction, which were abnormally high, while the results of the vertical direction revealed a significant difference. What we found from the subject B test image was that the overlap between the lower eyelid and the pupil was substantial. We did not accurately identify the area near the center of the pupil, which could have led to a wrong decision. We will deal with this problem by taking photos of the eyes from the front. As a consequence, the orientation and angle of the camera is the concern.

Experiment 2
As shown in Table 2, the difference between the outcomes of all 4 cases {D', J', M', T'} that did not fit the right device judgment was around 1Δ, which was subtle. We have grouped the causes into three patterns. As in Case B in Experiment 1, we will deal with one out of the four cases by changing the angle of the shooting of the camera.Two cases were irrelevant because they could not construct pseudo strabismus results. The cause of the remaining was a distortion of the pupil contour with the use of hard contact lenses, which we considered irrelevant to this system. Based on the above, we assessed the actual failure to be one out of four.

Table 2. Results

(1) Experiment 1			(2) Experiment 2		
subject	True Value (Maddox)	System results	subject	True Value (Expected value)	System results
A	normal	normal	A′	V4H0	V H.Ab.
B	V0H6	VH H.Ab.	B′	V2H0	V Ab.
C	normal	normal	C′	V0H2	H Ab.
D	normal	normal	D′	V0H2	VH Ab.
E	normal	normal	E′	V3H0	V H.Ab.
F	normal	normal	F′	V3H0	V H.Ab.
G	V0H0-1	normal	G′	V2H0	V Ab.
H	normal	normal	H′	V0H1	H Ab.
I	normal	normal	I′	V1H0	V Ab.
J	V0H0-1	normal	J′	V0H2	VH Ab.
K	normal	normal	K′	V4H0	V H.Ab.
L	normal	normal	L′	V0H1	H Ab.
M	VH0-1	normal	M′	V1H0	normal
N	normal	normal	N′	V0H2	H Ab.
O	normal	normal	O′	V3H0	V H.Ab.
P	V0H5	H H.Ab.	P′	V1H0	V Ab.
Q	normal	normal	Q′	V0H1	H Ab.
R	V3H12	VH H.Ab.	R′	V1H0	V Ab.
S	normal	normal	S′	V2H0	V Ab.
T	normal	normal	T′	V2H0	VH Ab.

V: Vertical , H: Horizontal
Ab.: Abnormal , H.Ab.: Highly Abnormal

5. Summary

To automate the strabismus/heterophoria screening, we have developed an alternative method that can detect not only horizontal but also vertical eye movements.
It is less susceptible to eyelids and eyelashes that make it crucial to recognize the vertical flow of the eye. Specifically, it only senses the motion of the center of the eye where eyelids and eyelashes cannot intervene. With this system, we measured the distance variance from the reference point of each pupil center position and then used it as an assessment criterion to devise a method for detecting anomalies. We have put in place an inspection framework based on the proposed approach and carried out a performance assessment test. First, we set 44 standard inspection data, four pseudo-data forms × 13=52, and 96 research data, and evaluation criteria.
We then carried out an anomaly judgment on 40 test data, which is 20 inspection data and 20 pseudo data, and compared the results with the right ones to test the efficiency of the proposed system. As a consequence of the experiment, the success rate of judgment on the data is 85 percent (34/40). However, we found four among the six failures irrelevant to this method because three were false pseudo data, and one was for the use of hard contact lenses. Therefore, the number of actual judgment failures of this system was two out of 40 cases, and the success rate was 95 percent (38/40). From the above, we defined the problems with the pupil center detection error for the infrared camera angle of the photograph and explained the future theme. [Acknowledgment] This work was supported by JSPS KAKENHI Grant Number JP19H00505

6. References

1. S. Hasebe, Objective Tests of Strabismic Deviation : Basics and Perspective, *Journal of the eye* 18(9), pp.1105-1110, 2001
2. Y. Takahashi, Y. Shinomiya, T. Noguchi, T. Niida, Study of the occlusion time required for stable fusion-free position by using an eye-tracking system, *Folia Japonica de Ophthalmologica Clinica* 9(3), pp.234-237, 2016
3. Y. Yamauchi,T. Yamashita, H. Fujiyoshi, Human Detection Based on Statistical Learning from Image, *IEICE TRANSACTIONS* Vol.J96-D, No.9, pp.2017-2040,2013
4. N. Uchida, K. Takatuka, K. Hinokuma, K. Hirata, H. Yamaba, N. Okazaki , Automated cover-uncover test system using active LCD shutter glasses, *23rd International Symposium on Artificial Life and Robotics (AROB)*, 2018.
5. N. Uchida, K. Takatuka, K. Hinokuma, K. Hirata, H. Yamaba, N. Okazaki, HMD-based Cover Test System for the Diagnosis of Ocular Malalignment, *Artificial Life and Robotics (AROB)*, Vol.24-3, pp332-337, 2019.
6. A. Oohira, Excyclotorsion and Eye Dominancy in Age-related Small-angle Vertical Strabismus, *Journal of Japanese Ophthalmological Society* 123(2) pp.161-166, 2019
7. K. Yokota, K. Murata, Y. Sugaya, K. Kanatani, Accuracy Comparison of Ellipse Fitting: From Least Squares to Hyper-Renormalization, *IPSJ SIG Technical Report* 111(379) pp.123-130, 2012

Basic Research on Video Production for Educational Support by Virtual Technology

Kodai Miyamoto, Taketo kamasaka, Kenji Sakoma, Makoto Sakamoto, Amane Takei
Faculty of Engineering, University of Miyazaki,
*Miyazaki-City, Miyazaki, Japan**

Tsutomu Ito
National Institute of Technology, Ube College,
Ube-City, Yamaguchi, Japan
Takao Ito
Guraduate School of Engineering, Hiroshima University,
Higashi-Hiroshima, Hiroshima, Japan

E-mail: hm16043@student-u.ac.jp, hm16011@student-u.ac.jp, hm15037@student-u.ac.jp,
fruits2000@yahoo.co.jp
**Corresponding Author*

Abstract

As a result of conducting a questionnaire about science classes to high school students in 2016, the percentage of high school students who answered "I like science" and "Science is important" is lower than other subjects. However, more than 80% of elementary and junior high school students said they like experiments and observations. In addition, the 2019 smartphone penetration rate survey found that it is popular among about 90% of students. In addition, VR technology has recently made remarkable progress. From the above, I researched the idea that creating a simulation application using VR technology using smartphones would change the way high school students think about science classes. In this paper, we have developed a simulation application for science experiments. Subjects were asked to experience the newly created app and complete a questionnaire. As a result, the average score is 4 out of 5 and it is not bad. But at the same time, a problem was found. The problem was that this app was a simulation app, so there wasn't much user operability, so I wanted a little more operability. I want to make apps in other fields while improving the problem.

Keywords: Education, science, chemistry, physics, experiment / observation, virtual reality, simulation app

1. Introduction

In the materials related to science surveyed by the Science Working Group in 2016, it was found that the percentage of high school students who answered "I like studying science" and "It is important to study science" is lower than other subjects. However, the percentage of elementary and junior high school students who answered "I like experiments and observations" exceeds 80%.

A survey of smartphone penetration in 2019 found that about 90% of students have it. Most schools have PC classrooms. In addition, each classroom has a personal computer for teachers. Under these circumstances, it is

thought that the educational effect can be further improved in various classes.
Based on the above, the purpose of this research is to contribute to education by creating a science simulation application using VR technology.

Fig. 1. Example of an English conversation learning app that supports VR

2. Physics experiment app

In this study, we tried to create physics simulations of "projectile motion (projectile motion)" and "falling body motion". Since these experiments require large experimental tools and it may be difficult to obtain accurate numerical values, we thought that they would be suitable as experiments to be performed in a virtual environment.

2.1. *Development environment*

This survey was conducted in the environment shown in Table 1.

operating system	Windows10
Programming language	C#
software	Unity 2019.2.15f1

Table. 1. Development environment

2.2. *Implemented function*

In this study, we implemented the following two physics experiment simulations.
1. Projectile motion
2. Falling exercise

2.2.1. Projectile motion

Projectile motion moves from the initial position.
After deciding the angle and speed and pressing Start Button, the object will be fired and the distance will be displayed. If it is difficult to check the current situation, you can zoom in and out on the ball with the zoom below.

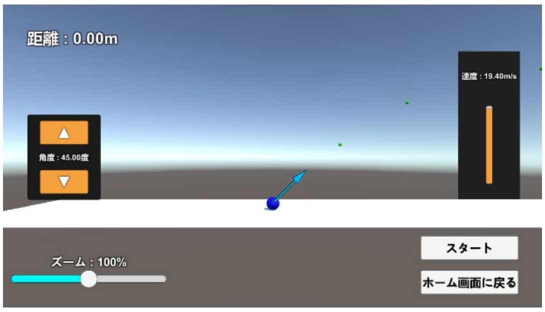


Fig. 2. Created projectile motion app

2.2.2 Falling exercise

I made a slope that controls the falling speed by rolling the ball while changing the angle of the slope. Determine the angle of tilt and press Start to start spinning the sphere. The tilt angle, ball speed, and ball position are displayed.

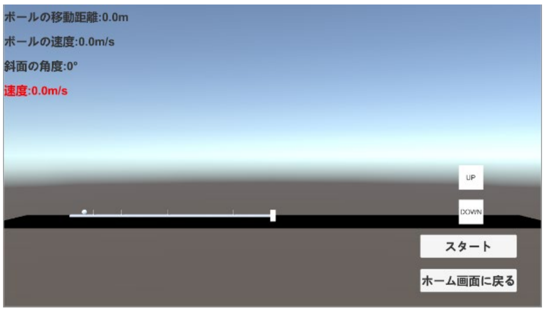


Fig. 3. Created Falling exercise app

3. Chemistry experiment app

In this study, we tried to make prototypes of chemical experiment simulations of "flame reaction" and "silver mirror reaction". Since these experiments use many solutions and experimental tools, it may be difficult to prepare and clean up, and since the experiments take time, we thought that they were suitable for solving the shortage of class time.

3.1. *Development environment*

This survey was conducted in the environment shown in Table 2.

operating system	Windows10
Programming language	C#
software	Unity 2019.2.15f1
	Blender 2.81

Table. 2. Development environment

3.2. *Implemented function*

In this study, we implemented the following two simulations of chemical experiments.
1. Flame reaction
2. Silver mirror reaction

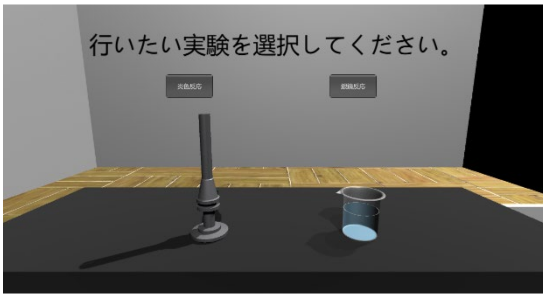


Fig. 4. Created scientific experiment simulation app

4. Evaluation experiment

An evaluation experiment was conducted to verify whether the developed science experiment simulation app was useful. We conducted a questionnaire to 10 people, including students in the laboratory, and evaluated the usefulness and usability of the system. The contents of the questionnaire are the following four points.

Evaluation item 1: Was it easy to operate?
Evaluation item 2: Was the result of the experiment easy to understand?
Evaluation item 3: Was it easy to imagine a physical phenomenon?
Evaluation item 4: Opinions and points to be improved (free description)

Evaluation items 1 to 3 will be evaluated on a scale of 1 to 5 points, and evaluation item 4 will be freely described.

4.1. Experimental result

The figure below shows the average of the evaluation results for evaluation items 1 to 3.
The average score for evaluation item 1 and evaluation item 2 was 4.4 points.
The average score for evaluation item 3 is 4.8 points, which is higher than that of evaluation item 1 and evaluation item 2.

For evaluation item 4
- good point

"Physics experiment"
- ✓ Easy to understand the trajectory of the sphere

"Chemical experiment"
- ✓ Easy to understand the flow of the experiment

There were opinions such as.

- Improvements

"Physics experiment"
- ✓ Display of the distance of the object that was skipped one before
- ✓ Change the trajectory of the ball from a point to a line

"Chemical experiment"
- ✓ Addition of tutorial
- ✓ Change the touched part to make it easier to understand

There were opinions such as.

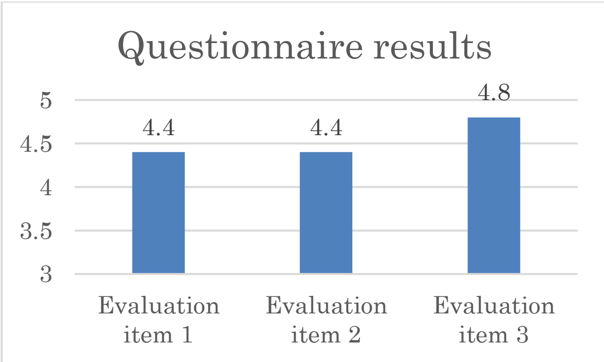


Fig. 5. Questionnaire average score

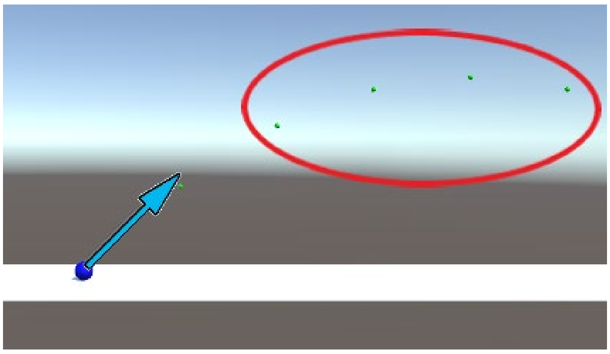


Fig. 6. Examples of questionnaire improvement points

5. Consideration

Regarding evaluation item 4, there were many opinions about the expansion of functions. Extension to VR was also mentioned as an extension of the function. With regard to expansion to VR, effects such as immersive feeling and easy image of physical phenomena can be expected. However, since VR-specific equipment is required, I think we must devise ways to make it easy for anyone to use.

6. Future tasks

If we can create a simulation app that is similar to the teaching materials that students usually handle, we can support education not only in science but also in various subjects. Also, before creating an app, you need to think about what kind of app can support student education.

This time, it could not be expanded to VR. If it is possible to expand to VR, it will be possible to realize a more immersive feeling by actually expressing the sound and science room during the experiment, so this is a future issue.

7. Summary

As mentioned at the beginning, VR technology has grown remarkably in recent years. In this research, we tried to create an application that can perform science experiments using virtual reality, but as mentioned in future tasks, special equipment is required to reproduce VR. It has become a trial of only CG images. However, from the results of the evaluation experiment, it was found that a useful application was created. Ultimately, I think it is to create a system that will help support education.

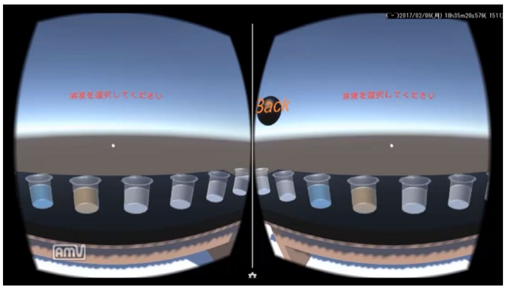


Fig. 7. Extension to VR

References

1. Curriculum Subcommittee Science Working Group "Materials on Science" , February 5, 2016.
2. National Academic Achievement / Learning Situation Survey , 2015.
3. "[Latest version] The smartphone penetration rate in 2019 is released to the public by age group, region, and age group! It's a smartphone-only era! I predicted how marketing should change in the future." Marketing Research Camp.
4. Aimi Kitamura "Unity 5 Textbook 2D & 3D Smartphone Game Introductory Course", SB Creative Co., Ltd., 2016.
5. "AEON provides VR English conversation app ..." Hospitality "simulated experience improves ability".

There is a movement towards the development of hula costume CAD

Taketo Kamasaka, Kenji Sakoma, Kodai Miyamoto, Makoto Sakamoto*, Amane Takei
Faculty of Engineering, University of Miyazaki,
Miyazaki-City, Miyazaki, Japan

Tsutomu Ito
National Institute of Technology, Ube College
Ube-City, Yamaguchi, Japan

Takao Ito
Graduate School of Engineering, Hiroshima University
Higashi-Hiroshima, Hiroshima, Japan

E-mail: hm16011@student.miyazaki-u.ac.jp, hm15037@student.miyazaki-u.ac.jp, hm16043@student.miyazaki-u.ac.jp,
fruits2000jp@yahoo.co.jp
**Corresponding Author*

Abstract

In recent years, three-dimensional computer graphics (3DCG) technology has been applied in various fields. One of the problems in this technology is the problem of contact between clothes and other objects (such as the body). In this paper, we verified each simple animation using the three methods of calculation. As a result of the verification, the FB Euler method creates a simple but stable animation Was completed. In this paper, we were able to create simple animations.

1. Introduction

In recent years, 3D computer graphics (3DCG) technology has been applied in various fields such as AR / VR technology, movies, games, and virtual fitting of clothes[1].
In this research, we will use a physical model to simulate cloth, which is applied to 3DCG animation.
In this research, we focus on past research and work on a simulation of a simple cloth (for example, a skirt) that considers contact with other objects (the body of a hula dancer)[2,3,4].

2. Simulation method

The Modeling methods for expressing cloth can be roughly divided into two types: geometric models and physical models.
In order to perform dynamic simulation, we will use a physical model that considers the change process this time.

2.1. *Mass spring model*

In the mass spring model, in order to express the anisotropy of the cloth, the shape of the cloth is a spring that connects the mass points arranged in a square grid and the mass points adjacent to each other with the direction along the warp and weft threads of the cloth as the axis. (See Fig. 1.).

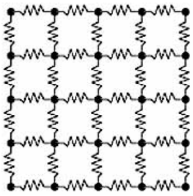


Fig. 1. Mass spring model

2.2. *Mechanical properties of cloth*

Of the six mechanical properties of cloth listed by KES, four are considered when modeling the shape of cloth: tensile properties, shear properties, bending properties, and weight properties (See Fig. 2.)[5].

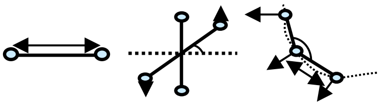


Fig. 2. Mechanical propeties of cloth

2.3. *Power on mass points*

The resultant force $\boldsymbol{F_i(t)}$ of the force applied to the particle i at time t is as follows from the external force such as wind force $\boldsymbol{f_{wind}}$ and gravity $\boldsymbol{f_{gravity}}$ and the force generated from the internal structure of the cloth such as the damping force $\boldsymbol{f_{damper}}$ against the vibration of the spring and the force $\boldsymbol{f_{stretch}}$ by the spring. Can be expressed in

$$\boldsymbol{F_i}(t) = \boldsymbol{f_{gravity}} + \boldsymbol{f_{dampar}} + \boldsymbol{f_{stretch}} + \boldsymbol{f_{wind}}$$
$$= m_i\boldsymbol{g} + \boldsymbol{r_i}(t) + \boldsymbol{f_i}(t) + \alpha\cos(\beta)\sin(\gamma) + \alpha\sin(\beta)\sin(\gamma)\,. \quad (1)$$

2.4. *Equation of motion*

Based on the equation of motion, the acceleration $\boldsymbol{a_i}(t)$ is expressed by the following equation.

$$\boldsymbol{a_i}(t) = \boldsymbol{F}_i(t)/m_i\,. \quad (2)$$

2.5. *Numerical calculation method*

In this study, Euler method, FB Euler method, and Runge-Kutta method were used from the explicit formulas[6].

2.5.1. *Euler method*

Assuming that the mass point moves linearly with constant acceleration between time t and time $t + \Delta t$ the velocity $\boldsymbol{v_i}$ and the coodinates $\boldsymbol{x}_i$ can be calculated by the following equations using the force acting on the particles calculated in Section 2.2.

$$\boldsymbol{v}_i(t+\Delta t) = \boldsymbol{v}_i(t) + \boldsymbol{a}_i(t)\Delta t$$
$$\boldsymbol{x}_i(t+\Delta t) = \boldsymbol{x}_i(t) + \boldsymbol{a}_i(t)\Delta t. \quad (3)$$
$$(\Delta t\text{: time step})$$

2.5.2. *FB Euler method*

It is a numerical calculation method that uses both the Euler method and the backward Euler method at the same time. Velocity applies the Euler method for current acceleration, and position applies the backward Euler method for future speed.

$$\boldsymbol{v}_i(t+\Delta t) = \boldsymbol{v}_i(t) + \boldsymbol{a}_i(t)\Delta t$$
$$\boldsymbol{x}_i(t+\Delta t) = \boldsymbol{x}_i(t) + \boldsymbol{a}_i(t)\Delta t. \quad (4)$$
$$(\Delta t\text{: time step})$$

2.5.3. *Runge-Kutta method*

At first glance, the amount of calculation seems to be four times, but since the calculation accuracy is high, the time step width Δt can be increased, and more stable simulation is possible.

$$k\boldsymbol{v_1} = \boldsymbol{a_1}\,\Delta t, k\boldsymbol{x_1} = \boldsymbol{v}(t)\Delta t$$
$$\boldsymbol{a_2} = \frac{F\left(\boldsymbol{x}(t) + {k\boldsymbol{x_1}}/{2}\right)}{m} - c\boldsymbol{v}(t) + \boldsymbol{g}$$
$$k\boldsymbol{v_2} = \boldsymbol{a_2}\,\Delta t, k\boldsymbol{x_2} = \left(\boldsymbol{v}(t) + {k\boldsymbol{v_1}}/{2}\right)\Delta t$$
$$\boldsymbol{a_3} = \frac{F\left(\boldsymbol{x}(t) + {k\boldsymbol{x_2}}/{2}\right)}{m} - c\boldsymbol{v}(t) + \boldsymbol{g}$$
$$k\boldsymbol{v_3} = \boldsymbol{a_3}\,\Delta t, k\boldsymbol{x_3} = \left(\boldsymbol{v}(t) + {k\boldsymbol{v_2}}/{2}\right)\Delta t$$
$$\boldsymbol{a_4} = \frac{F\left(\boldsymbol{x}(t) + {k\boldsymbol{x_3}}/{2}\right)}{m} - c\boldsymbol{v}(t) + \boldsymbol{g}$$
$$k\boldsymbol{v_4} = \boldsymbol{a_4}\,\Delta t, k\boldsymbol{x_4} = (\boldsymbol{v}(t) + k\boldsymbol{v_3})\Delta t$$
$$\boldsymbol{v}(t+\Delta t) = \boldsymbol{v}(t) + \left(\frac{1}{6}k\boldsymbol{v_1} + \frac{1}{3}k\boldsymbol{v_2} + \frac{1}{3}k\boldsymbol{v_3} + \frac{1}{6}k\boldsymbol{v_4}\right)$$
$$\boldsymbol{x}(t+\Delta t) = \boldsymbol{x}(t) + \left(\frac{1}{6}k\boldsymbol{x_1} + \frac{1}{3}k\boldsymbol{x_2} + \frac{1}{3}k\boldsymbol{x_3} + \frac{1}{6}k\boldsymbol{x_4}\right). \quad (5)$$
$$(\Delta t\text{: time step})$$

The motion of mass points in the virtual world can be simulated by these equations.

2.6. *Collision with other objects*

In this study, we consider the collision between the cloth and the body. Since a large number of mass points and the body must be judged, a penalty method that takes a simple calculation is used. This time, the mass point is regarded as a point and the bone is regarded as a capsule, and collision judgment is performed.

2.6.1. *Expression of the body*

A capsule is a cylinder with hemispheres attached to both ends, defined by a line segment and radius r, and is the part within the distance and radius r from this line segment.
Calculating the position of the capsule with the back, hips, thighs, and shin bones as the axis of the cylinder did not perform collision detection correctly(see Fig. 3.). Therefore, this time, in order to increase the number of capsules and improve the accuracy of collision detection, the back and shin bones are divided into two, and the thigh bones are divided into three.

2.6.2. *Penalty method*

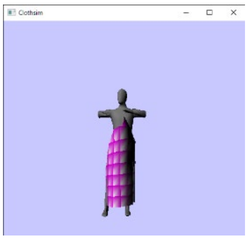


Fig. 3. Execution result by FB Euler method

The penalty method is one of the rigid body simulation methods. When objects collide with each other and a slight sunk occurs, a force (penalty force) proportional to the sunk value can be applied to obtain the behavior after the collision (See Fig.3.).
From this, it is possible to push back the lattice points colliding with the cloth to the outside of the body by the force of the spring so as not to be sunk into the body.

2.6.3. *Collision detection with capsule point*

The reason why capsules are used for collision detection is that the judgment only needs to be replaced. The collision between the capsule and the point is divided into a columnar part and a hemispherical part at both ends. Let the point x for collision determination be the intersection of the perpendicular and the line segment drawn from the point x to the line segment: the point y is obtained. At this time, the coordinates a and b at both ends of the line segment are set, and since the point y is on the line passing through the coordinates a and b, it is expressed by the following equation.

$$\boldsymbol{y} = \boldsymbol{a} + s(\boldsymbol{b} - \boldsymbol{a}). \tag{6}$$

$(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{y}$: 3D vector, s: Scalar)

Since the perpendicular line is orthogonal to this line, it is necessary to satisfy the following equation.

$$(\boldsymbol{y} - \boldsymbol{x}) \cdot (\boldsymbol{b} - \boldsymbol{a}) = 0. \tag{7}$$

$(\boldsymbol{x}$: 3D vector)

The only unknown is s, and when s is found, the point y is found, and it is possible to determine the collision between the cylinder and the foot. At this time, if the point y exists on the line segment and the distance r' between the point x and the point y is smaller than r, it can be seen that the point y collides with the cylinder. Since $r - r'$ is the distance from the surface of the cylinder to the point x, this value is calculated using the penalty method.
If it is found that the cylinder does not collide, then the collision between the point and the sphere is determined. The collision between a point and a sphere can be determined by examining whether the distance between the points a and b and the point x is within r.
Collision determination between a capsule and a sphere (radius r'') is based on the same principle, and the above r is replaced with $r + r''$ for calculation and determination.

3. Simulation environment

This program was created using the DirectX SDK, which is a standard CG library, in a program development environment using the C / C ++ language. Microsoft Visual Studio Community 2019 was used to create the program this time. It ran on Intel Core i7 (Windows 10 Pro).

4. Simulation result

The FB Euler method is used as the numerical calculation method. The simulation results are shown in the figure below (See Fig. 4.).

Fig. 4. Execution result by FB Euler method using the proposed method

Simulations were performed using the Euler method and the Runge-Kutta method. The Euler method resulted in the cloth moving violently and moving away from the body (See Fig. 5.). In the Runge-Kutta method, the movement of the animation became heavy, and the result was that the contact between the foot and the skirt caused the animation to dig in (See Fig. 6.).
In addition, all the calculation methods resulted in the animation moving being heavier than the execution result when the bones were not divided.

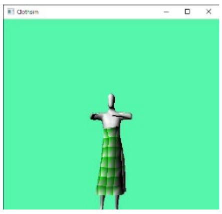


Fig. 5. Execution result by Euler method using the proposed method

Fig. 6. Execution result by Runge-Kutta method using the proposed method

5. Consideration

As a result of the simulation, the collision judgment considering the contact between the skirt and the body was stable by the FB Euler method as mentioned in the execution result of Section 4, and a simple animation could be performed. However, there are some issues:

- The parameters must be changed manually.
- Examination of numerical calculation method and contact force calculation method by implicit solution method.
- Examining the method of texture mapping to the skirt.
- Examination of a method with a small amount of calculation

6. Consideration

The purpose of this study was to dynamically express the cloth and to try a simple animation considering the contact between the body and the skirt with reference to past studies.
This study, the completed animation is a simple animation, the number of quality points is small, and there are many parts that cannot be implemented such as texture mapping and texture expression of cloth. It is still difficult to do a realistic clothing simulation.
Furthermore, we must pay attention to how to put clothes on 3D characters, how to express the difference in cloth materials, how to judge collisions with higher accuracy, and how to calculate less.
In addition, including the improvement measures mentioned above, we will work on CAD production that can simulate various movements of hula and dress design, etc., and express more realistic cloth, which will be a practical research. I want to aim for that.

References

1. Zorah Lahnar, Daniel Cremers and Tony Tung DeepWrinkles: Accurate and Realistic Clothing Modeling, *15th European Conference on Computer Vision*, 2018, pp. 1-18.
2. Ayumi Taniue, Attempt to create cloth animation with OpenGL, *Memoirs of the Faculty of Engineering*, University of Miyazaki, 2007, pp. 299-304.
3. Seira Nagatomo, Attempt of cross simulation by mass spring model, *Memoirs of the Faculty of Engineering*, University of Miyazaki, 2009, pp. 327-332.
4. Minaho Sakumoto, Attempt of cloth animation considering contact with an object, *Memoirs of the Faculty of Engineering*, University of Miyazaki, 2011, 223-227.
5. Mari Inoue, Research on the Mechanics of Fibrous Materials, SEN'I GAKKAISHI(Textiles and industry). Vol. 64. No.8, ,2008, pp.246-251.
6. Mikihiro Satou, Research on interactive cloth simulation, *Outline of the 37th Academic Lecture, Faculty of Industrial Science and Technology, Nihon University*. No.20, 2004, pp.23-26.

Analysis and Construction of Elements of the Stage Performance Structure in a Kabuki-dance

Miku Kawai
*Faculty of Software and Information Science, Iwate Prefectural University,
152-52, Sugo, Takizawa, Iwate, 020-0693, Japan*

Jumpei Ono
*Faculty of Software and Information Technology, Aomori University,
2-10-1, Seishincho, Edogawa, Tokyo, 134-0087, Japan*

Takashi Ogata
*Faculty of Software and Information Science, Iwate Prefectural University,
152-52, Sugo, Takizawa, Iwate, 020-0693, Japan
g031p035@s.iwate-pu.ac.jp, j.ono@aomori-u.ac.jp
ogata@iwate-pu.ac.jp*

Abstract

We conducted a detailed analysis of the stage performance structure of *Kyōganoko Musume Dōjōji* and reproduced it on the animation tool KOSERUBE with the music, lyrics, and images aligned. In this paper, 11 scenes of *Kyōganoko Musume Dōjōji* are organized. The actual flow is compared with the movement on the system. The goal of this paper is to understand the entire stage structure of *Kyōganoko Musume Dōjōji* in detail. In addition, this research aims to use the stage performance structural system for an integrated narrative system that the authors are developing. We believe that this system will be the basis for the construction of the Kabuki stage. Further, we want to apply it not only to Kabuki but also to other genres, such as computer games and commercials.

Keywords: Dōjōji, Stage Performance Structure, Narrative generation, Story

1. Introduction

Kabuki is a valuable cultural heritage, one of the elements of Japanese culture with a long history. Therefore, many studies have been conducted regarding the preservation and inheritance of such a heritage. In addition, Kabuki incorporates various performing arts from the genres of dance and classical performance, such as Noh and Kyogen. In recent years, we have worked actively to incorporate video technologies such as Computer Graphics. Additionally, in the vortex of COVID-19, a new performance form called Zūmu Kabuki was created.

Zūmu Kabuki is a Kabuki stage in which each actor takes a video on the "Zoom" video calling app, and the videos are put together to form a single stage. This is also known as online Kabuki. In fact, from the end of June to the middle of July, the *Zūmu Kabuki Chūshingura* was performed. In this way, Kabuki continues to develop by utilizing artificial intelligence and digital technology.

For example, last year, Hatsune Miku and Nakamura Shido co-starred at the Nico Nico Chōkaigi*, where a Kabuki called "Hana kurabe sen bon zakura" was

*A festival organized by DWANGO Co. Ltd. The festival has the content "Chō Kabuki" (https://chokabuki.jp/2019minamiza/english/).

performed. This is a stage that combines the world view of Hatsune Miku's representative song *Senbonzakura* with Kabuki (Minamiza, Kyoto, Aug. 2019). In addition, this year, the Ningyo Joruri *Koimusume Tsumugi no Labyrince*, using Hatsune Miku, is being held at the Tomioka Silk Mill (Tomioka, Gunma, Japan/Oct. 2020).

In Kabuki research, there have been many studies of the introduction of video technology. The heritage of Shosagoto (the body actions of actors) using motion capture has been analyzed in some studies. However, our research is different.

・Oda and Genda study the characteristic motion in Kabuki using a motion capture system. Omoto et al. create a model of Minamiza in Kyōto (Minamiza is a traditional Kabuki theatre).[1]

・A study (Kobayashi, 2017) compares Western dance in the ballet *La sylphide*, and the Japanese dance *Kyōganoko Musume Dōjōji* to investigate the creation of dances. This research focuses on "Furi <swing>," mainly analyzing small movements such as walking and hand expressions.[2]

・Using a commercially available robot (Desktop robot "Premade AI"), the stage of Ningyō Jōruri *Hidakagawa Iriaizakura* is reproduced remotely by the Tokyo University of Technology.[3]

We investigated and analyzed several of the elements that make up Kabuki and developed a system that simulates the interplay of stories. We use a system called KOSERUBE that partially reproduces the visual image of the stage performance structure of *Kyōganoko Musume Dōjōji.* In addition, we have conducted the following research so far.

- We have surveyed and analyzed Dōjōji stories and *Kyōganoko Musume Dōjōji* in Kabuki.[4]
- We have developed a system that simulates the stage performance structure of *Kyōganoko Musume Dōjōji.*[5]
- We have integrated narrative generation systems, including explanation generation, with the above simulation system.[6]

The current study focuses on the stage structure and aims to display the entire stage. Specifically, the purpose is to perform a detailed analysis of the stage structure and reproduce the stage performance structure. Therefore, this study does not necessarily analyze the body actions of the actors in detail. *Kyōganoko Musume Dōjōji* (The Maiden at the *Dōjōji* Temple) is one work in many works based on the "legend of Dōjōji," told in narratives and picture scrolls. This is a latter-day description of the legend of Anchin and Kiyohime. The oldest legend of the Dōjōji work is *Dainihonkoku Hokekyō Genki* (Miraculous Tales of the Lotus Sutra from Japan).[1] In the story, Anchin lies and is burned to death by Kiyohime; however, the power of the Lotus Sutra frees him from suffering.

We have conducted the following research so far.

1. As a study of Kabuki-dance: Analysis of *Kyōganoko Musume* Dōjōji.[4]
2. Analyze the stage performance structure of *Kyōganoko Musume Dōjōji:* We have developed a performance structure reproduction system.[5]
3. A combination of Unchiku, an explanation generation system, and an animation system.
4. Story generation related to love and sexuality using a Japanese classic tale.

We extracted the components of the story for use in our explanation generation system for a Kabuki-dance representation system.

KOSERUBE[7, 8] is an animation tool developed by the authors. We develop a system that reproduces the stage performance structure of *Kyōganoko Musume Dōjōji* with KOSERUBE. This research aims to automate the reproduction system of the stage performance structure and connect the system to story generation.[9] To that end, this paper focuses on the dance scenes of *Kyōganoko Musume Dōjōji* and analyzes their stage elements. *Kyōganoko Musume Dōjōji* focuses on dance, but it involves scenes with narratives. However, in this study, we focus only on the dance scenes. Therefore, we have not analyzed all the scenes in *Kyōganoko Musume Dōjōji*.

2. Analysis of *Kyōganoko Musume Dōjōji*

Dōjōji is a temple of the Tendai sect in Hidakachō, Hidaka, Wakayama, Japan. This temple is the subject of the legend of Dōjōji, which has been recorded for over a thousand years. Although this legend has changed its shape over the long history, we think it is still known to many people. Of the versions of this legend, we focus on the Kabuki dance *Kyōganoko Musume Dōjōji*, which is considered the legend of modern Dōjōji.

2.1. *Kyōganoko Musume Dōjōji*

Kyōganoko Musume Dōjōji is a representative work of Kabuki dance, known as a work played by representative female onnagata such as Nakamura Utaemon 6th and Tamasaburo Bando 5th. The ancient legend that led to the Kabuki dance *Kyōganoko Musume Dōjōji* was transmitted through Buddhist narratives, Noh, texts, picture scrolls, etc.

Kyōganoko Musume Dōjōji is a work that is a later development of the legend of Anchin and Kiyohime. Unlike the Kiyohime legend, which depicts a woman's obsession, she does not bring it to the fore in this version, using various kimonos and props to show her daughter in love, making it a completed work of Musume Dōjōji as a Kabuki dance. That is, it basically inherits the form of the Noh *Dōjōji* intact, but the legend of Dōjōji is not specifically shown in the verse. The theme is to show a changing dance.

The foundation of Dōjōji dance had already been laid in Noh's *Dōjōji*, and in *Kyōganoko Musume Dōjōji* it is transformed into a gorgeous and bright work, positioned as the completion of the modern transformation.

2.2. *The legend of Dōjōji*

The story of the legend of Dōjōji (also known as the legend of Anchin and Kiyohime) is Kiyohime, a woman living in Kumano who is betrayed by a Buddhist monk (Anchin). Her strong anger transforms her into a snake, and she burns the monk together with a bell in Dōjōji.

The legend of Dōjōji took a male point of view of a woman as evil, but since the Muromachi period, it has changed to the point of view of a woman's sadness.

2.3. *Analysis of Kyōganoko Musume Dōjōji*

Table 1 (following the references) shows the elements of *Kyōganoko Musume Dōjōji* and their implementation methods. The table includes the following five items for each scene:

(1) [Actual stages] Flow and characteristics of the dance on stage (the contents in brackets indicate scene characteristics);

(2) [Music] Music used in the system, based on the Kineie Shamisen score, and created with a free composition tool ("Wagaku Hitosuji," https://jonkara.net/soft/wagaku/). However, the system uses the actual music because there are no scores for "Michiyuki," "Ranbyōshi," and "Kaneiri." In other scenes, we only reproduced the music for the shamisen. Since the sound of the shamisen music feels higher than that on the actual stage, there is a lingering sense of discomfort compared with the actual music;

(3) [Lyrics] Excerpts from the beginning and end of the lyrics (only the "Ranbyōshi" is shown in full). The lyrics are romanized because they are written in old Japanese;

(4) [Images] Images of appearing characters, the main stage, and stage equipment. The props are listed in parentheses in the "Shirabyōshi Hanako";

(5) [Commands] The command is written in the code. Commands used in the system (input sentences instructing the reproduction system to create the character animation or the background).

3. An Animation Mechanism based on the Structural Analysis of Scenes

This section describes the generation of stage performance structures based on our analysis of *Kyōganoko Musume Dōjōji*. As we mentioned in Section 1, our goal is to create a system that automatically generates a staging structure from certain inputs. This study is a preliminary phase. To study the system for automatic generation, we created an animation from the content shown in Table 1. First, this section describes an animation tool. Next, it shows the flow of creating an animation.

This study uses the user interface of KOSERUBE, which has two modules. The first module converts the generated stories into animation code specific to KOSERUBE, based on which the second module plays the animation.

In this study, we used the second module in the user interface and the date base of images. Moreover, additional music and images were prepared for the animation of *Kyōganoko Musume Dōjōji*. We manually created the animation code based on the analysis shown in Table 1. Next, we described how to create the animation code.

Figure 1 shows an example of an animation. The system reproduces a performance by combining a background and a character image on another image that imitates the play's stage. In addition, the music in *Kyōganoko Musume Dōjōji*, along with the text showing the story and lyrics of the music, are displayed at the bottom of the screen. Our research is not on computer graphics. Rather, the goal is to reproduce the elements that make up the entire stage. For this reason, the paper does not pursue reality and the animations that we created are below par.

Fig. 1. Example of a visualized stage performance structure (Ranbyōshi).

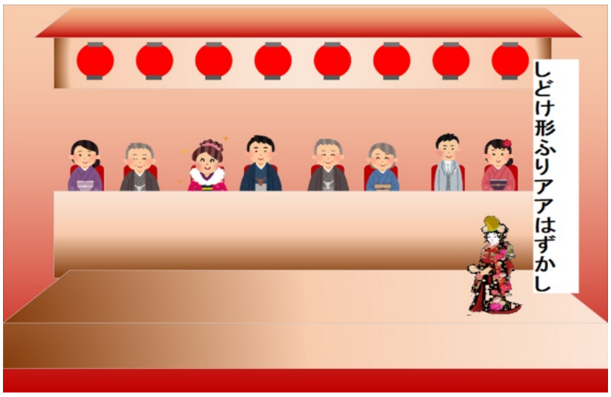


Fig. 2. Example of a visualized stage performance structure (Michiyuki).

The screen structure shown in Figures 1 and 2 uses background and human images stored in the image database of KOSERUBE. The screen consists of four layers: text, character, stage, and background. Here, the dancing figures are the characters, and the instrumentalists, monks, and bells are included in the background.

The system uses two kinds of music. First, the music of the actual *Kyōganoko Musume Dōjōji* and second, the electronic music created from a musical score. The musical score used was *Nagauta Kyōganoko Musume Dōjōji: Shamisen Bunka-fu* by Kineie.[10] To create the music, we used a free composition tool called "Wagaku Hitosuji."

Next, we explain the details of the animation code. The animation code is a set of instructions to the user interface describing the combination of images and texts, detailed instructions for character animation, and the timing of playing music. The user interface section dynamically composes images, plays animations, and plays music based on the animation code.

Figure 3 is an example of character animation based on the animation code. The code represents an animation that makes the character jump in Figure 3. The "mov" command in Figure 3 is a primitive command for the character's image to move. We used multiple "moves" to represent the dances of the characters. For example, "mov, 11, v, −50, −25" means, from left to right, "Command," "Character ID," "Direction of movement," "movement distance," and "movement speed." For example, "mov, 11, v, −50, −25" moves "Shirabyōshi Hanako," with a character ID of 11. The movement distance is the −50 coordinate in the vertical direction relative to the current coordinates. The movement speed is −25 (every frame, the coordinates of the image move relative to the current coordinates by −25 pixels).

In Figure 3, the coordinates of the character image are moved up and down by multiple "moves." The "Code" in Figure 3 is a combination of "mov" for moving up and "mov" for moving down repeated three times, so that "Shirabyōshi Hanako" jumps off three times.

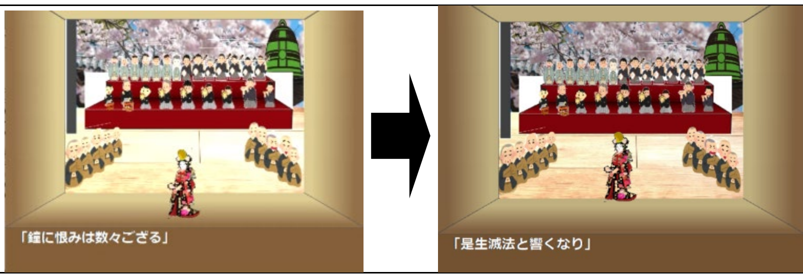
Code	Animation
mov, 11, v, −50, −25 mov, 11, v, 50, 25 mov, 11, v, −50, −25 mov, 11, v, 50, 25 mov, 11, v, −50, −25 mov, 11, v, 50, 25	

Fig. 3. An Example of Code for Visualizing Animation.

4. Discussion

The future potential of the animation system used in this paper and how to use it are described below.

4.1. *Results*

We have analyzed the stage performance structure of *Kyōganoko Musume Dōjōji* and reproduced it with animation. In addition, the story of the legend of Dōjōji is organized and summarized in a table. We analyzed and organized the stage performance structure, including such elements as musical instruments and stage backgrounds, and created animation using KOSERUBE.

We also analyze the history and story of the legend of Dōjōji, focusing on the transformation of the story from the legend of Dōjōji to *Kyōganoko Musume Dōjōji* and the personality of the characters.

The future of the animation system and the prototype system using the story of Dōjōji are described below.

4.2. *Future Work: Combination of INGS and the Animation System*

The authors believe that this system will be the basis for creating the Kabuki stage. Figure 4 is an image of the system. The system outputs the stage performance structure by automatically generating the stage components based on the input story. The story entered in the system is prepared based on two methods, creation by hand and automatic generation. We have developed a system that automatically generates stories, the Integrated Narrative Generation System (INGS). We are studying a system that automatically generates the stage performance structure by combining the reproduction system with INGS. Finally, we want to apply the system not only to Kabuki but also to other genres, such as computer games and commercials.

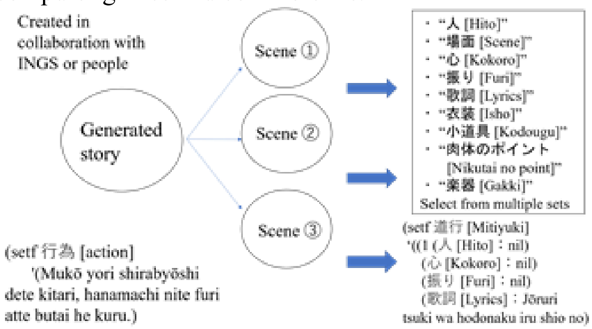


Fig. 4. Schematic of the stage structure reproduction system.

4.3. *Future Work: Parallel Generation System*

As mentioned in Section 2, the story of the legend of Dōjōji has changed during its long history. In addition, the character of Kiyohime has changed. Therefore, we reproduced the character of Kiyohime, which changes with the transformation of the story, with 2D and 3D animation.

In the legend of Dōjōji, the outline of the story is the same, but there are turning points in the story, such as the scene where Kiyohime turns into a snake and the way it ends. Scenes from the legend of Dōjōji, such as the turning point and turning into a snake, were reproduced with 2D animation and 3D animation. The goal of this system is for all scenes to be seamlessly represented using CG.

In particular, based on branches, many story scenes are generated in parallel, using the story units of the legend of Dōjōji. The system can also make one story line by selecting a possibility at each branch point.

Figure 4 shows an image of the prototype system,[11] in which the 3D and 2D generation results for 16 scenes flow. In the first and last scenes, a part of the 2D animation system is played for about 30 seconds. In the 2D animation system, it represents a scene of *Kyōganoko Musume Dōjōji*.

The flow of the system starts from the first scene, after which the story is divided into multiple scenes. The scene expressed in 3D is in the *Hokkekyō Genki*[12] and *Konjaku Monogatari*,[13] containing the narratives of the legend of Dōjōji and the story of *Dōjōji Engi*. Then, returning to the first scene, the story ends.

In the scene where only the code is displayed, only the structure of the generated story is flowing because 3D has not been created yet.

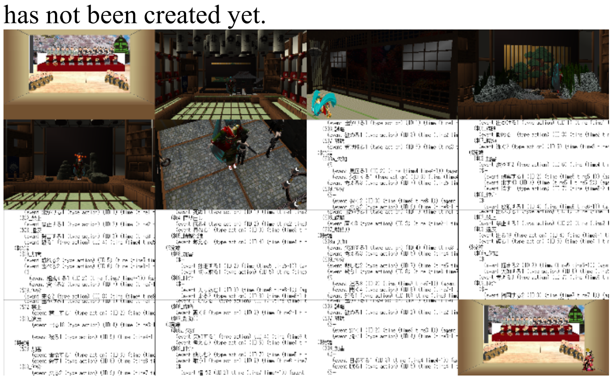


Fig. 5. Images of stage structure reproduction system.

5. Conclusion

In this paper, 11 scenes of *Kyōganoko Musume Dōjōji* are organized to provide a detailed description of the entire stage structure of a Kabuki dance performance. In this study, we conduct a detailed analysis of the stage performance structure of *Kyōganoko Musume Dōjōji* and reproduce it with the animation tool KOSERUBE by aligning the music, lyrics, and images. The purpose of this analysis is to develop a system for reproduction of the stage performance structure of the dance. In the future, we will study its generation method based on the results of the analysis and develop an approach to build a more detailed reproduction system.

Acknowledgements

The research for this chapter was supported by the Japan Society for the Promotion of Science (JSPS KAKENHI), Grant No. 18K18509.

References

1. N. Omoto, K. Hasegawa, H. Motojima, S. Nagata and S. Tanaka, 3D model of Kyoto Minami-za theater and its application Prof, IPSJ SIG Computers and the Humanities, 2004, pp. 109-112.
2. Kobayashi. S, Approach to choreography "Kiyohime": a comparative study of Japanese traditional dance and western ballet for female (Doctoral dissertation), Osaka University of Arts, 2018.
3. R. Dong, Y. Chen, D. Cai, S. Nakagawa, T. Higaki and N. Asai, Robot motion design using bunraku emotional expressions: focusing on Jo-Ha-Kyū in sounds and movement, Advanced Robotics, 34(5), 2019, pp.299-312.
4. M. Kawai, J. Ono and T. Ogata, Analysis and Generation of the Narrative Elements in a "Dojoji" Story. *Proceedings of the 3rd International Conference on Computational Intelligence and Intelligent Systems*, 2020, pp.44-45.
5. M. Kawai, J. Ono and T. Ogata, Analyzing the Stage Performance Structure of a Kabuki-Dance, Kyōganoko Musume Dōjōji, Using an Animation System, I*n E. Ishita, N. L. S Pang & L. Zhou (Eds.) Digital Libraries at Times of Massive Societal Transition (LNCS 12504)*, pp. 248-254.Switzerland: Springer Nature, 2020.
6. J. Ono, M. Kawai and T. Ogata, Implementation of an Explanation Generation Mechanism Using Attribute Frames and a Noun Conceptual Dictionary. *Proceedings of the 3rd International Conference on Computational Intelligence and Intelligent Systems*, 2020, pp.48-50.
7. S. Imabuchi, J. Ono, J. Endo, Y. Kurisawa, R. Michimata, M. Kamada, T. Akimoto and T. Ogata, KOSERUBE: Folk Tales of Iwate: An Integrated Generation System of Language/Music/Image Based on Propp-based Story Grammar Transactions, *Proc. of the Japanese Society for Artificial Intelligence*, 2012.
8. S. Imabuchi, T. Akimoto, J. Ono and T. Ogata, KOSERUBE: An Application System with a Propp-based Story Grammar and Other Narrative Generation Techniques, *Proc. of the Symposium on Cryptography and Information Security*, 2012.
9. S. Imabuchi, T. Akimoto, J. Ono and T. Ogata, KOSERUBE: An Application System with a Propp-based Story Grammar and Other Narrative Generation Techniques, *Proc. of the Symposium on Cryptography and Information Security*, 2012.
10. Kineie Yashichi IV, Nagauta Kyōganoko Musume Dōjōji: Shamisen Bunka Fu Hōgakusha Beppu, Oita, Japan, 1952, Feb. 2018, pp. 174-177.
11. M. Kawai, J. Ono and T. Ogata, Dual Story Generation Based on Love and Extreme Emotions, *Proceedings of 2020 The 3rd International Conference on Computational Intelligence and Intelligent Systems*, 2020.
12. M. Inoue and S. Ōsone (Collation Eds.), *Dainipponkoku Hokkekyō Genki,* Ōjōden Hokkegenki, Iwanami Shoten, Tokyo, 1974, pp. 43-219.
13. K. Mabuchi, F. Kunisaki and Toru Konno, Konjalkumonogatarisyu vol.1., Shogakukan, 1919.

Table 1. Stage performance structure.

Scene name	Actual stages	Music	Lyrics	Images	Commands
Michiyuki	Shoke walks from Hanamichi. Talk about Kane's memorial service on stage. A stage device imitating the door to the temple will be prepared on the stage. The narrator and the shamisen player come out and sit on a long table.	Existing music (Jōruri)	"Tsuki ha hodonaku hairu Shiono (omission) Hirari bōshi no fuwafuwa to"	・Characters: Shoke (12 people) ・Katarite ・Musicians: Shamisen [Japanese guitar] ・Place on stage: Main stage, Hanamichi ・Stage equipment: Dōjōji, Kane, Cherry blossoms	Not reproduced at the time of this paper.
	Shirabyōshi Hanako comes out of Hanamichi with a fan. Perform a dance along with Shirabyōshi. After the dance, head to the stage.	Existing music (Noh)	"Shidokenari furi ā hazukashi ya (omission) Dōjōji ni koso tsukinikere"	・Characters: Shirabyoushi Hanako, Shoke (12 people) ・Katarite ・Musicians: Shamisen [Japanese guitar] ・Place on stage: Main stage, Hanamichi ・Stage equipment: Dōjōji, Kane, Cherry blossoms	
Ranbyōshi	Put on the Eboshi given by Shoke. Move from the stage to Hanamichi according to Noh. Show off the dance. After that, return to the stage from Hanamichi. Perform a dance at the center of the stage (Mai has many small movements of wrists, feet, and props.)	Existing music (Noh)	"Hana no hoka ni ha matsu bakari hana no hoka ni ha matsu bakari kuresomete kane ya hibikuran"	・Characters: Shirabyōshi Hanako, Shoke (12 people) ・Katarite ・Musicians: Hyōshiita ・Place on stage: Main stage, Hanamichi ・Stage equipment: Dōjōji, Kane, Cherry blossoms	"Hana" command makes it look like he is going back and forth between the stage and Hanamichi.
Chūkeinomai	Continuing from Ranbyōshi, perform a dance at the center of the stage in tune with Kanedukushi.	Create from score (Nagauta)	"Kane ni urami ha kazukazugozaru (omission) Shinnyo no tsuki wo nagameakasan"	・Characters: Shirabyōshi Hanako (Have Chūkei), Shoke (12 people) ・Katarite ・Musicians: Shamisen [Japanese guitar], Taiko [Drum] ・Place on stage: Main stage ・Stage equipment: Dōjōji, Kane, Cherry blossoms	Not much movement, many small dances.
Teodori	Reappear on the stage without props. Perform at the center of the stage along with Hauta (Shōfu no Erothishizumu).	Create from score (Nagauta)	"Iwazu kataranu waga kokoro (omission) Miyakosodachi wa hasuppana mono ja e"	・Characters: Shirabyōshi Hanako, Shoke (12 people) ・Katarite ・Musicians: Shamisen [Japanese guitar], Taiko [Drum], Kodutumi, Fue [Flute] ・Place on stage: Main stage ・Stage equipment: Dōjōji, Kane, Cherry blossoms	Not much movement, many small dances.
Mariuta	Perform dance at the center of stage in accordance with Rō-dzukushi (Yūjo) (choreography like playing with Kakū no Mari).	Create from score (Nagauta)	"Koi no wake sato (omission) Omoi sometaga en ja e"	・Characters: Shirabyōshi Hanako, Shoke (12 people) ・Katarite ・Musicians: Shamisen [Japanese guitar], Taiko [Drum], Kodutumi, Fue [Flute], Kane ・Place on stage: Main stage ・Stage equipment: Dōjōji, Kane, Cherry blossoms	By repeating the vertical movement, it is possible to express that a ball is attached.
Hanakasaodori	Put Hanakasa on the head and hands at the center of the stage. Perform the dance according to Wakitebushi. Shirabyōshi Hanako and disappears from the stage.	Create from score (Nagauta)	"Ume to-san-san sakura wa (omission) Kawairashi-sa no hanako"	・Characters: Shirabyōshi Hanako (Hanakasa), Shoke (12 people) ・Katarite ・Musicians: Shamisen [Japanese guitar], Taiko [Drum], Fue [Flute], Kane ・Place on stage: Main stage ・Stage equipment: Dōjōji, Kane, Cherry blossoms	The character does not move much because the movement of Hanakasa is the main movement.
	All Shoke line up on stage with Hanakasa. Show off the dance.			・Characters: Shoke (12 people) ・Katarite ・Musicians: Shamisen [Japanese guitar], Taiko [Drum], Fue [Flute], Kane ・Place on stage: Main stage	

Scene name	Actual stages	Music	Lyrics	Images	Commands
				・Stage equipment: Dōjōji, Kane, Cherry blossoms	
Kudoki	Appear with Tenugui at the center of the stage. Show off Mai along with Koi to Shitto no Kouta.	Create from score (Nagauta)	"Koi no tenarai tsui minaraite (omission) Sawaraba ochin fuzeinari"	・Characters: Shirabyōshi Hanako (Tenugui), Shoke (12 people) ・Katarite ・Musicians: Shamisen [Japanese guitar], Taiko [Drum], Kodutumi ・Place on stage: Main stage ・Stage equipment: Dōjōji, Kane, Cherry blossoms	Tenugui is wrapped around the body or held in the mouth. In this way, the character is not moving because the focus is on using props.
Yamadukushi	Appear with Kakko close to the center of the stage. Show Mai while hitting Kakko according to Yama-Zukushi.	Create from score (Nagauta)	"Omoshiro shikinonagam eya (omission) Tsuki no kaobase Mikasayama"	・Characters: Shirabyōshi Hanako (Kakko), Shoke (12 people) ・Katarite ・Musicians: Shamisen [Japanese guitar], Taiko [Drum], Kodutumi, Hyōshiita, Kakko ・Place on stage: Main stage ・Stage equipment: Dōjōji, Kane, Cherry blossoms	Rotational movements using Kakko.
Teodori	Reappear on the stage with nothing. Perform at the center of the stage along with Hauta (Shōfu no Erotishizumu).	Create from score (Nagauta)	"Tada tanome (omission) Niku terashi hodo itoshirashi"	・Characters: Shirabyōshi Hanako, Shoke (12 people) ・Katarite ・Musicians: Shamisen [Japanese guitar], Taiko [Drum], Fue [Flute] ・Place on stage: Main stage ・Stage equipment: Dōjōji, Kane, Cherry blossoms	Not much movement, many small dances.
Shuzudaiko	Around the center of the stage, hit Suzudaiko and perform a dance along with taueuta (Saotome odori).	Create from score (Nagauta)	"Hana ni kokoro o fukamigusa (omission) Hiki ka dezuite zo use ni keru"	・Characters: Shirabyōshi Hanako (Suzudaiko), Shoke (12 people) ・Katarite ・Musicians: Shamisen [Japanese guitar], Taiko [Drum], Kodutumi, Hyōshiita, Suzudaiko ・Place on stage: Main stage ・Stage equipment: Dōjōji, Kane, Cherry blossoms	Suzudaiko is difficult to reproduce because it turns its wrist and rings.
Kaneiri	Shirabyōshi Hanako defeats all Shoke. Kane comes down on stage. Shirabyōshi Hanako climbs on Kane and gazes around her like a snake.			・Characters: Shirabyōshi Hanako, Shoke (12 people) ・Katarite ・Musicians: Shamisen [Japanese guitar], Suzudaiko ・Place on stage: Main stage ・Stage equipment: Dōjōji, Kane, Cherry blossoms	Not implemented because it is not a dance part.

Unchiku Generation with Moving Illustration Using Kabuki Knowledge

Jumpei Ono
Faculty of Software and Information Technology, Aomori University,
2-10-1, Seishincho, Edogawa, 134-0087, Japan

Miku Kawai
Faculty of Software and Information Science, Iwate Prefectural University,
152-52, Sugo, Takizawa, 020-0693, Japan

Takashi Ogata
Faculty of Software and Information Science, Iwate Prefectural University,
152-52, Sugo, Takizawa, 020-0693, Japan
E-mail: j.ono@aomori-u.ac.jp, g031p035@s.iwate-pu.ac.jp
t-ogata@iwate-pu.ac.jp

Abstract

We developed an explanation-generation system based on narratology. We also developed an *unchiku* generation system as an application. After creating a prototype, we then considered a new system that would combine the *unchiku* generation system with an animation system. We implemented the animation system in our kabuki study. This paper considers the possibilities of the *unchiku* generation system.

Keywords: *Kyōganoko Musume Dōjōji, moving illustration, narrative generation, unchiku*.

1. Introduction

Drawing on the foundations of narratology and literary theory, systematized thought, and philosophy, we developed a program that generates narratives.[1,2] We then designed a mechanism for *unchiku* generation[3,4] for integration into the narrative generation system.

We study explanation generation in the context of narratology. *Narrative* means not just the progression of a plot but also the surplus or excess exposition that can interfere with its rapid progress. In Gérard Genette's theories on narrative discourse,[5,6] such techniques—for example, explanation and description—are seen as halting the temporal progression of the narrative. *Explanation* is the act of justifying various objects or events in a story (e.g., characters, setting, physical objects, abstract ideas, etc.) and their relationships; it often clarifies but remains separate from the progress and development of the storyline. The *unchiku* is a form of explanation that uses deep content to share knowledge or teach a lesson to a disinterested audience.

We also study kabuki narrative generation by comprehensively examining the structure, methods, and techniques of narrative in kabuki to refine our narrative generation system.

In our research system of narrative generation,[1,2] the kabuki form is central to the method of narrative generation. We have placed the "work performed on stage" at the center and presented a list of its multiple components,[7] extracted representative kabuki methods and techniques, and analyzed them in detail through prototypes[8,9] to further our narratological research in the kabuki tradition.[10]

We focused our analysis specifically on the kabuki dance *Kyōganoko Musume Dōjōji*,[11–14] a later tale of the legend of Dōjōji that has been performed by many excellent *onnagata* (male actors playing female characters) since ancient times with new content added beyond the original legend. In our earlier work[13,14] we used a system we developed called KOSERUBE,[15] an animation tool for a story generation system, to create visual representations of the structure of the stage performance.

Other researchers have also built narrative generation systems based on the work of Pablo Gervás[16] as part of the broader computational creativity movement that focuses on narrative applications in the field of robotics—for example, developing socially assistive robots with a narrative approach.[17] Our particular research proposes a method of combining moving illustrations on the screen as a tool to inform users of *unchiku*. In this paper, *story* means the combined storyline and explanation, and the *unchiku* is the detailed information within the explanation.

There are many possible ways of using moving illustrations. We previously described our approach to one of these methods.[13,14] In this paper, we examine that method's opposite: using images as explanations of texts. The overall concept of using pictures to explain the text is not new—encyclopedias and dictionaries have done this for centuries. However, this paper proposes the use of an automated animation tool to explain *Kyōganoko Musume Dōjōji* text.

In describing our prototype, we will clarify the concept and examine the issues and usage for the future. Our final goal is to integrate the explanation using video and the explanation of the video and to treat each as a function.

2. Two Background Systems

This section describes the animation system of the *Kyōganoko Musume Dōjōji* and the *unchiku* generation system. The modules of the prototype system use these two systems.

2.1. *The Animation System of the Kyōganoko Musume Dōjōji*

The first system is an animation system that depicts the stage performance structure. The data on *Kyōganoko Musume Dōjōji* used in this study came from our analysis so far. First, we focused on three items in the stage structure table—*kokoro* (core conceptual theme), *furi* (performance), and lyrics—that are described in *Musume Dōjōji.*[1] We analyzed the elements in the stage structure table to build a storyboard.[2]

Referring to an analysis of *Musume Dōjōji*[18] by Tamotsu Watanabe, a contemporary kabuki researcher and critic, we conducted a detailed analysis of the "stage performance structure" (narrative outline) of the play. The main components were the characters, background (stage set), music (instruments, performers, and genres), lyrics, dialogue, and the core spiritual themes of each scene.

Musume Dōjōji recounts the legend of Kiyohime and Anchin. In the legend, a young woman named Kiyohim falls in love with Anchin, a Buddhist monk, who rejects her. She turns into a snake and kills him. After more than a thousand years of telling this story through oral histories, picture scrolls, and other reading materials, the story has transformed into the present-day *Kyōganoko Musume Dōjōji*. Our narrative generation system is based on its Kabuki dance. We summarized the stage performance structure of *Kyōganoko Musume Dōjōji* in an analysis table and created a storyboard to develop our animation system.

We have previously written about our analysis of the stage performance structure of *Kyōganoko Musume Dōjōji* in detail[11] and how we developed an animation system to express it called KOSERUBE, a folktale-style story generation system.[15] With the system, users can select folktales and characters, and it automatically generates a story based on the selections. The system also generates an animation for the generated story, displaying the lyrics at the bottom of the screen; the background and characters change from scene to scene. Our animation system can also reproduce and generate music[12] based on the score of *Syamisen Bunkahu Nagauta Kyōganoko Musume Dōjōji*[19] and recorded music stored in the system.[20] The prototype system uses the analysis of *Kyōganoko Musume Dōjōji*, the story of *Dōjōji Enki*, and the attributes of the characters. Details are given in Section 3.1.

2.2. *The Unchiku Generation System*

The second system is an *unchiku* generation system. *Unchiku* means deeply preserved knowledge. *Unchiku wo tareru* (to tilt one's *unchiku*) means putting all your energy into that knowledge. In recent years, *unchiku* is often used to mean "trivia." However, *unchiku wo tareru* (to draw upon one's *unchiku*) and *unchiku wo hikerakasu* (to show off one's *unchiku*) mean showing off knowledge learned from undesirable situations. The story uses *unchiku* as one of its rhetorical techniques. A story has a rhetorical device that slow or stop the progress of the plot in the story by, for example, pausing to describe something in the story. For the characters in the story, "tilting one's *unchiku*" is the act of explaining something (e.g., an object, character, setting, or event, etc.) in detail. We have developed the *unchiku* generation system with the aim of establishing a "starting point and means for systematically collecting, accumulating, and utilizing kabuki knowledge."[3,4]

As mentioned earlier, *unchiku* is one rhetorical technique used in *Musume Dōjōji*. We have previously discussed various aspects of narrative discourse, including the "distance" between the storyteller and the story,[6] based on the structuralist theories proposed by Genette.[5] We position *unchiku* as an explanation for excess quantities verbosity. Originally, we extracted information from Wikipedia to develop our *unchiku* generation module,[3,4] but we have refined our *unchiku* generation in the prototype system. Details are given in Section 3.2.

3. *Unchiku* Generation with Moving Illustration

Combining the *unchiku* generation with the insertion of moving illustrations tells the story as animation. Our goals for the prototype system were (1) to clarify the system image and (2) to analyze plans and issues for future full-scale development. This section describes the data structure and the system.

3.1. *Knowledge Structure for the Proposed System*

The prototype system limits the text and animation material of the unchiku to *Kyōganoko Musume Dōjōji*. For this reason, we have provided English supplements to the actual data and generated examples for this paper.

Figures 1 and 2 show the knowledge for *unchiku* generation: Figure 1 relates to elements in the story; Figure 2 relates to the scene of *Kyōganoko Musume Dōjōji* associated with the knowledge listed in Figure 1. Figure 3 shows the story used in this paper.

(道成寺
『京鹿子娘道成寺』である。『京鹿子娘道成寺』とは歌舞伎舞踊の代表的な作品である。昭和の六代目中村歌右衛門や五代目坂東玉三郎など、代表的な女形が演じる作品として知られている。古くからある伝説が、仏教説話や能、絵巻等の読み物を経て、歌舞伎舞踊の『京鹿子娘道成寺』に至る。[*Kyōganoko Musume Dōjōji* is a representative work of Shosagoto. It is a work played by representative onnagata such as Nakamura Utaemon VI and Bandō Tamasaburō V from the Showa period. An old legend leads to the kabuki dance *Kyōganoko Musume Dōjōji* through reading materials such as Buddhist narratives, Noh, and picture scrolls.]

Fig. 1. The *unchiku* knowledge of the *Dōjōji*

(乱拍子 [Ranbyōshi])
(人 [Character] 白拍子花子 [Shirabyōshi Hanako])
(心 [Core mental theme] 娘 [Young girl])
(振り [Performance] 白拍子の舞 [Shirabyōshi no mai])
(歌詞 [Music genre] 能[Noh])
(楽器 [Instrument] 拍子板 [Hyōshi])
(場面 [Scene] 桜の木（道成寺）[Cherry blossoms (Dōjōji)])
(衣装 [Costume] 赤 [Red])
(小道具 [Prop] 中啓 [Chūkei])
(肉体のポイント[Focus point in the dance] 足 [Foot]))

Fig. 2. The unchiku knowledge of the *Kyōganoko Musume Dōjōji*

年老いた僧と安珍が熊野に参詣へ行く。年老いた僧と安珍がある家に到着する。清姫が安珍に会う。清姫が安珍に惚れる。安珍は清姫を恐れる。安珍は「安珍と清姫が再会する」という嘘をつく。安珍はある家を過ぎる。清姫は嘘を知る。清姫は激怒する。清姫は蛇体に変身する。清姫は安珍を追う。安珍は道成寺に到着する。年老いた僧と安珍が熊野に参詣へ行く。年老いた僧と安珍がある家に到着する。清姫が安珍に会う。清姫が安珍に惚れる。安珍は清姫を恐れる。安珍は「安珍と清姫が再会する」という嘘をつく。安珍はある家を過ぎる。清姫は嘘を知る。清姫は激怒する。清姫は蛇体に変身する。清姫は安珍を追う。安珍は道成寺に到着する。[An old monk and a young monk, Anchin, go to visit Kumano. They arrive at a house. Kiyohime meets Anchin. Kiyohime falls in love with Anchin. Anchin is afraid of Kiyohime. Anchin tells a lie that "Anchin and Kiyohime will meet again." Anchin avoids the house. Kiyohime knows a lie. Kiyohime gets angry. Kiyohime transforms into a snake. Kiyohime chases Anchin. Anchin arrives at Dōjōji Temple.]

Fig. 3. The *Musume Dōjōji* story

3.2. *Prototype System*

In the prototype system, the *unchiku* generation is supplemented by inserting moving illustrations to produce animation: the story and *unchiku* appear as text with animation in the middle of the sentences associated with the words. Figure 4 shows the process of the system and its two modules. The first module creates a story with the *unchiku*. The other generates animations keyed to parts of the generated story. The system also contains a textual knowledge base and an image knowledge base. The system receives a story, then animates the story and the knowledge contained in it.

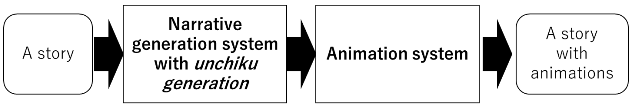


Fig. 4. The flow of a *unchiku* generation system with moving illustrations

The knowledge generation module takes a story and adds to it from the knowledge database. The *unchiku* generation module, using a story as its input, compares words in the story with information in the knowledge base, then generates an *unchiku* based on what it finds. Figure 2 is an example of knowledge retrieval for *unchiku* generation from *Musume Dōjōji*. Figure 3 is an input example showing how the module inserts the retrieved knowledge from Figure 2 into the *Musume Dōjōji* story in Figure 3. We made two changes to the system presented in Section 2.1: one, the knowledge generation module now adds a symbol to the output sentence to clearly distinguish between events and knowledge, and two, it explicitly states that an animation will be inserted immediately after the pronouncement.

The following is the flow of *unchiku* generation:

1. The module receives a story from the user or story generator.
2. The module searches the knowledge for elements and keywords from the story to find relevant knowledge.
3. The system generates text presenting the retrieved knowledge.
4. The system inserts the generated knowledge into the story.
5. The system generates knowledge related to the generated knowledge.
6. The system inserts the generated knowledge into the story.
7. The system specifies the animation associated with the generated knowledge.

The knowledge retrieval and generation are repeated until there is no more relevant knowledge available. After that, the inserted animation will be related to the last generated knowledge.

The animation module receives the story from the *unchiku* generation module and generates a code for animation appropriate to the story. The module expresses the generated code and story as an animation as described in Section 2.

In the text representation, events are shown in black and insights in red. An animation is inserted at the end of the inscription.

4. Discussion

Figure 4 depicts the *unchiku* with moving illustration generated based on Figure 3. The *unchiku* is related to *Musume Dōjōji*. The content is as follows:

> The story about *Dōjōji* is *Kyōganoko Musume Dōjōji*. *Kyōganoko musume dōjōji.* It is a representative work of Shosagoto. It is a work played by representative *onnagata* such as Nakamura Utaemon VI and Bandō Tamasaburō V from the Showa period. An old legend leads to the kabuki dance *Kyōganoko Musume Dōjōji* through reading materials such as Buddhist narratives, Noh, and picture scrolls. For example, the scene of *Ranbyōshi* is as follows: *hito: Shirabyōshi Hanako*; *kokoro:* daughter; *huri: Shirabyōshi no mai*; lyrics: *Noh, gakki: hyousiita*; scene: cherry tree *(Dōjōji)*; *ishō:* red; *kodougu*: *Chūkei, nikutai no*; *point:* The legs. Figure 5 represents *Ranbyōshi.*

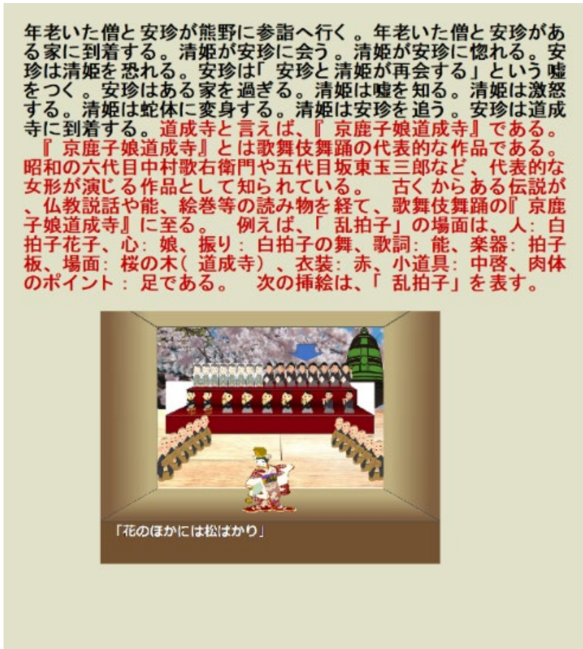


Fig. 5. *Unchiku* generation with moving illustration

In the future, we have two challenges for the system. First, we need to expand the Yunnan information we have in the knowledge base from our analysis of *Kyōganoko*

Musume Dōjōji. [11,12,13,14] Second, we need to expand the animations. We have created animations for Dōjōji,[13,14] and our goal is to use those animations.

5. Conclusion

We have developed an advanced explanation-generation system with two functions using *Musume Dōjōji* as our test model. First, the system inserts an *unchiku* explanation into a story. Second, it generates an animation to explain the *unchiku*. These two functions combine to deliver a more fully fleshed out explanation of the narrative.

Acknowledgement

This study was supported by JSPS KAKENHI, 18K18509.

References

1. T. Ogata, *Toward an Integrated Approach to Narrative Generation: Emerging Research and Opportunities*, IGI Global, PA, USA, 2020.
2. T. Ogata, *Internal and External Narrative Generation Based on Post-Narratology: Emerging Research and Opportunities*, IGI Global, PA, USA, 2020.
3. J. Ono, K. Fukuda, M. Kawai, and T. Ogata., Considering unchiku generation mechanism through system implementation, *Proc. of the 37th Annual Meeting of Japanese Cognitive Science Society*, Japan, Tokyo, 2020, pp. 776–786.
4. J. Ono and T. Ogata, Unchiku Generation Using a Narrative Explanation Mechanism. In E. Ishita, N. L. S Pang & L. Zhou (Eds.) *Digital Libraries at Times of Massive Societal Transition* (LNCS 12504), pp. 240–247. Switzerland: Springer Nature.
5. G. Genette, *Narrative discourse: An essay in method* (J. E. Lewin, Trans.), Cornell University Press, NY, 1972.
6. T. Akimoto, and T. Ogata, Towards a discourse mechanism in narrative generation system: Proposal of a System Introducing Narrative Discourse Theory and Reception Theory, *Cognitive Studies*, 20(4), 2013, pp. 396–420.
7. T. Ogata, Kabuki ni mukete (1): Shiisei to henshūsei no monogatari kara tajūsei to jitsuzon no monogatari he [Toward kabuki (1): From narratives of arbitrariness and editing to narratives of multiplicity and real existence], in *Jōhō monogatariron: Jinkōchinō, ninchi, shakai katei to monogatari seisei [Informational narratology: Artificial intelligence/cognition/social process and narrative generation]*, eds. T. Ogata, Y. Kawamura and A. Kanai, Hakutō Shobō, Tokyo, Japan, 2018, pp. 187–208.
8. T., Ogata, Kabuki ni mukete (2): Tajū monogatari kōzō no shosō [Toward kabuki (2): Aspects of multiple narrative structures]. in *Jōhō monogatariron: Jinkōchinō, ninchi, shakai katei to monogatari seisei [Informational narratology: Artificial intelligence/cognition/social process and narrative generation]*, eds. T. Ogata, Y. Kawamura and A. Kanai, Hakutō Shobō, Tokyo, Japan, 2018, pp. 209–244.
9. T. Ogata, Kabuki as Multiple Narrative Structures and Narrative Generation, in *Post-Narratology Through Computational and Cognitive Approaches*, eds. T. Ogata and T. Akimoto, IGI Global, PA, USA, 2019, pp. 192-275.
10. T. Ogata, Kabuki as a synthetic narrative: Synthesis and expansion. in Internal and external narrative generation based on post-narratology: Emerging research and opportunities, T. Ogata, IGI Global, PA, USA, 2019, pp. 315-403.
11. M. Kawai and T. Ogata, An analysis of Kyōganoko Musume Dōjōji for the generation of the staging structures of Kabuki from narrative plots, *SIG-LSE Proc.*, 63, 2019, pp. 25–61.
12. M. Kawai and T. Ogata, Analyzing the stage performance structure of Kabuki-dance "Kyōganoko Musume Dōjōji", *Proc. of the 34th Annual Conference of the Japanese Society for Artificial Intelligence*, 2020, 3D1-OS-22a-04.
13. M. Kawai, J. Ono and T Ogata, The Stage Structure of a Kabuki-dance: Survey, Analysis, and Representation of *Kyoganoko Musume Dojoji*, *Proc. of the 37th Japanese Cognitive Science Society*, 2020, pp. 776–785.
14. M. Kawai, J. Ono and T. Ogata, Dual Story Generation Based on Love and Extreme Emotions, *Proceedings of 2020 The 3rd International Conference on Computational Intelligence and Intelligent Systems*, 2020, pp. 9–10.
15. T. Akimoto, S. Imabuchi, J. Endo, J. Ono, Y. Kurisawa, M. Kamada and T. Ogata, Development of a generation / expression system of narratives in the style of a folk tale, KOSERUBE version 1, *Transactions of the Japanese Society for Artificial Intelligence*, 28(5), 2013, pp. 442–456.
16. P. Gervás, Propp' s Morphology of the Folk Tale as a Grammar for Generation, *Workshop on Computational Models of Narrative*, 2013, pp. 106–122.
17. K. L. Koay, D. S. Syrdal, K. Dautenhahn, and M. L. Walters, A narrative approach to human-robot interaction prototyping for companion robots, *Journal of Behavioral Robotics,* 11(1), 2020, pp. 66–85.
18. T. Watanabe, *Musume Dōjōji*, Shinshindo, Tokyo,1986.
19. Y. Kineie, *Syamisenbunkahu nagauta Kyōganoko Musume Dōjōji*, Hougakusya, Tokyo, 1952.
20. Kyōganoko Musume Dōjōji, *Bandō Tamasaburō butō shū Vol.1* (DVD), Syotsiku Home Video, Tokyo, 2003.

Implementing Story Units of Japanese Folktales with Conceptual Dictionaries

Takuya Ito
Graduate School of Software and Information Science, Iwate Prefectural University, Sugo 152-52, Takizawa, Iwate, Japan

Jumpei Ono
Faculty of Software and Information Technology, Aomori University, Seishincho 2-10-1, Edogawaku, Tokyo, Japan

Takashi Ogata
Faculty of Software and Information Science, Iwate Prefectural University, Sugo 152-52, Takizawa, Iwate, Japan
E-mail: g231r003@s.iwate-pu.ac.jp, j.ono@aomori-u.ac.jp, t-ogata@iwate-pu.ac.jp

Abstract

This paper reports the development of story units that are knowledge representing story structure for story generation. A story structure in our narrative generation study is formed by a tree that consists of events connected based on the relationships, such as the causal and temporal relations of events. We have been implemented story units based on the "Type of Folktale" in the *Nihon Mukashi-banashi Taisei* (*The Concluded Compilation of Japanese Folktales*). The current main theme is to use our conceptual dictionaries, including verb and noun conceptual dictionaries, to combine the current version of story units to our "integrated narrative generation system," which is a general framework of our narrative generation study. In this paper, we explain the above mechanism.

Keywords: List four to six keywords which characterize the article.

1. Introduction

In a study of folk tales and folktales from around the world, Propp collected a genre of Russian folk tales called magical folktales, and analyzed the tales structurally in terms of commonalities among the collected stories[1]. Propp called the structural commonalities "functions," which are related to the classification of characters, and in this respect are like the main action of the main characters in the definition of motifs.

This research is underpinned by narratology. We analyze existing stories and propose story units for generating a story using the results of the analysis in the Integrated Narrative Generation System (INGS), which is being developed by the authors[2,3].

Story units are in the prototype stage, and the verbs representing events in the Story Unit were created according to the descriptions of the folktales to be analyzed, and do not correspond to the descriptions in the verb concept dictionary that INGS refers to when generating stories.

In this paper, we aim to map the verbs in the story unit to the verb concept lexicon of INGS in order to generate a story using the story unit. In this paper, we aim to map the verbs in story unit to the verb concept dictionary of INGS. Specifically, we examine the descriptions of the verbs in story unit, and if the verbs in story unit can be mapped to the existing verb concepts in the verb concept dictionary, we register the verbs in story unit as the verb concepts of INGS. This correspondence enables us to use the rich sentence generation mechanism and about 110,000 noun concepts in INGS, and to generate a variety of stories from a story unit.

2. Motif and Form in the study of folktales

A motif is a unit that contains one major action of a major character in the composition of a story, as well as actions that directly correspond to that action [3].

In addition to the research on motifs mentioned in section 1, Aarne and Thompson collected folk tales mainly from various parts of Europe and produced the International Folk Tale Abstracts [4,5] (compiled by Aarne and expanded and revised by Thompson, so it is called the Aarne-Thompson type index (AT index)).

To analyze Aarne-Thompson's International Folktale Abstracts, it was noted that there were differences in the characters and actions of the different categories of folktales in the International Folktale Abstracts.

There are several efforts by Seki. The folktales used in this study are the efforts of Keigo Seki. The details are described below.

3. Seki's Study of Folktales

In *Nihon Mukashi-banashi Shusei* (*Collection of Japanese Folktales*), Seki classified about 8,700 folktales collected in various parts of Japan. Later, Seki and his colleagues expanded the number of tales to about 35,000 and updated the classification in *Nihon Mukashi-bahashi Taisei* (*The Complete Japanese Folktales*)[7-18]. Seki summarized the contents of 827 "type of folktale," which were the result of the classification. The 827 types of folktales based on the AT index[5,6]. The types of folktale is a classification of Japanese folktales as defined by Seki and a summary of their contents.

[Seki, 78a-e, 79a-d, 80a] has collected and classified about 35,000 folktales from various parts of Japan in *Nihon Mukashi-banashi Taisei*. There are 12 volumes of *Nihon mukashi-banashi Taisei* in total, and volumes 1 to 10 [7-16] contain the collected folktale themselves, volume 11 [17] contains materials on the type of folktale, an index, and a bibliography, and volume 12 contains contributed articles on folktales [18].

Each type has a summary of the contents of one or more folktales corresponding to the type. These types of folktales are systematized, with the largest categories being "animal folktales," "authentic folktales," and "funny stories," which are based on the AT index. In addition, these three classifications have "divisions" that group folktales according to their motifs, with 11 divisions for "animal folktales," 16 for "authentic folktales," and 12 for "funny stories" (a total of 39 divisions). Under each category, 827 types of folktales are assigned. The three major categories are further divided into 39 divisions. 827 types of folktale belong to one of the divisions, and each division contains a minimum of 2 and a maximum of 66 types of folktale.

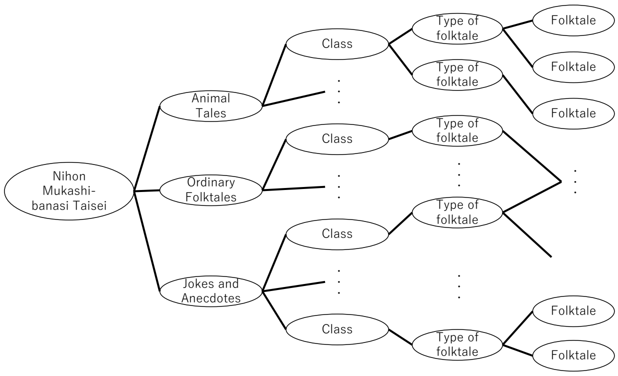


Fig. 1. The hierarchy of folktales in the "Nihon Mukashi-banashi Taisei"

Figure 2 shows an example of a type of folktale. type of folktale has four items.

- Serial number: The serial number for type of folktale (e.g., 一四二 (One hundred forty-two))
- Title: Motifs representing the characteristics of folktales (e.g., 蛇息子 (The snake son))
- AT Classification Number: These are the numbers corresponding to the motifs in the AT index. Some types of folktales do not have corresponding numbers. (e.g., AT 四三三 B)
- Content: The structure of a type of folktale. It consists of one or more sentences. The sentences are numbered roughly by scene. It may also contain branching patterns of events and patterns of elements. For example, the structure is as shown in Figure 3.

一四二 A 蛇息子 (AT 四三三 B)
1、子どものない夫婦が(a)神に祈願して蛇の子を生む。(b)卵を拾い、それから蛇の子が生まれる。2、あまりに大きくなりすぎて村人に恐れられるので山に捨てる。3、この蛇の援助によって(a)雨を降らせる。(b)田に雨がかかるようにする。(c)蛇が禍をなして爺婆に仇を討たれる、4、そのために爺婆は富を得る。

Fig. 2. An example of a type of a folktale

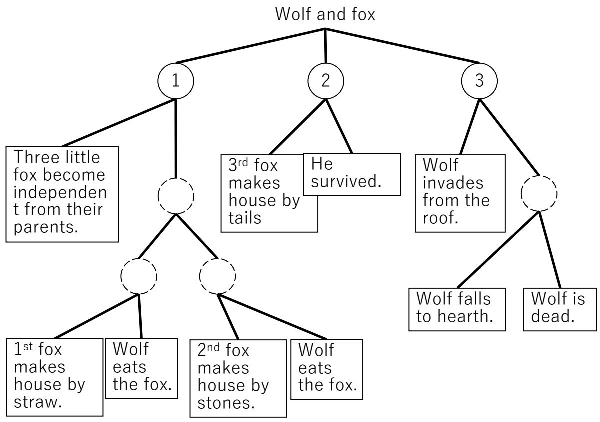


Fig. 3. An example of structure of a type of a folktale

The type of folktale is a structure that integrates the actual collected folktales. Figure 4 shows an example of collected folktales.

「昔、三匹の子狐がおった。一番兄狐が、もう一匹前になったからといって、親狐に藁から家をつくってもらった。ところが、ある日、・・・」[Once upon a time, There are three little foxes. An eldest brother fox is presented a house that makes by straw from their parent, because the eldest brother fox becomes an adult. But, one day, …]

Fig. 4. An example of motif variation

Each type of folktale has a serial number called a taisei number and a title that represents the most characteristic motif of the content (the bold Chinese numerals in the figure are taisei numbers).

4. First Step for Making Story Units

4.1 Story Units

A story unit is the knowledge that makes up a story. There are many kinds of story units. For example, there is a story unit that shows the structure of a large story. On the other hand, there is a story unit that shows the local structure of a story. We used "Type of folk tale" to create story units [19]. The created story units can be used to create folktale-like story. The story units can be used to create folktale-like story, and by combining the story units, it is possible to create a story that is different from the original story.

Section 4.2 and 4.3 show the procedure for creating story units.

4.2 The Structure of Type of Folktales

We first reconstructed the folktale type into a form suitable for the program. We split the content of the type of folktale so that an item contains a verb. Table 1 shows a form of splitting.

- Serial number: Referring Figure.2 in the section3.
- Title: Referring Figure.2 in the section3.
- Content number: The specific content of the folktale type is summarized by scene, with the events contained therein.
- Sentence number: Indicates a sentence break in type of folktale.
- Event: Events segmented by verb units

4.3 Making Story Units

To make a story unit, we create a case structure for each of the “event” descriptions in Table 1.

This story unit consists of three scenes, each of which is represented by A, B, and C. The events in each scene are numbered, and there are six events in this story unit. Each event is composed of a verb and its case information based on the way events are described in INGS. The descriptions of verbs and case elements in the Story Unit are based on the descriptions in type of folktale, and do not correspond to the verb and noun concepts in INGS. However, in the case of verbs, a temporary number 1 is

Table 1. The example of a structure form of type of folktale

大成番号	モチーフ(昔話)	内容番号				文番号	事象(内容)
474	鬼の面					①	ある女(男)が山(化け物屋敷)で鬼の面をかぶっている。
						②	化け物がそれを見て逃げる。
						③	女は
					(a)	①	宝物をとって
							帰る、
							&または&
					(b)	①	化け物屋敷の主人になる。

written after the verb in order to correspond to the verb concept.

The branching structure of the Type of folktale is represented by a description starting with "or", as shown in Figures 5.4a and 5.4b. In addition, there are branching structures that affect the choice of subsequent branching structures. In order to preserve the pairing relationship, there are examples where several different story units were created from the same type of folktale. Therefore, there are 912 story units that we have created.

The Story Unit is created based on the description of the type of Folk. In the case structure of verbs involving a dialogue such as “say,” the content of the dialogue is inserted directly into the object case. In this case, the elements of the object case cannot be changed or expanded without changing the content of the description. Therefore, we consider a more flexible use of the story unit by creating a nested structure of events.

Among the case elements of the story units, there are 761 cases that can be nested. Table 2 shows the classification of nested structures.

Table 2. The type of nested structure in the story units

Category	Amount	Definition
An action	217	Contents consisting of a single action
Actions	50	Contents consisting of multiple actions
Avoid	1	Contents that attempt to avoid a certain action.
Prohibition	15	Contents that prohibit certain actions
Propose	10	Contents of a proposal for an action
Desire	33	Contents that want to an action
Speak	4	Contents that characters speak
Denial	29	Contents that negate an action.
Impossibility	3	The impossibility of an action.
Command	68	To command an action.
Rewording	3	The action is difficult to understand, so it is replaced by a different verb expression.
合計	433	

5. Combining Story Units and a Verb Conceptual Dictionary

We combine the story unit shown in section 4 with a verb concept dictionary.

5.1 Goal of Combining Story Units and a Verb Conceptual Dictionary

The combination of verb concept dictionaries provides two advantages to story units. (1). Through the sentence pattern of the verb concept, single sentence generation and sentence generation with conjunction in INGS will be available, and the following functions will be available--“Ordering words,” “Transforming inflected form,” “Transforming notation,” “Multi-sentence.” (2). Story units can use the system structure of the verb concept dictionary. For example, the story units can manipulate events using the systematic structure of the verb concept dictionary, and to use the co-occurrence information of the verb concepts associated with the verb concept dictionary.

5.2 Verb Conceptual Dictionary

A concept here is data that shows the meaning and classification of a word. A concept dictionary is a kind of ontology of words belonging to a certain part of speech. In this section, two types of concept dictionaries are explained. One is the verb concept dictionary, and the other is the noun concept dictionary. The verb concept dictionary is a dictionary that systematically registers verb concepts. The verb concept dictionary has about 10,000 verb concepts. A noun concept dictionary is a systematic dictionary of noun concepts. The Noun Concept Dictionary has about 110,000 noun concepts.

Figure 6 shows an example of a verb concept. A verb concept has a sentence pattern, a case frame, and a constraint. Constraint describes the constraints for

```
(motif0669 (鬼の面 (被る1 見る1 逃げる1 (or (取る1 帰る1) なる1)))
        (
         ($継起
          (A
           (1 (event 被る1 (agent (&sc 女)) (object (&sc 鬼の面)) (location (&sc 山))))
           (2 (event 見る1 (agent (&sc 化け物)) (counter-agent (&sc 女))))
           (3 (event 逃げる1 (agent (&sc 化け物))))
           (or
            (4a (event 取る1 (agent (&sc 女)) (object (&sc 宝物)))
                 (event 帰る1 (agent (&sc 女))))
            (4b (event なる1 (agent (&sc 女)) (to (&sc 化け物屋敷の主人))))
           ))))
```

Fig. 5. An example of a story unit

selecting a case frame from the noun concept dictionary, and is-a information for systematically classifying verbs.

5.3 The Method for Combining Story Units and a Verb Conceptual Dictionary

This paper describes the binding of verbs contained in story units to a verb concept dictionary. We organized the types of verbs contained in story units and implemented methods according to the classification.

5.4.1 Counting kinds of verbs in the story units

The 912 story units we created have 1107 different verbs. These verbs can be divided into those that can be easily combined with the verb concept dictionary and those that require some work to be done in order to be combined with the verb concept dictionary. First, we classified the verbs in the story unit in terms of the challenges in binding. Table 3 shows the breakdown of the classification, the number of units corresponding to the classification, and an example.

Table 3. The category of task in the story units

Task	Amount	Example
(A). Direct connect	615	食べる
(B). New registration	363	あさる
(C). Notation	121	あきらめる
(D). Passive form	13	追われる
(E). Negative form	5	承知しない
Total	1117	

5.4.2 (A). Direct connect

(A) is a verb that can be easily combined with the verb concept dictionary. For verbs that have the same notation, once the description of the Story Unit is used as is, the sentence generation mechanism of INGS is used to expand the sentence into a natural sentence, and if a natural sentence that matches the content of the Story Unit is expanded, it is used as is. In case of no case, we selected an appropriate verb concept number from the verb concept dictionary and changed the number of the Story Unit verb to that number.

5.4.3 (B). New registration

For the verbs in (B), we first checked whether there were any verb concepts with the same meaning but different notation in the verb concept dictionary, and if so, we processed them as the task in (C). If there was no verb with different notation, we registered it as a new verb concept in the verb concept dictionary.

For each new verb, we created a sentence-pattern and a case-frame based on the case structure of the target verb in the story unit. case-frame refers to the case structure of the verb in the story unit, and case-frame refers to the case of the target verb in the story unit. The sentence-pattern was created by specifying the particle corresponding to each case used in the case-frame, and combining the "case number" and the "particle corresponding to the case". In the tentative data, we did not give any constraints on the cases to be used, and all is-a to classify verb types were set to "physical action". In this example, the case registered in the original story unit is the agent case, which represents the subject, and the object case, which represents the object.

5.4.4 (C). Notation

This problem stems from the linguistic nature of Japanese. The Japanese language can express homophonic words with different notations. For example, “食べる (taberu)” which means “eat,” can be expressed as “たべる,” “タベル,” etc. In story units, “気づく (notice)” can be expressed as "notice," "notice," "notice," etc.

In the story unit, there is a verb “気づく(kiduku)” which means “notice.” However, the verb concept

```
(set '通行する 1
    '((name 通行する 1)
      (sentence-pattern "N1 が N2 を 通行する")
      (case-cons-set
        ((case-frame ((agent N1) (location N2)))
         (constraint ((人 -死人 -人間〈人称〉 -準人間)
                      (場所 -世界 -景 -宇宙 建造物 -屋根 -柱・梁〔はしら・はり〕 -壁 -家屋{部分〈要素{その他}〉}[アンテナ] -家屋{部分〈要素{その他}〉}[避雷針] -建具 -塀)
                   ))))
       (is-a (v 物理的移動))))
```

Fig. 6. An example of a verb concept

dictionary registers it as "気づく." Therefore, we changed the notation of the verb in the story unit side to that of the verb concept dictionary.

5.5.5 (D). Passive form and (E). Negative form

(D) and (E) are issues that are addressed by revising the structure of the story unit. (D) is dealt with by using the original verb concept of the verb and placing the person who is the target of the action in the counter-agent case. (E) is expressed by adding a not structure to the case structure of the verb.

6. Conclusion

In this paper, we combined story units and the verb conceptual dictionary. Therefore, the story units can use sentence generation in the INGS. And the system structure of the verb concept dictionary.

In the future works, we have two tasks. At first, we have to combine story units with the noun conceptual dictionary. By combining with the noun conceptual dictionary, story units can manipulate the case of story units based on the systematic structure of noun concepts. At second, we revise the new verb concepts. The newly registered verb concepts have been created according to the structure of the folktale. In particular, we have to revise the constraints of the verbal concepts in order to make them more generic.

References

1. Propp
2. T. Ogata, *Toward an Integrated Approach to Narrative Generation: Emerging Research and Opportunities*, IGI Global, PA, USA, 2020.
3. T. Ogata, *Internal and External Narrative Generation Based on Post-Narratology: Emerging Research and Opportunities*, IGI Global, PA, USA, 2020.
4. [小澤 97] 小澤俊夫：昔話入門，ぎょうせい，1997.
5. [Aarne 61]Aarne, A.: The types of the folktale: a classification and bibliography, "FF communications," 184, Helsinki: Academia Scientarum Fennica, 1961.
6. [Aarne 69]Aarne, A.: 関啓吾訳.『昔話の比較研究』. 東京：岩崎美術社, 1969. (Antti A.: Leitfaden der vergleichenden. Marchenforshung, 1913.)
7. [Seki 1979] Seki, K., Nomura, J., & Ōshima, H. (Eds.): Nihon mukashibanashi taisei: vol. 12 kenkyū hen [Complete collection of Japanese folktales: Vol. 12, researches]. Tokyo, Japan: Kadokawa Shoten, 1979.
8. [Seki 1980] Seki, K., Nomura, J., & Ōshima, H (Eds.): Nihon mukashibanashi taisei: Vol. 11 shiryō hen [Complete collection of Japanese folktales: Vol. 11, data]. Tokyo, Japan: Kadokawa Shoten. 1980.
9. [Seki 1979a] Seki, K.: Nihon mukashibanashi taisei vol.1: dobutsu monogatari [The Concluded Compilation of Folktales]. Tokyo, Japan: Kadokawa Shoten, 1979.
10. [Seki 1978a] Seki, K.: Nihon mukashibanashi taisei vol.2: honkaku mukashibanashi (1) [The Concluded Compilation of Folktales]. Tokyo, Japan: Kadokawa Shoten, 1978.
11. [Seki 1978b] Seki, K.: Nihon mukashibanashi taisei vol.3: honkaku mukashibanashi (2) [The Concluded Compilation of Folktales]. Tokyo, Japan: Kadokawa Shoten, 1978.
12. [Seki 1978c] Seki, K.: Nihon mukashibanashi taisei vol.4: honkaku mukashibanashi (3) [The Concluded Compilation of Folktales]. Tokyo, Japan: Kadokawa Shoten, 1978.
13. [Seki 1978d] Seki, K.: Nihon mukashibanashi taisei vol.5: honkaku mukashibanashi (4) [The Concluded Compilation of Folktales]. Tokyo, Japan: Kadokawa Shoten, 1978.
14. [Seki 1978e] Seki, K.: Nihon mukashibanashi taisei vol.6: honkaku mukashibanashi (5) [The Concluded Compilation of Folktales]. Tokyo, Japan: Kadokawa Shoten, 1978.
15. [Seki 1979b] Seki, K.: Nihon mukashibanashi taisei vol.7: honkaku mukashibanashi (6) [The Concluded Compilation of Folktales]. Tokyo, Japan: Kadokawa Shoten, 1979.
16. [Seki 1979c] Seki, K.: Nihon mukashibanashi taisei vol.8: warai banashi (1) [The Concluded Compilation of Folktales]. Tokyo, Japan: Kadokawa Shoten, 1979.
17. [Seki 1979d] Seki, K.: Nihon mukashibanashi taisei vol.9: warai banashi (2) [The Concluded Compilation of Folktales]. Tokyo, Japan: Kadokawa Shoten, 1979.
18. [Seki 1980] Seki, K.: Nihon mukashibanashi taisei vol.10: warai banashi (3) [The Concluded Compilation of Folktales]. Tokyo, Japan: Kadokawa Shoten, 1980.
19. [Ono, 2019] Ono, J., A., Ito, T., & Ogata, T.: Mukahi banashi no motif no program ka to motif kozo no hikaku. In Proceedings of the 61th Special Interest Group on Language Sense Processing Engineering, pp. 51-62, 2019.

Story creation and cognitive biases

Jun Nakamura
Faculty of Global Management, Chuo University, Higashinakano 742-1
*Hachioji-shi, Tokyo 192-0393, Japan**
E-mail: jyulis.77f@g.chuo-u.ac.jp
www. Global.chuo-u.ac.jp/English/

Abstract

A wide variety of stories can be created from the same material. The author asked several teams to create stories, focusing on cognitive biases. As expected, a variety of stories were created by each team, and cognitive biases were identified in the process of creating them. In other words, if it is assumed that cognitive bias is a driving force behind the creator's intentions, then it can be concluded that the creator must have a certain strong intention when creating a story.

Keywords: Story, cognitive bias, analogy, and creativity

1. Introduction

Studies of "creativity" are difficult to replicate. This research has sometimes been said to be inherently dubious. However, human prosperity is nothing but the accumulation of creativity derived not only from destructive innovation but also from gradual creative innovation that makes it sustainable.

The word "creativity" can be used to describe a wide range of products, depending on what one is creating. The product may be an object, a service, or a new technology. Whatever the case, to attract people, it is important to tell a "narrative story", and I believe that an attractive story makes a deep impression on people and inspires them to remember it.

I have been studying the process of designing and generating stories. In doing so, I have focused on how the characteristics of stories change depending on the presence or absence of external stimuli (Nakamura, 2020). This paper is focused, first, on the effects of availability heuristics, termed cognitive biases in this paper, such as the selective use of memory. I explore the process by which different teams given the same data set create different stories.

The second chapter touches on related research and experimental tools. Next, Chapter 3 describes stories' impression points and items for assessing cognitive biases. This is followed by Chapter 4, which introduces an experiment based on participants' self-assessment. The results of the experiment are then presented in Chapter 5. This is followed by a discussion aimed at analyzing and interpreting the results of the experiment, and the final chapter offers a conclusion.

2. Related research and experimental tool

As an area of research in support of activities to create stories, the aim was to develop games that enhance composition and creativity to reconstruct combinations of words and thereby create new concepts. The importance of the meaning and role of words in the design process, which is a creative activity, has been noted (Nakakoji and Yamamoto, 2010). The game developed is aimed at connecting words by analogy (Holyoak and Thagard, 1996) and discovering new expressions in various combinations. It is called the Analogy Game (Nakamura and Ohsawa, 2007).

This type of game falls into the category of "serious games" in the broadest sense of the word. Serious games are defined as "games used to solve problems in education and other areas of society" (Sawyer 2001). Examples would include training in firefighting activities, emergency and medical training, and language learning. Overseas, the use of serious games to address the issue of the educational use of games has been attracting attention (Fujimoto, 2006). Since this analogy game was aimed at fostering composition and creativity, it is of a type that players explore and structure by themselves (Quinn, Beattie, McNaught, and Wills, 1994).

The analogy game is introduced as an experimental tool in the following manner. At the beginning, players are given 20 word cards that have been pre-classified (panel A in Figure 2). The word cards are clustered into several categories that are given relevant concept names commonly used in daily life, such as "fruits", which would cover bananas, strawberries, and melons. In this paper, such concepts are regarded as ready-made concepts.

After the game begins, players attempt to dismantle the ready-made concepts and reconstruct new combinations by moving the word cards around (Figure 2 B). Words that are grouped together at this stage are highlighted in the same color. This is because the system recognizes groups by their color code. Players then generate new concept names for the reconstructed groups. If one word card is left out of these groups, the composition must be reviewed again. The game ends when all of the word cards are classified into one of the new groups (Figure 2 C). For example, suppose that services and medical care were categorized as "businesses", sushi and baseball as "Japan", and cars and barbecue as "family travel". Players could recombine these ready-made concepts by combining service, medical, and barbecue into "compassion" and by bundling sushi, baseball, and cars into "overseas expansion", thereby creating concepts that are completely different from the original concept.

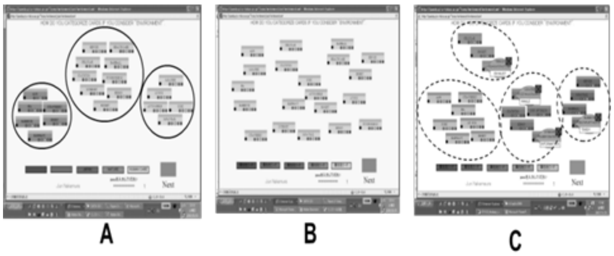


Fig. 1. Flow of the Analogy Game

3. Story impression points and assessment

The aim of the experimental tool mentioned in the previous chapter was to support creative activities (Hori, 2007). The present chapter explains impression points in story creation as a deliverable self-assessment through debriefing after the game.

3.1. *Story impression points*

It has been suggested that creativity can be evaluated from the viewpoints of originality and novelty as well as practicality and appropriateness (Paletz and Peng, 2008). Therefore, the author concluded that "Innovative" and "Feasible" are appropriate as impression points for created stories. In addition, three impression points, Usefulness, Scale expansion, and Barrier to rival, were added, with reference to the value proposition design proposed by Imazu† (2016), which excluded quantitative evaluation points. Therefore, the following five items were identified as the impression points for created stories in this paper: Innovation, Feasibility, Usefulness, Scale expansion, and Barrier to rival.

3.2. *Assessment*

Next, the assessment items for cognitive biases are introduced. In relation to cognitive biases, self-awareness is metacognition (Brown, 1978). This paper aimed to use metacognition to explore the source of story creation. For this purpose, as described in the first chapter, the participants were asked to self-assess relevant characteristics, considering the availability heuristics such as people's selective use of memory, as forms of cognitive bias. In this case, what are the types of cognitive bias? Bazerman and Moore identified 11 types of bias that can occur during decision making (Baserman and Moore, 2013). In this paper, four of these were selected for consideration in creative activities, excluding quantitative items.

- Ease of Recall: cognitive bias that occurs toward easily recalled material based on the clarity of memory
- Retrievability: cognitive bias toward an object to be remembered that is structurally easy to retrieve
- Confirmation trap: cognitive bias based on the assumption that one's own hypothesis is correct
- Anchoring effect: cognitive bias toward available information as a starting point and tendency not to move far beyond that anchor

Based on the above, the assessment items for cognitive bias were organized in this experiment as follows.

Assessment item 1 (memory versus hypothesis)
1-X = Recalling someone's *memories of past events* affected your ideas
1-Y = Not your memories but your team's *own tentative hypothesis* affected your ideas

Assessment item 2 (stuck to the cards versus took off from cards)
2-X = You were affected by the meaning of a given set of cards and continued to *stick with* that until the end
2-Y = You were influenced by a given set of cards, but the idea *took off from* that

Assessment item 3 (not shared versus shared)
3-X = During the team discussion, a unique idea arose that was *not shared* by other individuals
3-Y = The ideas were sufficient using only the information *shared* among team members

Assessment item 1-X was combined with Ease of recall and Retrievability because both were types of cognitive bias related to ones' memory of the past. On the other hand, assessment item 1-Y asked the participants whether they felt a cognitive bias suggesting that they believe their hypothesis is correct as a confirmation trap or as a memory of the past.

Assessment item 2-X asked whether the participants recognized the anchoring effect, which is a type of cognitive bias, and assessment item 2-Y asked whether they recognized the expansion of their ideas without their being anchored to a given card.

Assessment item 3 was not a straightforward question of cognitive bias, but it asked for "bounded awareness" in conducting teamwork. In this experiment, the participants spent more than 2 hours in teamwork, in effect, and the amount of information in the conversation was considerable. In sorting this information during the discussion, the participants had to process unconsciously. In the post-experimental review, assessment item 3-X asked participants whether they felt that the information that they considered useful (i.e., the cognitive bias of the confirmation trap was also present here) was ignored or overlooked. In contrast to 3-X, assessment item 3-Y asked whether the respondents felt that the information that they found useful was sufficiently shared and agreed upon within the team.

4. Experiment

In this paper, we conducted an experiment using the analogy game introduced in Chapter 2 in the following manner:
Experiment date: Friday, November 27, 2020
Participants: Twenty-one second-year students of the Faculty of Global Management, Chuo University
Experimental method: Twenty-one people were divided into six teams and presented with a set of word cards on the screen as follows. Capital letters in the following indicate the names of the clusters, and lower-case letters indicate individual words on the cards.

- TRIP ADVISOR: discount ticket, foreign tour, backpacker, guide, word of mouth
- FACEBOOK: search, friends, like!, share, network
- IKEA LIFE: living room, do it yourself, easy to store, Europe, wardrobe
- STARBACKS: yen400, third place, extra job, coffee, steamer

The above set of word cards is the same as that used in our previous experiment (Nakamura, 2020). The difference between the previous experiment and the present one is that this was an experiment not with one team but with six teams. Participants were instructed to devise a story connecting several newly reconstructed clusters (a set of word cards) using the experimental tool. The story was to be expressed in PowerPoint using a copy of the screen created with the experimental tool. The story was then presented to the participants. After the presentation of the story, participants were asked to do the following two tasks:

- Give a score to stories created by teams other than your own, taking into account the impression points mentioned in Chapter 3. A total of 20 points was the upper limit of the score. However, the same number of evaluation points should not be given to more than one team.
- Assess what cognitive biases were present in response to your own team using the assessment items described in Chapter 3. Participants were instructed to answer questions X and Y for each of the assessment items (1, 2, and 3), for a total of 10 points, and determine the ratio of X and Y that they think applies to them.

5. Results

Even though the word cards provided in the experiment were the same for all teams, different teams created different stories. The impression point scores for the six teams are shown in Figure 2, with team 4 receiving the highest rating.

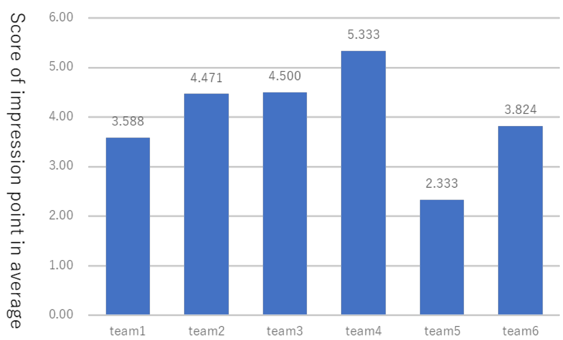


Fig. 2. Impression points rating by team

The team 4 story that received the best ratings was the following:

Story theme. *Four things that the students on Working Holiday need might be Cheap clothes, Information, Relax and Tour.*

Story. Since they are working in foreign countries to earn money, study a language, and learn the culture at the same time, it is better to save clothing money and travel around if they have time. To save money on their clothing, it would be better for them to purchase or borrow seasonal clothes at a sales market or garage sale. Also, since it will be difficult to get information in foreign countries, where they do not have friends or family, they might need to use an online community to get various kinds of information during their working holidays. Finally, it is important to relax after work every day.

An interesting slide from the team 4 presentation is shown in Figure 3.

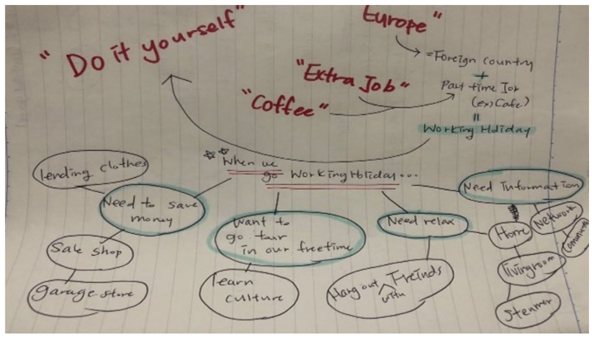


Fig. 3. A slide representing the thinking process

This is a picture of a note on the thinking process underlying creating a story. The note was written down and projected onto one of the presentation slides. The explanation given for this slide was as follows:

(1) The first four words that stood out to me were, “Do it yourself”, “coffee”, “extra-job”, and “Europe”. I decided to connect the four.
(2) Think of it this way, it's a working holiday! That is said, a part-time job in a foreign country is indeed a working holiday!
(3) On a working holiday, you have to do it all by yourself.
(4) As a mini-story derived from the above, on a working holiday we need to save money, we might want to go on a tour in our free time, we might need to relax, and we might need information.

The results of the self-assessment from all of participants are shown in Table 1.

Table 1. Self-assessment of cognitive biases.

	Team1	Team2	Team3	Team4	Team5	Team6	AVE	STD-EVP
1-X	3.00	4.25	**7.00**	3.00	2.67	0.50	3.40	1.96
1-Y	7.00	5.75	3.00	**7.00**	7.33	9.50	6.60	1.96
2-X	3.00	4.25	5.33	2.67	5.00	5.50	4.29	1.11
2-Y	7.00	5.75	4.67	**7.33**	5.00	4.50	5.71	1.11
3-X	0.75	4.00	2.67	2.00	4.67	0.25	2.39	1.60
3-Y	9.25	6.00	**7.33**	**8.00**	**5.33**	9.75	7.61	1.60

6. Discussion

Focusing on teams 4 and 3, which had the highest ratings, what kind of cognitive biases did the team members feel? In team 4, the most remarkable assessment item that the team member reported was item 2-Y, which was unique compared to the other teams. This indicates that they did not stick to the given key words until the end, and they perceived themselves to have made a leap in the meaning of the words. This may be due to the so-called *polysemy* effect, in which a given key word has multiple meanings (Costello and Keane, 2000). It could be said that blending a meaning that is different from the original meaning (Fauconnier, 1994) produced a creative story. In fact, if we look at the thought process of team 4, as depicted in Figure 3, the approach to the mini-story that connects the first four key words (Europe, Extra-Job, Coffee, and Do it yourself) was noteworthy. From this mini-story, the team expanded the idea and connected the remaining key words to form the whole story.

The second most highly rated team was team 3, which, like team 4, showed characteristic cognitive biases that were different from the other teams. Team 3 had a higher rating on item 1-X than the other teams, i.e., they felt that their memories were more effective in generating stories than were their own free ideas and hypotheses. The possible reasons for this are discussed below.

In the case of team 3, it might be presumed that memory was given priority because the consensus was that teamwork must take precedence over mutually agreed-upon ideas, resulting in the hypotheses of various participants cancelling each other out. However, team 4, the most highly rated team, gave the opposite answer. That is, team 4 responded with a free hypothesis that was not anchored in their own memory. This contrast is quite interesting. Given that teams 4 and 3 did not respond uniquely to the question, it is difficult to figure out the story-generating algorithm. However, at least from the above results, we can say that the cognitive biases of assessment items 1 and 2, whether X or Y, are characteristically different to those of the other teams (i.e., representing highest or lowest awareness).

Now, we discuss assessment item 3. For both team 3 and team 4, assessment item Y was ranked higher than item X. This implies that both teams seemed to be in a good environment for the members to share their opinions with each other. On the other hand, for team 5, which had the lowest score, X was higher than Y, indicating that individual opinions were not reflected. That means that no chemical reaction yielding new ideas occurred in the mix of different opinions from the team members.

7. Conclusions

Cognitive biases were evident throughout the experiment in the process of creating stories. In other words, if we assume that cognitive biases are the driving force behind the creator's intentions, then it is essential that the creator have a certain strong intention to yield a memorable story.

The experiment in this paper, however, is just a case study of a limited number of people, and there are still many works to be addressed before the findings can be generalized. In the future, the author intends to focus more on the process of creating stories.

Acknowledgement

This study was carried out with the cooperation of undergraduate students of Chuo University, Japan. Also, the Analogy Game was upgraded and developed further with the support of Mr. Tano, President of Expert Gig Co., Ltd. This research was supported by Kakenhi (JSPS KAKENHI Grants Numbers 17K03872 and 19K03062).

The English in this document has been checked by at least two professional editors, both native speakers of English. For a certificate, please see: http://www. textcheck. com/certificate/rTtVeD.

References

1. J. Nakamura, Effect of external activate factors serving as clue: creating new products or services ideas with storytelling, T. Ogata and J. Ono (eds.), *Bridging the Gap Between AI, Cognitive Science, and Narratology With Narrative Generation*, IGI Global, Hershey, PA, 2020.

2. K. Nakakoji and Y. Yamamoto, Wording as external representations to interact with: its role and nature throughout a design process, *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, **17**(3), 2010, pp. 474-490.
3. K.J. Holyoak and P. Thagard, *Mental Leaps*, MIT Press, MA, 1996.
4. J. Nakamura and Y. Ohsawa, The visualization of concept creation throughout the analogy game, *Proceedings of the 2nd Oukan Conference, Transdisciplinary Federation of Science and Technology*, 2007, pp. 157-160.
5. B. Sawyer, Serious Games: Improving Public Policy through Game based Learning and Simulation, in *Foresight and Governance Project, Woodrow Wilson International Center for Scholars*, *Publication 2002-1*, retrieved on Dec.5, 2020, at https://www.scribd.com/document/38259791/Serious-Games-Improving-Public-Policy-through-Gamebased-Learning-and-Simulation.
6. T. Fujimoto (2006) Kaigai niokeru serious game no kanousei (in Japanese), in Proceedings of the Conference on Simulation & Gaming, Autumn, 2006.
7. C. N. Quinn, K. Beattie, C. McNaught, S. Wills, Designing educational computer games, *Proceedings of Interactive Multimedia in University Education: Designing for Change in Teaching and Leadership IFIP Transactions*, 1994, pp. 45-47.
8. K. Hori, *Creativity Support, Theory and Application of AI Approach toward Creativity*, Ohmusha, Tokyo, 2007.
9. S. B. Paletz and K. Peng, Implicit theories of creativity across cultures, *Journal of Cross-Cultural Psychology*, **39**(3), 2008, pp. 286-302.
10. M. Imazu, *Hashirinagara kangaeru sinkijigyou no kyoukasho* (*in Japanese*), Kanki Pub., Tokyo, 2016.
11. A. L. Brown, Knowing when, where, and how to remember: a problem of metacognition, in R. Glaser, (Ed.), *Advances in Instructional Psychology*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1978.
12. M. H. Bazerman and D.A. Moore, *Judgement in Managerial Decision Making*, John Wiley & Sons, Inc., MA, 2013.
13. F. J. Costello and M. T. Keane, Efficient creativity: constraint-guided conceptual combination, *Cognitive Science*, **24**(2), 2000, pp. 299–349.
14. G. Fauconnier, *Mental Spaces*, Cambridge University Press, Cambridge, UK, 2008.

Time in an Aesthetic Experience of a cup of Sake

Hiroki Fukushima
Kyushu Women's University, 1-1, Jiyugaoka, Yahata-hnishi Ku
Kitakyushu City, Fukuoka, Japan

Abstract

There are two broad genres of artworks depending on the nature of their appreciation: arts of space and arts of time. In the aesthetic appreciation of taste, which is the focus of the experience of a cup of Sake? Is there the dimension of time in the experience of taste, and if so, what is it? Because taste has never been considered a formal subject of aesthetics, there is little discussion of describing and appreciating the aesthetic experience. This study deals with the concept of time in the appreciation of Sake, with some reference to the concept in narrative studies. As with wine, the time associated with Sake is multifaceted, and each is intrinsic to the aesthetic experience. Consider some examples: Time from the Sake is brewed to the time it is opened on the table (so-called "vintage"; for some years). The time from opening the bottle to pour into the glass (for minutes to tens of minutes). The time between pouring into the glass and entering the mouth (for a few minutes). Time to swallow in the mouth (a few seconds). The flavor felt after swallowing ("after flavor"; tens of seconds). This study focuses on the period from the Sake is placed in the mouth to the time it is swallowed. The purpose of this study is to reveal how time is narrated (explicitly or metaphorically) in the tens of seconds of a sake tasting.

Keywords: Sake, Taste, Temporal Arts, Plastic Arts.

1. Introduction

Taste has been largely neglected in traditional aesthetics. Since Sibley discussed the aesthetics of taste and smell in 1959, several philosophers have tried to theorize the aesthetics of taste and smell and have introduced artistic activities of smell, but not sufficient.

The few researchers who deal with the aesthetics of taste and smell must devote the early part of their works to a masochistic discussion of how taste and smell have been neglected in aesthetic research.

In this study, we temporarily put aside the question that whether taste and smell are the subjects of aesthetics. Then, if the art works of taste and smell are established, what kind of character they can have is discussed.

2. The Temporal Art and the Spatial Art

Consider the position of taste art. There are many genres in art such as painting, music, architecture, drama, literature, and so on. There is a long history of discussions on how to classify them systematically. Max Schasler (1819 -1903), a German philosopher, sets two opposing principles for the artworks: *das Simultane* and *Sukzessive*. The former corresponds to the plastic arts and the latter to temporal art. Architecture, sculpture, and painting are classified into plastic art, and music, mimic, and poetry are classified into temporal art. This classification is based on the conflict between material elements and ideological connotations.

The opposition between temporal art and spatial art was made clear in Lessing's (1729-1781) *Laocoon: An Essay on the Limits of Painting and Poetry*. In this debate, Lessing criticizes the tendency to equate poetry and painting, as typified by Horatius' proverb, *"ut pictura poesis*". The opposition between poetry and painting, he argues, is that painting and sculpture depict the decisive moment of a subject in a spatial extension, while literature depicts a story in a temporal extension. If we follow Lessing's argument, spatial art can be further divided into two categories: three-dimensional (e.g., architecture and sculpture) and two-dimensional (e.g., painting).

The classification of temporal art and spatial art may seem unassailable, but E. Souriau argued against it. Souriau criticizes the disconnection between temporal and spatial art. In his main work, *La Correspondance des arts* (1943), he compares music and painting, and points out the similarities between their various elements.

3. Taste and The Genre of the art

Expressions of taste can be broadly classified into verbal and non-verbal expressions. Verbal expressions are, as typified by sommeliers, the descriptions of the characters and impressions of a taste experiences. Linguistic data can be collected from publications such as books and magazines, or through verbalization experiments and interviews.

There are few practical attempts of non-verbal expression of taste. In the narrower sense, non-verbal expression can be defined as a work of art in which a person draws a picture or plays music to describe the flavors of all along the time a wine enters the mouth to the time it is swallowed. Such artworks will express a single, subjective, individual experience of eating and drinking. On the other hand, in a broader sense, non-verbal expression can include labels, posters, and commercial images of luxury goods such as wine, *sake*, cigarettes, etc. These can be thought of as depictions of the character of the flavors of the product, but they do not represent a specific eating or drinking experience. The fact that a bear is depicted on a chocolate package does not mean that the chocolate tastes like bear meat.

3.1. *Multiple Time aspect in the Appreciation of Taste*

In this section, I will analyze how time is discussed in the representation of taste.

In the appreciation of taste, the most crucial time is the few seconds between when the food enters the mouth and when it is swallowed into the back of the throat. However, those few seconds are only a small part of the many aspects of the concept of time relating to the appreciation of taste.

As with wine, the time associated with Sake is multifaceted, and each is intrinsic to the aesthetic experience:

- The time between when the liquor is brewed and when it is opened on the table. (so-called "vintage" for some years).
- The time from opening the bottle to pour into the glass (for minutes to tens of minutes).
- The time between pouring into the glass and entering the mouth (for a few minutes).
- Time to swallow in the mouth (a few seconds).
- The flavor felt after swallowing ("after flavor"; tens of seconds).

Among these time aspects, this study focuses on the period from the Sake is placed in the mouth to the time it is swallowed.

3.2. *Describing the Time of the Taste*

Case 1: Dominio IV in Oregon

Figure 1. Riesling 2014, Dominio IV

Oregon winery Dominio IV's imagination series features taste drawings on wine labels. In this drawing, called "shape tasting," the impression of taste is represented by a two-dimensional picture along a time axis. The shapes and colors of the drawing synaesthetically represent the taste elements of the wine. The time axis and the number of seconds are at the bottom of the painting. The time goes from left to right, symbolically depicting the flavors as they appear, spread, move, unfold, change, and linger in the mouth.

At first glance, it appears as if the artist is freely drawing the flavor images with paint, but in this series, the vertical and horizontal axes of the canvas are defined first. The wine label is oriented horizontally, and the horizontal axis direction is a time axis with seconds. In the picture of Riesling shown in Figure 4, the leftmost point is 0 seconds, and the changes in taste over 60 seconds are drawn in 15-second increments. The vertical axis is shown as the extent of the palate breadth. The middle of the vertical axis represents the center of the oral cavity. The 1/4 width range from the center is depicted as the "narrow" area. The outer drawing area shows how wide the flavor spreads. Thus, in the example of fig. 1, a time variation can be seen as follows; the flavor first spreads widely in the mouth -> gradually converges narrowly from 15 seconds to 30 seconds -> leading to the final taste. In the imagination series, the drawings are described vertically symmetrically, with some exceptions.

In addition to the time axis and the spatial axis of the oral cavity, another feature is that what the shapes (i.e., circles, lines, and arrows in the drawings) represent are defined. Again, in figure 1, the example of Riesling 2014, we can see a note at the bottom of the figure. According to the notation, the red circles represent "fruit" flavors, and the yellow curve represents "fruit-based acidity." Then, the brownish arrows indicate "linear acidity."

In this way, the definitions of the drawing elements enable us to decode the drawings. The 2014 Riesling is expected to taste as follows; fruit-based acidity and sweet fruit flavors that fill the mouth as soon as you put it in -> and then gradually fade after about 15 seconds -> After 30 seconds, a linear acidity emerges -> a lingering acidity can be enjoyed from 45 to 60 seconds."

The expression of wine flavor is often given in a language-based manner, as typified by sommeliers. And such descriptions are often printed on the back labels of wines. In the Imagination Series of Dominio IV, captions are provided just below the illustrations. Additionally, verbal explarnations of the flavors are also provided. This series is an eloquent example of the possibility of the non-verbal description of the taste.

Case 2: Sake Art Label Project

Figure 2, Sake Art Label project

The attempt to present the image of taste in pictorial form has also been practiced in sake.

In the past, sake labels were generally designed by selecting a background from a template and placing the brush-written brand name. In recent years, however, with the development of distribution and the boom in local sake and *ginjo* sake, stylish and unique designs have been seen on sake labels. This shift represents a turning point. Up until now, it was enough to know the name of the brand in order to select the "usual" sake from the local brewery, such as *Tsukasa-Botan* or *Tosatsuru* in Kochi Prefecture. But now, with the development of transportation, consumers tend to seek unique local sake all over Japan.

Against such a background, it has become common for sake breweries to express their thoughts and the characters of their brand into their label design.

One of the notable efforts to express the image of sake flavor through labels is the "Sake Art Label Project," which was realized in 2014 through a collaboration between the Obata Shuzo Brewery (Sado City, Niigata Prefecture, the representative brand "Manotsuru") and the Isetan Mitsukoshi Group (Figure 6).

The project is an effort to break away from the former sake labels style, as a list of technical terms and express the image of flavor as an art label. Sake labels generally contain not only the name of the brand but also the variety of rice, such as "Yamadanishiki," the standard, such as

"Junmai" or "Ginjo," the method of production, such as polishing ratio or "Namagenshu," and numerical values, such as amino acid content or alcohol content. While this information is useful for predicting the flavor, it also requires a certain level of knowledge to decipher, which can put off beginners.

In the sake art label project, the opinion exchange meeting which invited cooking party, copywriter, and general consumers was carried out, and 8 kinds of sake were tasted blindly (concealed brand names). About 40 participants answered a questionnaire to answer images, colors, and objects associated with each brand according to their preferences. At the tasting site, Takeshi Hirashima, an illustrator, was present and drew illustrations based on the comments of the participants. Finally, the brewery, the buyer of Isetan Mitsukoshi, and the illustrator selected the two highly evaluated brands, and the illustrator designed the images of each flavor on the labels.

4. Conclusion

The characteristic of painting is that it has no temporality. Of course, some philosophers such as Souriau insist on the concept of time in paintings. And in the plastic arts, aesthetic experiences can occur chronologically. However, unlike music, the object itself does not change its form in time.

There are many efforts to express time on canvas. These efforts are not avant-garde. Manga uses a method of developing time through the combination of multiple panels of pictures. Or the illustrated handscrolls "Emaki-mono" (literally 'picture-scrolls'), traditionally produced in Japan and China, are pictorial artworks in which the story unfolds chronologically. The oldest picture scrolls in Japan date back to the Nara period (Einga-kyo, 7C-8C). From the 10th century onward, picture scrolls have been actively produced in Japan in various genres, including war stories, diaries, fairy tales, and comic stories. Anna Willmann points out that; Reading a handscroll can become an almost cinematic Anna Willmann points out that; Reading a handscroll can become an almost cinematic experience as the viewer scrolls through a narrative from right to left, rolling out one segment with his left hand as he re-rolls the right-hand portion.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 20K20127.

References

1. Lessing, G. E. (1853). *Laocoon: an Essay on the Limits of Painting and Poetry*. Longman, Brown, Green, and Longmans.
2. Souriau, E. (1949). Time in the Plastic Arts. *The Journal of Aesthetics and Art Criticism*, 7(4), 294-307. doi:10.2307/426722

Extension of Clinical/Psychological Approach Using Post-Narratology: Possibility of application on Artificial Intelligence and Robot

Kai Seino
Department of Social Rehabilitation of Research Institute, National Rehabilitation Center for Persons with Disabilities, 1, Namiki 4, Tokorozawa City, Saitama Prefecture, 359-8555, Japan

E-mail: seino-kai @rehab.go.jp
www.rehab.go.jp/english/index.html

Abstract

A purpose of this research is to examine the psychological approach from a narrative viewpoint and discuss expanding collaboration with AI and Robot. Methods are a literature review and theoretical examination. As a result, psychological narratology was confirmed to have expanded beyond the so-called narrative therapy. In addition, AI or robot research in the clinical fields have been identified. This will give suggestions to the extension of psychological narratology.

Keywords: Narratology, Computational Narratology, Counseling, Clinical Psychology, Disability, Psychiatry

1. Introduction

This research examine the potential of combining the narratology as a problem-solving method with techniques such as artificial intelligence (A.I.) or Robot, and the applicability of clinical areas. As a general term, “narrative” is understood as the “act of discussing and being discussed using language (Spoken Language and written language, etc)” (Yamada, 2007). “Narrative” can also be defined more narrowly as in the term “narratology,” which refers to the reproduction of real or fictional events and is always accompanied by time constraints (Prince, 2003).

Bruner (1986) suggested that the narrative provides individuals with a method for understanding and thinking about their everyday lives and interactions with others. In recent years, the narrative approach has been utilized in clinical domains such as medicine, nursing, psychology. The narrative may be thought of as an evidence base containing conventional, objective data.

In the clinical domain, the approach is key to any narrative, and is called the narrative approach. For example, representation is a narrative therapy in psychotherapy. However, the conceptualization of a narrative in recent years has not been limited to the narrative approach and is considered to have a wider extent and possibility. One key to understanding the wider possibility of a narrative is the term “narrative generative system,” which is advocated by Ogata,. The narrative generative system conceived by Ogata not only considers the structure or form of a story, but also incorporates the process by which the story is constituted and received. In addition, “computational narratology,” one of the underlying theories of the narrative generative system, is a concept resulting from the fusion of computational technology and story. Authors have divided narratology into literary narratology and psychological narratology from the viewpoint of a narrative generation system (Seino 2016). A literary

narratology considers the contents and form of a story, whereas psychological narratology is called a story treatment in the clinical domain and is a narrative approach.
Against the backdrop of Ogata's theory, Seino et al (2016, 2018) proposed a new technique for data mining designed to provide support for persons with disabilities. The technique provides a method for the analysis and generation of a story, through a combination of text-mining technology and the viewpoint of literary narratology. The results demonstrated that a theory could be established based on a narrative generative system. However, to investigate the practical implications of the system for the theory and method of a psychological narratology, a discussion of technology (e.g., computer, software, coding) is omitted.
Therefore, the psychological approach is presented in this research from the viewpoint of a narrative, and aims at clarifying the possibility of a psychological narratology. disabilities in order to investigate the possibility of using AI and robot for narrative research into the clinical field. Method in this research is literature review. Medical treatment and support for any illness that exists in the clinical domain is defined as a psychological approach. In this research, the possibility of extending the psychological approach through collaboration with AI or robot has been examined. Specifically, we review the research and practice of psychology, psychiatry, and a proximity domain in recent years and their use of narrative. Our discussion is broadly focused on the approach and practice relevant not only to narrative therapy but also to a narrative, or to narratology. We will examine the potential of AI and robot using psychological narratology from applications of AI and robotics in research on actual diseases and support for people with disabilities.

2. A Definition and Theory of a Narrative

2.1. *Definition of "Narrative"*

A narrative has been defined in multiple ways, and there are also many theoretical positions on the topic (Riessman, 1991). For example, one definition of narrative is that it is an explanation of the experience of a continuous story or people. A narrative may also be communicated using several methods (Clandinin, 2007). Narratology, which is the theory of a narrative, is defined in the narrowest sense in the literary theory of "literary narratology." Moreover, a narrative may be used outside of the field of literature, as in a clinical domain such as psychology, medicine, nursing, and social work that provides services to support persons. The combination of narratology and the use of a narrative in a clinical domain is called "psychological narratology.

2.2. *Literary narratology and "computational narratology" as an application*

The narrowest definition of narratology is regarded as a literary theory, which is the study of the nature, form, and function of narrative (Prince, 2003). Ogata (2010) organized narratology into five categories that establish the theoretical underpinnings of narratology: (1) "The Poetics" of Aristotle; (2) structuralism, which developed based on the ideas of Saussure; (3) the literary art movement of Russian Formalism ; (4) British and American literary theory; and (5) the literary theory of structuralism.
Narrative comprises "story" and "discourse." According to narratology (Prince, 2003), the story is the content of the narrative while the discourse concerns the expressive side. Narratology employs an analysis of both story and discourse. Propp (1968) described the fundamental components of any Russian folktale. In particular, he defined the narratological notions of "function" and characters' "roles." "Function" is a character's act defined from the viewpoint of its significance in the story. Furthermore, Genette (1980) divided discourse into three categories: (1) "tense," which is the temporal relation between discourse and story; (2) "mood," which is the reproduction of the story through the discourse; and (3) "voice," which is the relationship between the narrating and the narrated and the relationship between the narrating and text.
In applied studies of narratology, some researchers analyze and generate stories based on the theories constructed by Propp and Genette (Akimoto & Ogata, 2013; Ogata, 2011). Regarding these applied studies, Ogata (1999) suggested a study framework called "computational narratology." This research offered a literary study frame that fused an understanding of narratology with the generation of narrative through the use of AI and cognitive science, thereby broadening the concept of computational narratology.

2.3. *Interdisciplinary narrative generative system*

In this section, "interdisciplinary narrative generative system," which is derived from Ogata's narrative generation system, is outlined. Considerable research on story generation has been conducted in the domains of cognitive science and AI. Ogata and Kanai (2010) have completed advanced research on the interdisciplinary approach for the narrative generative system, which introduced narratology and literary theory. In addition, Ogata (2011) proposed the concept for the narrative generative system, which continues to be developed at two levels: at a broad level, the narrative generative system constitutes a method for symbolically understanding a human being and society in aggregate; and at a narrow level, a narrative generative system consists of a computer program. Recently, research in the clinical domain, which is based on Ogata's concepts, has been progressing. Specifically, the research is on disease and disability, and the results will be used for clinical support (Aoki et al, 2018; Seino et al, 2016, 2018).

2.4. *The Research Framework for narrative*

Several studies have suggested how to treat a narrative. In a macroscopic position, the framework of a narrative includes all life activities. In a micro position, a narrative (story) expresses the causalities of an event (Riessman, 1993). In addition, the narrative approach is defined as “the method of making the concept of a narrative key and presenting a certain phenomenon” (Noguchi, 2005).
Numerous disciplines study narratives. Ogata (2018) classified the research in the entire discipline with the narrative into three more categories (Table 1).

Table 1. The division between a discipline and narrative

	Classification	The contained discipline
1	Symbol level [Literature, art, and an entertainments domain]	Narratology, poetics, rhetoric, literary criticism, linguistics, folklore, folktale study, and an old tale, entertainment research, Noh play theory, kabuki theory etc., cultural anthropology and folklore, philosophy and thought, movie theory and cinematic review
2	A brain, a nerve, and a psychological level [Psychiatry and a psychological domain]	Psychology, psychiatry and psychopathology, psychoanalysis, tale treatment, neuro-psychoanalysis, brain science (The author added a postscript.)
3	Social level [Social domain]	Sociology, history, business administration, economics, jurisprudence, marketing theory, and advertising theory

3. The Narrative in Clinical / Psychological Approach

In this section, we will examined psychological narratology from the viewpoint of a narrative, and discussed its extension and the possibility of applying it to other areas of research. The conventional clinical domain attaches importance to evidence is objective. The concept of an evidence-based approach originated from evidence-based medicine (EBM) in the 1990s (Guyatt, 1991; Sackett, Richardson, Straus, Rosenberg, & Haynes, 1997),. Sackett et al. (1997) define EBM as the “the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of the individual patient.” Conversely, the idea of “evidence-based” has been reasonably criticized (Gergen, 1994) . Based on the limits of objective facts, social constructivism, which provides the relativist conception of reality, has come to be recognized (Gergen, 1999). Against the backdrop of social constructivism, a trend that allows a reconsideration of EBM from the perspective of the narrative is Narrative Based Medicine (NBM) (Greenhalgh & Hurwits, 1998). Since the 1990s, the narrative approach has attracted greater attention in interpersonal help domains, such as medical care, nursing, and psychology. The development of epistemology and the methodology that attached importance to narrative is known as the “Narrative Turn.” (Denzin & Lincoln, 2000).

4. Review of a Psychological Narratology

The system of a psychological narratology is proposed in this section, which shows that the clinical / psychological approach is extensible through post-narratology. Post-narratology is a narrative generative system that includes story generation by A.I. In order to actuate story generation, a literature review is performed with a clinical/ mental approach considering that the extension to post-narratology is possible. The system then constructs the story through AI and authors argue about the ability of a bridge to be carried out and whether by re-extending the clinical/mental approaches through story generation, a clinical contribution is possible. These approaches are concretely aimed at the approach (research and practice) in clinical domains, such as psychology. A psychological narratology is the theory of the clinical/ psychological approach that is relevant to the narrative in the broadest sense that the author has proposed.

4.1.1. *Narrative approach in the counseling*

In the counseling field, a "narrative" has come to be associated with the narrative therapy since the 1980s. In a narrow sense, psychological narratology is narrative therapy in psychotherapy and family therapy. Narrative therapy was advocated by White and Epston (1990) and it developed through the context of family therapy in the 1980s. White and Epston (1990) applied a classification of Foucault (1980) to narrative therapy, aiming to excavate the new "alternative story" from the "dominant story" that restricts life. This excavation is conducted by "externalizing" and seeking a "unique resolution" for the problem. In addition, Spence (1983) described the occurrence in which a narrator is in treatment, and a hearer hears the talk of a narrator as a tale. In analytic psychology, the role of a "tale," like the myth, is touched upon by Kawai (2001), who presupposes that the process of psychotherapy is a tale. In addition, in Gestalt therapy, Polster (1987) presupposed that the recovery of harmony of the heart requires a dialog, and proposed the method of making achieving this harmony is by having a real person sit on an "empty chair" that is assumed to be there and engage in a dialog.

4.1.2. *Consideration of a proposal*

Narrative therapy creates a narrative that provides the client with a tool for problem solving. A typical narrative approach heads toward the solution to a question as the client completes a tale. We propose the following possibility from the viewpoint of post narratology: what is treated is the discovery or description of a "dominant story" and an "alternative story," and the technique of externalization. There are two implications of this suggestion. First, the "dominant story" and "alternative story" may be identified in order to construct the theory of a narrative generative system. Second, computational technology may be able to realize two stories.

4.1. *Self-help group*

4.2.1. *The group for mutual aid*

A self-help group is practiced among those who suffer from the same illness and offers support through peer relations. For example, self-help groups may be formed for individuals addicted to alcohol, drugs/psychotropic substances, or gambling, or who are dealing with the same life obstacle or chronic malady. In recent years, self-help group have come to be viewed as a form of a narrative approach (Noguchi 2005). "Alcoholics Anonymous (AA)" was the first self-help group, which was developed in the United States in 1935, and is for recovery from alcohol addiction. AA was effective for treating alcohol addiction, and over time, many other groups adopted the method used by AA to deal not only with addiction or dependence but also for other problems. The typical definition of a self-help group is that of Kats &E. I. Bender (1976), who defined a self-help group as a small group of companions (peers) who provide mutual aid for the achievement of a specific purpose. AA meetings are regularly scheduled and the members who gather at each meeting talk in turn about various topics. The talk may be of experiences from one's past or the utterance of a single word may constitute a speaking turn (Noguchi 1995). The self-group does not have a technical definition (Ito, 2000).

4.2.2. *Consideration of a proposal*

The first-line specialty of support of a self-help group is "experiential knowledge." Experiential knowledge is the knowledge based on a member's experience, and serves as a source of support for a self-help group. The knowledge of members is more practical and more comprehensive than the knowledge of professionals (Borkman 1976). Each member's "experiential knowledge" is discussed and the collective knowledge if the group is accumulated and utilized as one tale knowledge from a group. We can interpret sharing of knowledge as the process of making one big narrative from two or more members' narratives. The process may be adopted into theory as a narrative generative system and may be realized as a system. For example, studies have proposed collect many texts that were spoken and generating the tale based on some typical elements. In addition, Ito (2000) examined a self-help group from the perspective of an individual tale. The results of the study indicated that the tale should have been regarded as an object with the effect of showing not the collective knowledge system of a group, but of a participant's self. To determine the degree of change for each participant, it was presupposed that it was important to determine the standard by which a participant made a tale "the good talk." In other words, it was important for a tale or suggestion to achieve a set basis for valuation. An evaluation of the tale could be conducted based on how the group absorbed the tale and how it affected the production and consumption of future tales.

4.2. *Open Dialogue*

4.3.1. *Philosophy of discussing*

In recent years, Open Dialogue (OD) is another psychological approach that has attracted attention in Japanese psychiatry. OD started in Finland in the 1980s and is a method of therapeutic intervention for an acute-term mental disease. In OD, two or more staff conduct a crisis intervention within 24 hours of a request. Patients and families practice OD every day by sitting in a circle and performing "OD." OD seldom uses a drug but makes a patient's critical situation cease (Saio 2014). All matters are discussed and determined as a characteristic portion of OD at the members' meeting place. In addition, an expert team discusses a patient and a family (exchange) in their presence, as they reflect on what is being shared with them. Olson, Seikkula, and Ziedonis (2014) summarized the following key OD items:

4.3.2. *Consideration of a proposal*

OD is considered to be a narrative approach given that a conversation called a dialog becomes a key medical treatment. OD contains narrative therapy and a common feature of OD is that it has an ideological background. In addition, it is also a common feature with narrative therapy to use reflecting. The 12 elements of OD consider tale generation and its effect, and the process is thought provoking. A dialogue is performed by two or more members of various positions. In addition, open-ended questions are used as an interlocutory trigger. OD may be able to be realized as an automatic dialogue system or a narrative system by speech recognition. By pulling out two or more viewpoints and observing relationships, an automatic dialogue system may be able to return utilize an extraction method to return an element as the narrative is being generated.

4.3. *Tohjisha-Kenkyu (Self-directed studies)*

4.4.1. *Empowerment approach by oneself and a companion*

In recent years, the "Tohjisha-Kenkyu" approach that originated in Japan has attracted international attention. Tohjisha-Kenkyu may be introduced as an open dialog in Japan (Saito 2016) and is an approach commonly used with patients with schizophrenia. In recent years, it has also been developed as a treatment for persons with developmental disabilities. Tohjisha-Kenkyu is an empowerment approach, and it happened among the activity and liveing of the person with disabilities (Mukaiyachi, 2013).

4.4.2. *Consideration of a proposal*

Tohjisha-Kenkyu provides suggestions for examining a narrative generative system and how to connect narrative generation for problem solving. In Tohjisha-Kenkyu, it is assumed that the party concerned will themselves become a hero of difficulties. In addition, Tohjisha-Kenkyu allows the narrative to pass and has suggested that it is a narrative approach tied to the problem solving of the party concerned. In addition, Noguchi (2005) reported that Tohjisha-Kenkyu is "a community of narration" that makes it "the community of the tale." Many audiences are in a "narrative community," and new narration becomes more certain with an existing audience. A narrative community is the space wherein new narration is shared and established. In addition, the aspects of changing attitudes, and how to identify and position attitudes Tohjisha-Kenkyuare are also common in narrative therapy. Moreover, the conceptualization of one's own experience and that of a friend is similar to the notion of "experiential knowledge" in a self-help group. In addition, when considering tale generation, an improvisatorial, unique understanding or an idea is important and shows how to utilize the result, which creates a "relation" as a "behavior" in a scene received by the narrative generative system.

5. Psychological Approach for Application of A.I.

In this section, two studies of the clinical/ psychological approach and the possibility of narrative generation by A.I. are introduced.

5.1. *Narrative generation for employment support of persons with disabilities*

In this section, the authors' research is introduced, in which we extracted text elements and employed a system of story generation using A.I. Specifically, we analyzed a disabled person's free descriptive answer to an interview question using the method of "informatics tale analysis" to extract the elements of the story. This method combines the technology of text mining, and some theories of a literary narratology. Text mining involves the quantitative analysis of text data based on natural language processing or data mining. After we extracted the element of the story from the individual's text response, the system proposed creating a collectivestory. An example of the narrative is shown below. Please note that to protect respondents' privacy, some answers are altered, but the basic meanings remain unchanged. In the authors' study, both the wider and narrow senses of narrative were used depending on the particular contexts of the analysis.

[Concerning work description] In response to the question: "Describe your job in detail."

Example 1: Since my legs were injured, I have been assigned paperwork. I am now engaged in computer work. (Physical disability)

[Concerning care and support at the workplace] In response to the question: "Specify the care and support provided in your workplace as a means of assisting your career."

Example 2: At my workplace, employees often get acquainted with coworkers at informal meetings, such as drinks parties. I did not attend these parties because of my disease, so I had no opportunity to get to know them. The other day, I took the opportunity of participating in one, and it worked much more effectively than I had expected. (Mental disorder)

Example 3: I first wanted to work without mentioning my disease. The other day, a relative of a person who had supported me asked me to work in their office. When I began working there, I soon found that this relative did not care about my condition, so it was a disastrous year. From this experience, I realized that I should conceal my disease because people consider me to be inferior to healthy people. (Intractable disease)

The authors' study analyzed the descriptive texts that reflected the personal stories of persons with disabilities, and clarified social stories of disability based on the results of multivariate analysis involving text mining.

This is an example from our research entitled "Employment Support and Story Generation for Students with Developmental Disabilities." The results indicated that story generation tends to materialize through the combining of existing elements with those from previous research (Seino, Enomoto, & Miyazawa, 2018). The figure below shows a story generative system.

The analysis of job descriptions analysis results were used to create the image. For cases in which this system was created, it was important to consider how elements were combined. The probability of a story being generated is attained by calculating the probability that it may happen and the strength of the relationship between the text elements. In addition, in order to utilize the result of the support generation simulation, it is important to understand what kind of elements have been collected. For example, the elements of support needs may be collected based on responses to a specific problem, and it may be possible to examine what kind of problem and support needs occur at what time, thus enabling the preparation of appropriate support.

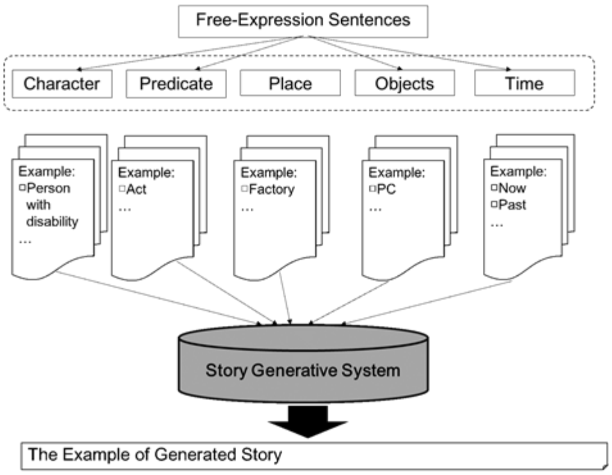


Fig. 1. The image of the story generative system

5.2. *Narrative generation for paper creation support of a student with a developmental disease tendency*

Joint research by Ogata and Aoki et al. (Aoki, et al. 2018) proposed students' paper creation support and narrative generative system based on the cognitive tendencies and action patterns of individuals with autism spectrum disorders. Aoki et al (2018) suggested that a paper creation might address the problem of a student with autism spectrum disorder. they explained why the problem occurs based on the cognitive tendencies and the action patterns of people with autism spectrum disorder, who may be sensitive to "a surprise and a gap" in the "discontinuous nature" of a story. Therefore, a story (paper creation), after making it continuous, stops the reaction of a student with autism spectrum disorder. The evaluation mechanism was set up and support was proposed for advancing a story by issuing directions on a macro and a micro level.

6. AI and ROBOT for interpersonal support and clinical practice

In this section, we review some of the practical applications of AI and robot in assisting patients and people with disabilities as a reference for considering AI and robotics applications in narratology.

6.1. *AI applications in psychiatric research*

Ikeda (2019) reviewed the use of AI in the psychiatric field.

According to this study, 393 articles were extracted from the article search database Pubmed using keywords related to psychiatry and AI. In the 382 abstracts obtained, the most frequently occurring psychiatric diagnoses were depression, schizophrenia, and Alzheimer's_disease. In addition, the research objectives of these articles were mainly the following four. (1) to improve the accuracy of medical treatment (e.g., diagnostic support, prediction of treatment effects and side effects, etc.), (2) to elucidate the etiology and pathogenesis of diseases (e.g., genome analysis, image analysis, etc.), (3) to develop new treatment methods (e.g., drugs and medical devices), and (4) to reconstruct diagnostic concepts and systems. The AI-related terms in those papers are, in descending order of frequency, support_vector_machine, random_forest, logistic_regression, deep_learning, natural_language_processing, decision_tree, convolution decision_tree, convolutional_neural_network, and naive_bayes.

The following is an overview of the specific studies.

6.2. *Use of machine learning in AI*

AI is being used in research to predict disease and identify contributing factors.

Rahman et al (2020) developed a method for predicting autism in newborns by machine learning the parents' age, socioeconomic status, medication, etc. from their medical records.

This algorithm clarifies, autism and the use of substances such as caffeine and certain antidepressants by parents are relevant. The results may also help to identify factors that may induce autism.

AI has also been used in research to differentiate between diseases and disease subtypes. Stevens et al (2019) analyzed behavioral data in autism and found two comprehensive behavioral profiles. Each profile had its own subgroups based on the severity of different traits.

Chand et al (2020) used machine learning to analyze brain scans of schizophrenia and clinical information such as age of onset and medication use. As a result, they identified two subgroups of schizophrenia based on different patterns of brain structure. One of the subtypes showed brain volume increasing almost normally in two regions, rejecting previous ideas that schizophrenia is associated with reduced brain volume.

Koike and his research group developed a machine learner that determines schizophrenia and developmental disorders through image analysis by machine learning of magnetic resonance imaging (MRI) brain structure images (Yassin et al, 2020).

Thus, machine learning research is expected to be applied to differential diagnosis and treatment prediction in clinical practice. It may also be useful for discovering subtypes of the same disease and understanding the reasons for differences in nature and severity.

6.3. *Robots for Supporting Children with Autism Spectrum Disorder (ASD)*

Hirokawa (2014a, 2014b) et al. have proposed a robot intervention activity for children with autism spectrum disorder (ASD), a developmental disability, to improve their social skills and promote their communication abilities. They have developed a robot manipulation interface that can be used by supporters for this purpose, as well as a method for quantitatively measuring face-to-

face behavior and facial expression during intervention activities, and have verified its effectiveness.
Jain et al (2020) developed a machine learning algorithm that automatically tells a robot when to encourage Autism Spectrum Disorder (ASD), as a therapeutic intervention. The reason for the use of robots in these proposals is that humanoid robots, with mechanical features that are highly compatible with ASD children and physical and social characteristics similar to those of humans, are expected to play an important role in interventions for ASD children.

7. Discussion

7.1. *Extension and possibility of a psychological narratology*

In this paper, the psychological narratology was divided into some items and reviewed. The author showed that the narrative approach in clinical domain, defined as psychological narratology, went beyond the narrative therapy in psychotherapy and was expanded to various experiments and research. Moreover, is thought that three features are among the psychological narratologies reviewed in this paper. As the first, many practice is not simple technology and method but stance, and a way of thinking. Since it has narrative approach against the background of social constructionism, fixation is avoided and relativized. As the second, the narrative can point out being made communally. The third point is all the tales are built through a dialog. They were not monologues and were constructed through colloquial expressions.
Based on these points, the narrative generative system is considered in both a narrow sense and a broad sense. The following possibilities exist regarding the automatic tale generation machine realized as a narrative generative system in a narrow sense, i.e., a computer program etc. (1) It is the construction of a fluid system. For example, a random element is taken in accepted. This system carries out narrative generation by combining elements. (2)Two or more persons are the systems that generate a tale communally. (3) It is a system of the tale generation through simple dialogue but not an input.
In addition, the following is mentioned as a suggestion for the practice of this research. A psychological narratology is not merely a methodology but is also the mechanism for converting of a relative sense of values. It is given significance with conversion (or improvement) of the worth of the tale of the party concerned and the party concerned. The following specific suggestions are based on such a system. (1) Consider the community itself which generates a tale and utilizes it to be a system and as well as its construction and practical use. (2) The idea of a system and the realization that this idea places more value on the process rather than the contents of tale generation. (3) Get participation and collaboration of the concerned parties towards realization and practical use of a system.

7.2. *Potential Applications of Psychological Narratology to AI or Robot*

Research had been accumulated on the use of AI and robot for clinical support and it was confirmed that algorithms had already been developed or started to be implemented on robots. In addition, the previous studies had the following four main objectives (Ikeda, 2019). (1) Improvement of medical treatment system, (2) Elucidation of etiology and pathogenesis, (3) Development of new treatment methods, (4) Reconstruction of diagnostic concepts and systems Firstly, by machine learning of existing data, the use of AI in clinical support has enabled to predict and identify diseases, discover subtypes, and identify factors in the development of diseases (Stevens et al, 2019; Rahman et al, 2020; Chand et al, 2020; Yassin et al, 2020).
If this method is applied to psychological narratology, applications to detect subtype, and to predict and identify tasks are considered throughout machine learning of speech, dialogue, and written text data.
Secondly, the use of robots for clinical support had been studied to improve the convenience and effectiveness of robots for supporting children with autism spectrum disorder (Horikawa, 2014, 2014b; Jain et al, 2020). If the application to psychological narratology is considered from this point, incorporating narrative elements into robot for autism spectrum disorders is possible. Also, it may be necessary to examine the effectiveness of robot support for non-autism spectrum disorders.

8. Conclusion

This research defines the narrative approach of a clinical domain as a psychological narratology, and examines its extension and possibility for application on AI or robot. As a result of the review, psychological narratology was confirmed to have widely expanded beyond the so-called narrative therapy of psychotherapy and family therapy. In addition, suggestions from these practices are thought-provoking and also utilize expansion, realization, and the system of the narrative generative system in a narrow sense and a broad sense for actual problem solving. In addition, AI or robot research in the clinical domain has been identified. This gives suggestions to the extension of psychological narratology. It is meaningful to consider the practice of a clinical domain from the viewpoint of a narrative, as well as in respect to clinical reduction. We expect that expansion of new discovery and practice will be performed because a supporter, and a researcher in connection with a theme, and the researcher of related domains, such as A.I., computing science, cognitive science, will argue and collaborate in the future.

References

1. Y. Yamada, Shitsutekishinrigaku no houhou: Katari o kiku [Method of qualitative psychology: Listening to narrative]. Shinyousha, 2007.
2. G. Prince, *A dictionary of narratology* (Revised ed.). University of Nebraska Press, 2003.
3. J. Bruner, *Actual minds, possible worlds*, Harvard University Press, 1986.
4. K. Seino, K, Y. Haruna and S. Ishizaki, Employment status and support needs of persons with disabilities in Japan: An analysis of narrative using narratology and text mining on a national survey. in T. Ogata, and T. Akimoto (eds.), *Computational and cognitive approaches to narratology* (pp .245-275). Information Science Reference , IGI Global, 2016.
5. K. Seino, Y. Enomoto and S. Miyazawa, Narrative analysis of employment support for students with developmental disabilities: Using an objective analysis of free-expression answers. In T. Ogata and S. Asakawa (eds.), Content generation through narrative communication and simulation (pp. 341-357). Information Science Reference , IGI Global, 2018.
6. C. D. Riessman, Beyond reductionism: Narrative genres in divorce accounts, in *Journal of Narrative and Life History*, 1(1), 1991, pp.41-68.
7. D. Clandinin, D. (ed.), *Handbook of narrative inquiry: Mapping a methodology*. Sage Publications, 2007.
8. T. Ogata (2010). “Monogatari seisei sisutemu” no haikei oyobi monogatari to bungaku no aida [Backgrounds of the “narrative generation system” and “relations between narrative and literature”]. In T. Ogata & A. Kanai, Monogatariron no jouhougaku josetsu: monogatari seisei no shisou to gijutsu wo megutte [An introduction to the informatics of narratology: about the thought and technologies of narrative generation] (pp. 186-258). Tokyo, Japan: Gakubunsha.
9. V. Y. Propp, Morphology of the folktale (L. Scott, Trans.). University of Texas Press, 1968 (Original work published 1928).
10. G. Genette, Narrative discourse: An essay in method (J. E. Lewin, Trans.). Cornell University Press, 1980. (Original work published 1972).
11. T. Akimoto, T and T. Ogata, Towards a discourse mechanism in narrative generation system: Proposal of a system introducing narrative discourse theory and reception theory. Cognitive Studies, 20(4), 2013, pp.396-420.
12. T. Ogata, Narrative generation system as the practice of “informatics of narratology.” Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, 23(5), 2011, pp.676-685.
13. T. Ogata, Proposal of computational narratology. In Proceedings of IPSJ SIG Computers and the Humanities (pp. 53-60), 1999.
14. S. Aoki, T. Ogata and J. Ono, Narrative generation related to cognitive patterns seen in ASD -From the perspective of “Surprise”-, In Proceedings of the 35th Annual Meeting of the Japanese Cognitive Science Society (pp. 652-660). (2018).
15. C. K. Riessman, Narrative analysis. Newbury Park, CA: Sage, 1993.
16. Y. Noguchi (ed.), Naratibu no rinshoshakaigaku [*Clinical sociology of narrative*]. Keiso shobo, 2005.
17. T. Ogata, Narratology of Narrative Generation or Post-Narratology, in *Cognitive Studies*, 25(2), 2018, pp.200-217.
18. G. H. Guyatt, Evidence-based medicine. American College of Physicians Journal Club, 1991. 114, A-16.
19. D. L. Sackett, W. S. Richardson, W. Rosenberg and R. B. Haynes, *Evidence-based medicine: How to practice and teach EBM*. New York, NY: Churchill Livingstone, 1997.
20. K. J. Gergen, *Toward transformation in social knowledge*. Sage Publications, 1994.
21. K. J. Gergen, *An invitation to social construction*. Sage Publications, 1999.
22. T. Greenhalgh and B. Hurwitz (eds.), *Narrative based medicine*, BMJ Books, 1998.
23. N. K. Denzin and Y. S. Lincoln (eds.). *Handbook of qualitative research*, Sage Publications, 2000.
24. M. White and D. Epston, Narrative means to therapeutic ends. New York: Norton. in *Transformative Learning Conference*, Oct. 1990, pp. 23–25, 2003 184/492
25. M. Foucault, Power/knowledge. (C. Gordon, L. Marshall, J. Mepham and K. Soper, Trans.). Pantheon Books ,1980. (Original work published 1972, 1975, 1976, 1977).
26. C. D. Spence , Narrative persuasion, in *Psychoanalysis and Contemporary Thought*, 6, 1983, pp. 457-468.

27. H. Kawai, Shinriryoho niokeru Monogatari no Igi [Significance of the "narrative" in psychotherapy], in *Japanese Journal of Psychotherapy* , 27, 2001, pp. 3-7.
28. E. Polster, *Every person's life is worth a novel*. W. E. Norton & Company Inc, 1987.
29. A. H. Katz and E. I. Bender (eds.), *The Strength in Us: Self-Help Groups in the Modern World*. New York: New Viewpoints, 1976).
30. T. Borkman, Experiential knowledge: A new concept for the analysis of self-help groups. in *Social Service Review*, 50 (3), 1976, pp.445-456.
31. T. Ito, Serufu herupu gurupu to kojin no monogatari [Self-help groups and personal stories]. in *Japanese Sociological Review*, 51(1), 200, pp. 88-103.
32. T. Saio, Open dialogue approach to acute psychosis: Is its effectiveness well-established? in *Clinical Evaluation*, 42（2）, 2014, pp. 531-537.
33. I. Mukaiyachi, Tojisha-kenkyu toha: Tojisha-kenkyu no rinen to kousei [What is a Tojisha-Kenkyu? The concept and structure of a Tojisha-Kenkyu]. Tojisya-Kenkyu network (https://toukennet.jp/?page_id=56) , 2013.
34. S. Ikeda, Current Trends, Issues and Potentiality of AI Technology in Psychiatry. in *The 33rd Annual Conference of the Japanese Society for Artificial Intelligence*, 2N5-J-13-03, 2019, pp. 1-4.
35. R. Rahman, A. Kodesh, S. Z. Levine, S. Sandin, A. Reichenberg and A. Schlessinger, Identification of newborns at risk for autism using electronic medical records and machine learning, in *Eur Psychiatry*, 63(1), 2020, e22.
36. E. Stevens, D. R. Dixon, M. N. Novack, D. Granpeesheh, T. Smith and E. Linstead, Identification and analysis of behavioral phenotypes in autism spectrum disorder via unsupervised machine learning. in *Int J Med Inform*, 129, 2019, pp. 29-36.
37. G. B. Chand, D. B. Dwyer, G. Erus, A. Sotiras, E. Varol and D. Srinivasan,et al. , Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning. in *Brain*, 143(3), 2020, pp. 1027-1038.
38. W. Yassin, H. Nakatani, Y.Zhu, M.Kojima and Owada K., et al., achine learning classification using neuroimaging data in schizophrenia, autism, in *Translational Psychiatry*. 10(1), 2020, pp.278.
39. M. Hirokawa, A. Funahashi, and K. Suzuki, A Doll-type Interface for Real-Time Humanoid Teleoperation in Robot-Assisted Activity: A Case Study, in *Proc. ACM/IEEE Intl. Conf. on Human-Robot Interaction*, 2014, pp.174-175.

M. Hirokawa, A. Funahashi, Y. Itoh and K. Suzuki, Design of affective robot-assisted activity for children with autism spectrum disorders, in *Proc. of The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 2014, pp. 365--370.

40. S. Jain, B.Thiagarajan, A.Shi, C. Clabaughand and M.J. Matarić, Modeling engagement in long-term, in-home socially assistive robot interventions for children with autism spectrum disorders, in Sci. Robot, 5(39), 2020, eaaz3791.

A Relationship between Narratology and Marketing

Akinori Abe
Faculty of Letters, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan

Abstract

We have been studying an expression of the taste of Japanese sake. Actually, the descriptions of the taste can be regarded as stories. Accordingly, we conducted an experiment, in which we asked participants to draw descriptions of the taste of the Japanese sakes and to design labels of them. From the results of the experiment, we will analyse the relationship between narratology and marketing of Japanese sake.

Keywords: Narratology, Taste of sake, Marketing.

1. Introduction

We have been studying an expression of the taste of Japanese sake (alcoholic drink) [1,2,3,4]. Since we do not have a dictionary of the taste of Japanese sake, we tried to create words or phrases for the taste of Japanese sake. Generally we do not describe the taste of sake as a still situation, but as a changing situation. During describing the taste, we will deal with time. We will describe a gradual change of the taste. Thus the descriptions of the taste can be regarded as stories. The descriptions are interesting as a set of corpus. However more interesting matter is that we may use the descriptions for the marketing. Since the descriptions are generated from the taste of sake, it should contain the taste. And if labels are designed based on the description, it should contain the information of the taste.

Usually labels of the Japanese sake are not attractive. At least, what labels should inform us are name of the sake. No attractive matter is necessary for selling the sake. However, according to the Jensen's theory[5], it is important to prepare a story for selling products. Accordingly we think it will better to include a story in the labels of sakes. At least labels of sakes should be attractive for a certain generation.

Therefore, we conducted an experiment, in which we asked participants to draw descriptions of the taste of the Japanese sakes and to design labels of them. The design should include the taste information of the sake.

We could collect several results. From the results of the experiment, we will analyse the relationship between narratology and marketing of Japanese sake.

2. Description of Sake as a Story

2.1. *Descriptions of the taste of sake*

In several books, we can find the descriptions of taste of sake. For instance, for the taste of Japanese sake, in “LOVE ♡ Japanese sake!, 2014, 2 Gakken,” we can find the following descriptions:

- Kamoshi-bito Kuheiji (醸し人九平次) Eau Du Desir 2012
  At the first moment in the mouth, strong fruit flavour can be felt. After a proper maturing time, the mouth is fragrant with the smell of honey and vanilla. A good flavour is spread and in the center of it splendid sour can be found... (口にした瞬間は果実のフレーバーが強く感じるものの、適度な熟成を経ると蜂蜜やバニラの香を覚えます。旨味は華やかな酸を軸に広がりますが....)
- Isojiman Junmai ginjou (磯自慢純米吟醸)
  Refreshing scent such as the scent of muskmelon and pear is elegant... (マスクメロンや洋ナシを思わせる爽快とした香りは気品すら感じさせるもの....)

The above descriptions do not include stories. Perhaps, for the evaluation, subjective descriptions will not be preferable. On the other hands, descriptions in several book contain the stories of production such as "Mr. Fernandez who was 79 years old had spent all his life for the wine production (El Vinclo 2010 by *Real Wine Guide Vol. 52 (2016, Winter)*)" and "This taste was born by the malted rice which the president Kimizuka in the winter puts his heart and soul into producing and the fermentation technique that pulls out the potential taste of Fusakogane (an eating rice only in Chiba prefecture) at the best (Naruka special junmai direct packing raw [white] (鳴海特別純米直詰め生【白】) the homepage of Nishiura liquors shop (酒舗にしうら))."

Previously, we described the taste of Japanese sake[1]. I will show one of the descriptions; The followings are records of two person's expressions for the taste of Japanese sake. We tasted the same sake (Shizengou seven (自然郷 セブン)) at the same time in the same room. We used different cup but the environmental condition will be almost the same.

- A: The first impression in my mouth was oh good taste. However, after that the taste disappeared as if water escaped from the side of the tongue. The first taste of sake did not spread and taste like water went through tongue then escape from my mouth between teeth. Very "interesting" taste.
- F: The first half of tongue can feel good taste. I felt oh good taste. However gradually the taste disappeared as if the taste was veiled in mist. At the last, I had an aftertaste such that sugar is diluted in water. I felt quite a few satisfaction. If I dare to describe the last half taste as a refreshing taste, it can be described that the taste is tightened.

In fact, we used the phrases for the flow of time. Though they do not contain the story of the producer nor process of production, the flow of time during the tasting can be regarded as a story.

2.2. *A label of sake*

Certain labels of Japanese sake contain certain illustrations. "The Japanese sake stories (日本酒ものがたり)" (http://sakemono.com/) by the Japanese sake × produce production project (日本酒×作家創作プロジェクト) is conducted as an intention to draw a story in the Japanese sake by artists. For instance, the illustration on the label of Hanahimesakura Shizuku-hime non-filtered junmai ginjou no-water-added sake (華姫桜 しずく媛 無濾過純米吟醸原酒) is shown in Figure 1. This is an anime-like illustration.

Though this sake is a non-filtered junmai ginjou no-water-added sake, its taste was fully matured opposite to

Fig. 1. A label of Japanese sake (華姫桜しずく媛無濾過純米吟醸原酒)

our expectation. According to the home page, her name is Sakurashizuku (桜雫) and she is 138 years old. Accordingly the sake might be matured sake. However, I do not think actual taste of the sake is coincide with the lady in the illustration. At least, the taste is not the similar to what I expected when I saw the label. This illustration might represent the image of the sake as the sake brewery imaged, but does not represent the taste of the sake. If they want to sell this sake to those like idols, it might be successful. Because the illustration contains the story which will be liked by idol lovers. This label includes a story with this lady. We would like to buy the sake with regard to the story read in this label. Thus a story in the product will play an important role in marketing.

3. Experiment

We conducted an experiment as follows;

3.1. *Objectives*

Sub-headings should be typeset in boldface italic and capitalize the first letter of the first word only. Section number to be in boldface Roman. We reviewed the role of narratology in the previous section. We concluded in the previous section that a story in the product will play an important role in marketing. Accordingly, we would like to see how can we generate stories from the product and how can we express the stories for the marketing.

3.2. *Participants*

Seven persons from 21 to 29 years old took part in the experiments. Some of them frequently drink Japanese sakes and describe the taste of them.

3.3. *Method*

We used the following six kinds of sakes: Ichijima junmai ginjou raw no-water-added sake (市島 純米吟醸生原酒), Azumaichi junmai daiginjou (東一 純米大吟醸), Azumaichi Yamadanishiki junmai-shu 64\% polishing (東一 山田錦純米酒 64\%磨き), Azumaichi Nero, Azumaichi junmai daiginjou drip squeezing (東一純米大吟醸 雫搾り), Inabazuru junmai ginjou gouriki freshly squeezed non-filtered raw no-water-added sake

Fig. 2. Japanese sakes used in this experiment

(いなば鶴 純米吟醸 強力 しぼりたて 無濾過生原酒)

We asked the participants to taste Japanese sake to describe the taste. The description could include texts and figures. Then we asked them to create catch copies for the sake. In addition, we asked them to design the label of the sake. The target person was young person in 20's. Those descriptions and drawings were drawn on the papers in A3 size. They used several pens such as coloured pens and pencils.

3.4. *Results*

We collected such results as shown in Figure 3.

First, we show some descriptions of the taste of sake. All descriptions were written in Japanese. We translated them in English.

- Ichijima junmai ginjou raw no-water-added sake
  1) Delicious. I like this because I can smell and taste rice. Since it contains spicy taste, the taste comes gradually. At first, I feel gentle taste, but in latter half I feel spicy taste.

  2) Jyowa jyowa--- The fragrance disappears suddenly and clearly. A refreshing fragrance which is going out from nose. For the summer season? I can not say it is a refreshing type. With snack it is not good. With a Camembert cheese, the taste becomes soft. For a party? It is difficult to drink without any foods.

  3) (fragrance) I like the fragrance.

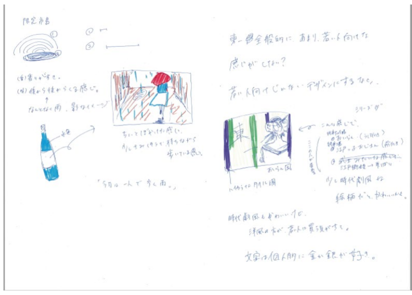


Fig. 3. One of results in this experiment

  (taste) The taste comes again and again. I feel somewhat an image of rain and shadow.

  4) (fragrance) I feel a straight alcoholic fragrance.

  (taste) Strong sour and slightly sweet. After taste includes spicy. Bitter taste is weak.

  5) Spicy banana

  mine) I feel fragrance of cloth. First sweet taste such as pears comes to mouth and spread slightly. Unfavourable taste flows out of mouth but it is not hate taste. With a duck, I can taste without conflictions. Smoked taste becomes weak.

- Azumaichi junmai daiginjou
  1) I feel sweet and floral feeling. When I open the sake, I feel a fragrance from the 20cm distance, the taste is soft and I feel umami of rice. The more I drink, I become aware of biting taste. If I drink a lot, spreads of the sake will change? It is easy to drink, so I can drink more and more. I think I can get drunk comfortably. It will be good for drinking when tired? But it is not good to drink before holiday. I might not be aware of having a hangover.

  2) Refreshing fragrance. The first attack is pliant sweet. It is rather bitter. The balance of the taste is good. After a while, I feel a bitter taste more. The target generation of the sake will be elder persons. Younger generation will not like this. The taste is bitter, but after taste does not remain long. So it will be good for with a meal. After a long time, the taste changes flat and I do not feel a bitter taste.

3) (fragrance) Fresh and young type. (taste) strong sour. At (1), a prickly taste comes and at (2) also a prickly taste comes. Then the taste spread to (3) rapidly. The taste remains more at (1) and (2) than at (3). It has a peculiar taste but is easy to drink. It is a feminine type sake.

4) (fragrance) Sweet fragrance of rice which comes my nose straightly and a straight alcoholic fragrance.

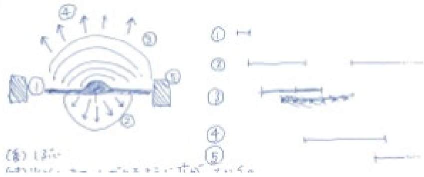
(taste) First, I feel a sweetness of rice. The taste becomes bitter as the sake enters the back of tongue. At the throat I feel a biting taste. It is easy to drink (because a bitter taste dose not come suddenly). After taste is sweetness of a rice and I feel a biting taste in the throat. At the back of throat, I feel a little bitter taste.

mine) Fragrance like a dust. A taste like a fruit (apple) comes in the shape of round and thickly. The taste does not spread and comes heavily. However, At the top of the tongue, spicy taste. In the mouth, a remaining taste is faintly sweet and clear. If it does not have spicy taste, it is clear taste.

- Azumaichi Yamadanishiki junmai-shu 64% polishing
  1) A sharp feeling. Cool. Good for summer. Cool feeling → I cannot feel the fragrance. The taste is sour? like a caramel. I feel bitterness rather than sweetness. But sweet. It's difficult to deal with expressing the taste. The taste spreads a little. The taste stops at the place where I feel like to go more. It the type of sake that tantalizes after appealing to me. It is like a quiet and cool pretty girl such as Rei Ayamaru (綾〇レイ).

  2) It is a normally delicious sake. It is neither good nor bad. It is better than Ichjima and Azumaichi Junmai daiginjou. If I do not have any sake, I will drink. I can drink with every foods.

  3) (fragrance) astringent. (taste) A prickly taste comes and spreads with numbing. After drinking the sake, at (2) and (4) the taste remains. In addition, at (5), the taste remains. The taste is like a little bitter chocolate remains in the mouth (A cacao degree is high.). The taste in (3) moves back and forth according to the way of putting it in my mouth. The taste in (3) spreads overall slowly. The after taste is a chocolate.



4)(fragrance) Rather strong fragrance.
(taste) Just after putting the sake in my tongue, I feel a spicy taste. I feel sweetness little and the after taste has a bitterness. But the taste is more refreshing than that of Azumaichi junmai ginjou. I feel a spicy taste at the top of my tongue.

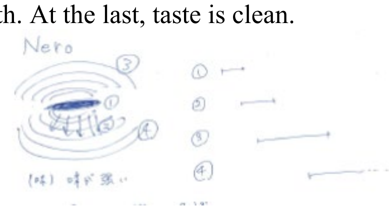
5) Middle: clear, a clear stream, acid, forest, summer, late summer, September, sun light through the trees
Later: Autumn rice field with swaying inaho, refreshed taste. A fine autumn day. A rim of the taste is clear.

mine) A little fragrance of an alcohol exists bottom of the sake. The taste is matured and at the bottom of the sake, an alcoholic taste sinks. At the upper side of the sake exists thin sweetness. I feel a spicy taste at the top of my tongue. I feel an astringency spreading around my cheek.

- Azumaichi Nero
  1) Its taste is similar to that of Azumaichi junmai daiginjou? If it is said as sweets, it will belong to the sweets. The taste is rather sweet and it comes gradually. The taste is not spreading wider than that of Azumaichi junmai daiginjou. I feel like a drinking sake more than Azumaichi junmai daiginjou drip squeezing.

  2) I can smell nothing. Its taste is calm and gentle. I feel carbonic acid. I think it is better to drink with yakitori. Perhaps with sashimi. After a little while, it becomes melting. And more time after, it becomes sticky.

  3) (taste) Its taste is strong. (4) remains ..., but is rather smooth. At the last, taste is clean.



4) (fragrance) I feel a straight fragrance. No sweet fragrance I feel.
(taste) The taste of sweetness of a rice and sour strongly comes. The after taste is slightly (gentle) spicy and bitter.

mine) I feel a closed fragrance. A fruit type of taste comes smoothly. It spreads very little and on my tongue sweetness remains. I feel little unfavourable taste. If I inflate the sake in my mouth, a little sour comes and the taste of the sake becomes very well balanced. At the last, I feel a spicy taste on my tongue, but I mind a little.

- Azumaichi junmai daiginjou drip squeezing
  1) The spreading of the taste is a little. I have a unremarkable impression. Spicy? I feel a prickly taste. I think I can recommend to everybody?

  2) I feel a rich fragrance. This is a hoping sake. The taste is not bad. I will drink later. A cheap taste. I can dare to drink it in a train alone. For a party. After several minutes, a bitter taste comes a little.

  3) (fragrance) I smell a little. pretty.
  (taste) like a melon. pretty. At (3) the taste comes with a heavy impact. “It is not just a pretty face.” I feel it is like a pearl.

  4) (fragrance) Fragrance of sweetness of rice and alcohol.
  (taste) I feel a taste of strong flavour of a rice. In addition, I feel sweetness and slightly acidity. On the throat, I feel relatively strong spicy taste. Even if I drink it with air, a spicy taste comes.

  5) Mowa--n, unformel, it has an outline.

  mine) I smell fragrance of cloth a little. On the periphery, clear sweetness exists. The taste enters in mouth cleanly. I feel a spicy taste over the sweetness. Its clear sweetness is like a delicious water and beautiful. However, I want a more spreading taste. Without the last spicy taste, it will become a non-stimulative sake? It it had umami or rich flavor, it will be the best.
- Inabazuru junmai ginjou gouriki freshly squeezed non-filtered raw no-water-added sake
  2) I drink it with sweets. The taste is not bad. (It might because I drink it with sweets.) It smells like fruits (green apple).

  mine) At first, the taste is closed. A sour taste stands out, but is not strong. It is easy to drink, but does not have a special feature.

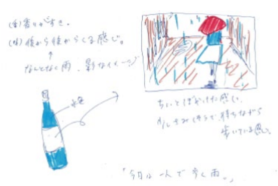
Next, we show some catch copies for sakes.

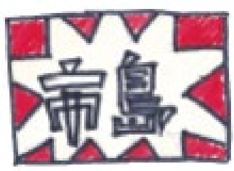
- Ichijima junmai ginjou raw no-water-added sake
  2) Best for the summer party... with cheese and ham.
  3) It's raining... today I walk alone.
  4) Spreading taste.
  5) If we get drunk, everything is tasteful.
- Azumaichi junmai daiginjou
  2) For the sake lovers!! The best in Azumaichi.
  3) Scared 怖い lady (gentle lady)
  4) Seven changes
  5) Flower blooms four times.
- Azumaichi Yamadanishiki junmai-shu 64\% polishing
  1) A friend with an evening drink.
  2) The entry item of Azumaichi!!
  3) The first step to adult.
- Azumaichi Nero
  2) It's very clear taste. So you can taste it with both meats and fishes.
  5) 25 years old, week end, girls party.
- Azumaichi junmai daiginjou drip squeezing
  2) Please taste slowly.
  4) Happiness of a drop.
  5) Metamon transforms Japanese sake. Drinking rice ball sake.

Several catch copies were collected. These copies are for younger generation. An interesting copy is that for Azumaichi junmai daiginjou. It might have changing taste, accordingly they generate phrases with words meaning change.

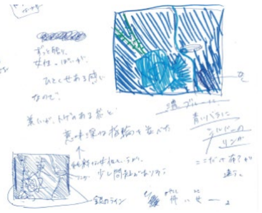
Then according to the catch copies, participants created labels for the sakes as shown below:

- Ichijima junmai ginjou raw no-water-added sake

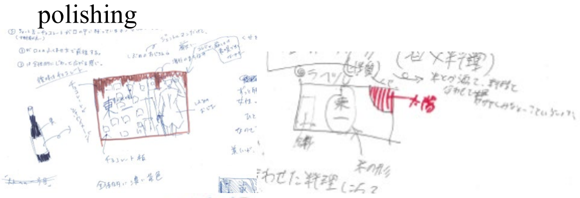




- Azumaichi junmai daiginjou

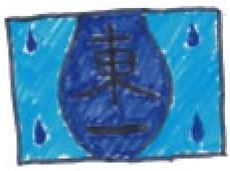


- Azumaichi Yamadanishiki junmai-shu 64\% polishing



- Azumaichi Nero

- Azumaichi junmai daiginjou drip squeezing



Some labels have stories.

4. Discussions

In the experiment, we collected several results. They are descriptions of the taste of the sake, catch copies for the selling the sakes, and labels of sakes. We expected stories in the descriptions of the taste. Some participants described changing tastes. We think these are stories for the sakes.

A sommeliere sometimes describes a taste of a wine with a scene such as “walking in a forest with stepping firmly dead leaves.'” This expression comes from the changing taste. We think for the description of the taste of Japanese sake, this type of expression can be made.

Reviewing the labels, we think a story of the taste is described in certain label. And the story (narratology) functions well in the labels. We think illustration in the label expresses the story of the taste of the Japanese sake. We think it will be better to express the story of the taste of the sake. It has a lot of information to select a product.

We did not conduct an experiment to evaluate the design of labels by somebody. This is a unsolved question for the future.

5. Conclusions

In this paper, we conducted series of experiments. First, we asked the participants to taste Japanese sake to describe the taste. The description could include texts and figures. Then we asked them to create catch copies of the sake. In addition, we asked them to design the label of the sake.

We could collect several expressions from this experiment. The expressions are very interesting. The more interesting results are labels designed by the participants. We think some labels can be used for the marketing.

We have not evaluate the designs. In the future, it is necessary to evaluate them by outsiders.

Acknowledgements

We thank Mr. Ippei Setou for offering us Japanese sakes (Azumaichi).

References

1. Akinori Abe: About the better expression of taste, Proc.of the 2nd. Int’l Workshop on Language Sense on Computer in IJCAI2017, pp. 28–34 (2017)
2. Akinori Abe: Is the expression of taste abstract or representational?, Proc. of SMC2018 Workshop on Narratology and Cognitive Content Generation, pp. 18–23 (2018)
3. Akinori Abe: Expression of the taste of Japanese sake and metaphor, Proc. of ICAROB2019, pp. 589–592 (2019)
4. Akinori Abe: Taste of Japanese Sake as Metaphor and Shape, Proc. of IJCAI2019 The 3rd. International Workshop on Language Sense on Computer, pp.29–34 (2019)
5. Rolf Jensen: The Dream Society: How the Coming Shift from Information to Imagination Will Transform Your Business, Mcgraw Hill (1999)

Layout decision system for multiple production lines using work-flow-line and GA

Masato Noda
Department of Intelligence Science and Engineering, Gifu University, Yanagido 1-1, Gifu, Gifu 501-1194, JAPAN

Hidehiko Yamamoto
Department of Mechanical Engineering, Gifu University, Yanagido 1-1, Gifu, Gifu 501-1194, JAPAN

Hirohumi Tsuji, Yasuhisa Terawa, Yoshinori Nakamura, Masayuki Tsuchida
InfoFarm co.,Ltd, Yanaizu-cho Distribution Center 1-8-4 Gifu, Gifu 501-6123, JAPAN

Katsuaki Yamada, Yukiyasu Kuriyama
Kai industries co., ltd, Oyana 1110 Seki, 501-3992, JAPAN

E-mail: y4525062@edu.gifu-u.ac.jp, yam-h@gifu-u.ac.jp

Abstract

We develop the system to decide the efficient layout of assembly production line by using Genetic Algorithm (GA). We call the system as System of Production-line-layout Decision by Chameleon-code and GA (PDCG). PDCG decides the efficient layout of production line by using GA, work-flow-line acquired by Chameleon code and the machine breakdown data. PDCG evaluates the layout efficiency by calculating the operator's walking time in order to fix the machine breakdown occurred on the production line. By the evaluation.

Keywords: Genetic Algorithm, Chameleon-code, Assembly production, Work-flow-line, Machine breakdown

1. Introduction

In order to improve the production efficiency, many Japanese companies adopt Kaizen. The line manually performed by several operators is needed to be also improved. In this study, we develop the system of Production-line-layout Decision by Chameleon-code and Genetic Algorithm (PDCG). PDCG determines the layout of multiple production lines that some operators work. The determination is given by using Genetic Algorithm (GA), and the work flow lines obtained from Chameleon-code. PDCG is applied to the razor assembly line to verify whether it is effective in improving work efficiency.

2. Work-flow-line with Chameleon code

Chameleon code is the new color barcode whose function has a high speed and precision recognitions as shown in Figure1.This study puts Chameleon code on the top of an operator's hat and obtains the operator's work flow line data by overhead cameras. The acquired work flow line data includes the name of the operator, the camera names, the dates and times, and the X and Y coordinates of the chameleon code.

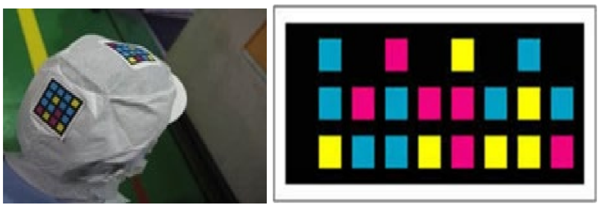


Fig. 1. Chameleon-code

3. PDCG outline

PDCG is the system to determine the layout of production lines using GA, based on the work flow line database acquired from the chameleon code and the machine breakdown database. Specifically, PDCG calculates the walking time of operator moves to fix machine breakdowns, evaluates the operator's walking time and determines the better layout where one operator manages some production lines. As shown in Figure 2, PDCG consists of the condition module and the GA module. Condition module reads the acquired work flow line database and sets various parameters for GA. GA module determines the better production line layout by using GA.

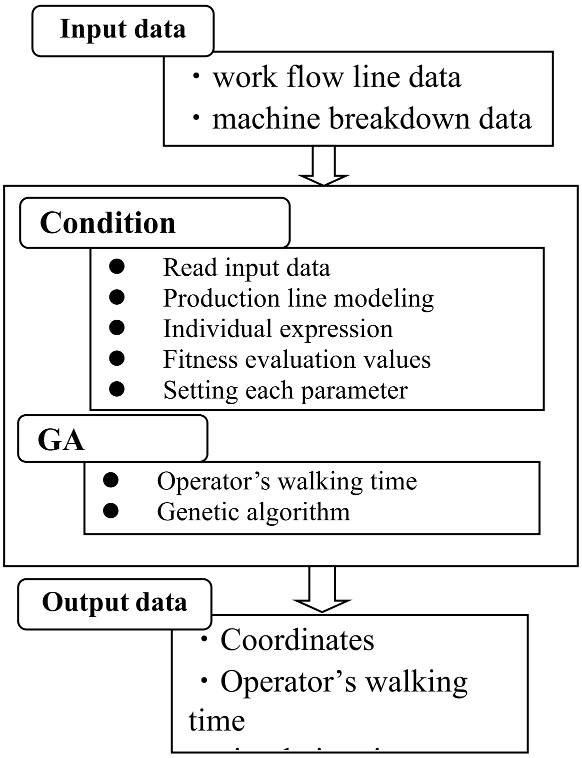


Fig. 2. PDCG outline

4. Individual representation

The individual representation of GA used in PDCG adopts the structural gene, as shown in Figure 3. We express the x-coordinate, the y-coordinate and the angle indicating the position of one production line as the three consecutive loci. When representing multiple production lines, these three loci are connected in series. Figure 3 shows the individuals in the case of three production lines. The x-coordinates, y-coordinates and angles are shown in Figure 4. In order to secure the movement path of the operator, we set the constraint that the width between each production line is an arbitrary value Z mm. Individuals that don't meet the constraint are deleted as lethal genes.

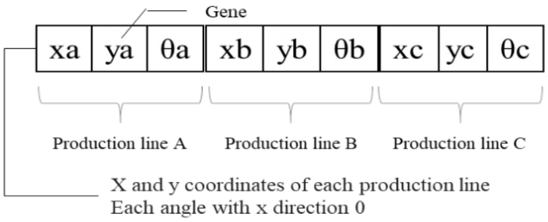


Fig. 3. Individuals expression

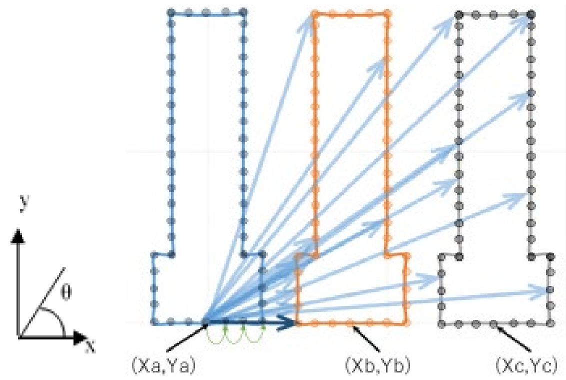


Fig. 4. How to judge lethal chromosome

5. Fitness evaluation values

PDCG adopts the walking time of the operator during the operating time as the evaluation values of fitness. If the walking time of individual i is t_i and the slowest walking time among the population of a generation is t_w, the fitness f_i of individual i is expressed by the following equation.

$$f_i = (t_w - t_i + 1)^2 \qquad (i = 1,2,3,\dots) \tag{1}$$

If the time difference between each individual and the slowest individual is taken as the fitness, the difference in fitness between an excellent individual and another individual becomes small. To make large difference, we give the difference a square. [1].

6. GA module algorithm

The GA module is the module to improve the efficiency of production line layout by GA. The algorithm of the GA module is shown below.

Step1: Generate the initial population.

Step2: Calculate the machine breakdown coordinates for each production line.

Step3: Calculate the walking time of the operator.

Step4: Calculate fitness.

Step5: Judge whether the end condition is satisfied or not, and if it is satisfied, go to Step 8. If not, go back to Step6.

Step6: Carry out selection, crossover and mutation operations.

Step7: Carry out Step 2 on the new generation's individuals.

Step8: The production line layout of the individual with the shortest walking time is adopted as the satisfying layout and is output as a solution.

We describe the details of the above algorithm.

Step1: Generate the initial population.

The individuals of the initial population are randomly selected from a range of coordinates that can be placed in the factory layout where the x- and y-coordinates and angles of the production line are applied. When there are multiple target production lines, the layout is randomly generated. There is a possibility that the production line layout may overlap or be too close. In this case, these layouts correspond to lethal genes. Not to include the lethal genes in the population, the following operations are carried out.

(1) Generate individuals. (2) As shown in Fig. 4, calculate the distance between the contour of a production line and the contours of other production lines. (3) If the calculated distance is less than Z mm, the individuals are deleted as a lethal gene. (4)Return to (1) and regenerate the individual.

Step2: Calculate the machine breakdown coordinates for each production line.

The coordinates whose machine gets breakdown in the factory layout change according to the coordinates of the production line layout. That is, since the coordinates of a machine breakdown occurrence for every generated individual are different, it is necessary to calculate each walking time. In PDCG, the coordinates of a machine breakdown occurrence of all individuals are calculated by the following operations.

(1) Search the work flow line data corresponding to the machine breakdown occurrence time in the machine breakdown data and find the coordinates of the machine breakdown that occurred in one production line. (2) Based on the found coordinates, calculate the machine breakdown occurrence coordinates for other all individuals. For example, if the coordinates and angles of one production line A of an individual generated in Step 1 are (x_a, y_a, θ_a) and the machine breakdown coordinates found in (1) are (x_e,y_e),the machine breakdown coordinates (X_e, Y_e) in production line A are expressed as the following equations.

$$X_e = x_a + x_e * sin\theta_a \qquad (e = 1,2,3,\dots) \qquad (2)$$

$$Y_e = y_a + y_e * cos\theta_a \qquad (e = 1,2,3,\dots) \qquad (3)$$

Step3: Calculate the walking time of the operator.

When a machine breakdown occurs, the operator moves from the waiting place to the place where the machine breakdown occurred to fix it. PDCG calculates the total walking time to get there and back.

Step4: Calculate fitness by using the equation (1).

Table.1 Mechanical breakdown example

Time	X coordinate	Y coordinate	Breakdown name A
411	3,513	4,046	Whether the blade is abnormal
565	3,858	2,046	Whether the blade is abnormal
890	4,128	881	On standby
1,156	2,651	3,468	Abnormal presence or absence of frame

Step5: Judge whether the end condition is satisfied or not. If it is satisfied, go to Step 8. If not, go back to Step6.

After the walking time of the operator is updated, if no better individual appears after certain generations, GA is finished.

Step6: Carry out selection, crossover and mutation operations.

Using the fitness calculated in Step 4 and using the fitness proportional method, select a set of the individuals as a crossover target. As the crossover, we adopt the two-point crossover. Mutations occur with a certain probability within the population.

Step7: Carry out Step 2 on the new generation's individuals.

Repeat steps 2 to 6 for the next generation population.

Step8: The production line layout of the individual with the shortest walking time is adopted as the satisfying layout and PDCG outputs it as a solution.

The coordinate data of the satisfying individual is output, and the layout is represented as an image based on it.

7. Simulations applications and results

PDCG was applied to the razor assembly line in the Kai Industries Co., Ltd. Since the shape of the razor assembly line is extremely complicated, we simplified it to the graphic model as shown in Fig. 5 and applied it to PDCG.

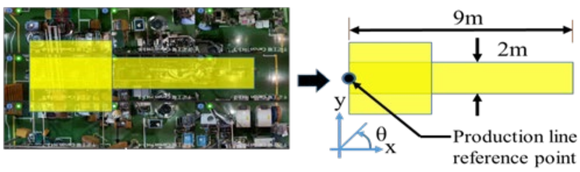


Fig. 5. Production line model

In this case, we acquired the work flow line data for 8 hours per one shift for 3 days with the chameleon code and also used the machine breakdown data of the same time. We adopt the GA parameters as follows. A mutation rate is 10%, a population size is 200 individuals, one individual is preserved as an elite. Selection and crossover methods were carried out by the fitness proportional method and two-point crossover. We also adopted b = 500mm and Z = 800mm.

The simulation results are shown below. The simulations were performed 20 times, and the layout obtained by PDCG is the solution. The layout is shown in Figure 6. The solution corresponds to the individual

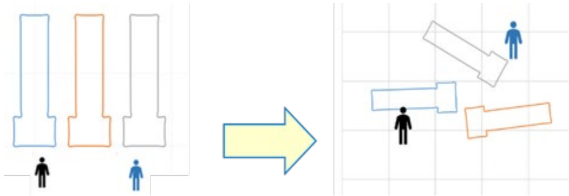


Fig. 6. Parallel layout and PDCG layout

with the minimum walking time. As a comparison, the walking time of the current layout of Kai Industries Co., Ltd. which is arranged in parallel was also calculated. The layout determined by PDCG shortened the operator walking time by about 63 minutes compared to the parallel layout. The layout determined by PDCG was effective in improving work efficiency.

Table. 2. Comparison of simulation results

	Walking time[sec]
PDCG layout	2,405
Parallel layout	6,164

8. Conclusions

After application simulations, it is ascertained that PDCG using the chameleon code of Infofarm Co., Ltd. is effective for the decision to find better layouts with multiple lines that one operator works.

9. Reference

(1) K. Honda, H. Yamamoto and T. Yamada, "Design System of Cell Type Assembly Machine with Dual Arms Robot by GA", The 2018 International Conference on Artificial Life and Robotics (ICAROB2018), Feb. 1-4, Beppu International Convention Center, Oita, Japan

Autonomous decentralized FMS's AGVs moving control

by mind change with deep learning

Ryunosuke Yamane
Department of Human Information Systems, Gifu University, Yanagido 1-1,
Gifu, Gifu 501-1194, JAPAN

Hidehiko Yamamoto
Department of Mechanical Engineering, Gifu University, Yanagido 1-1
Gifu, Gifu 501-1194, JAPAN

E-mail: y4525092@edu.gifu-u.ac.jp, yam-h@gifu-u.ac.jp, yamat@gifu-u.ac.jp

Abstract

This study describes the control method of Automated Guided Vehicles (AGV) movement by using a mind model in order to avoid AGVs interferences by using two types of mind , the arrogant mind and the modest mind model, the interferences can be avoided by repeating the two types of mind changes. This study develops the new mind model. The model includes deep learning. By the mind of the model of this new method, we can improve the decrease of the route interference time.

Keywords: Autonomous decentralized FMS, AGV with a mind, Deep learning

1. Introduction

We have developed the control of movements of Automated Guided Vehicles (AGVs) in an autonomous decentralized flexible manufacturing system (AD-FMS). The AGVs have a minimum model of mind (MUM) [1]. However, in the past research, if the number of AGVs operating in the factory increases, it is difficult to deal with efficient mind changes and it takes time to avoid route interferences. As a result, production efficiency decreases. In order to solve the above mentioned problems, we develop the new mind model including deep learning system. We call the model as Minimum Unit of Mind with Deep learning (MUMD). We compare the AGVs efficiencies of MUMD with the AGVs efficiencies of our developed two mind models, MUM, and MUMN by applying the three types of the mind models to FMS constructed in a computer. MUMN is including m a neural network into the MUM.

2. Autonomous Decentralized FMS

The example of AD-FMS is shown in Fig.1. The factory floor is divided into a grid pattern where an AGV moves. The system of AD-FMS is carrying the parts to the warehouse or machining centers (MCs). The AD-FMS does not have a management mechanism that controls the whole system. The agents (MCs, AGVs and Warehouses) which are the components of the AD-FMS exchange knowledge and determine the behavior of each agent.

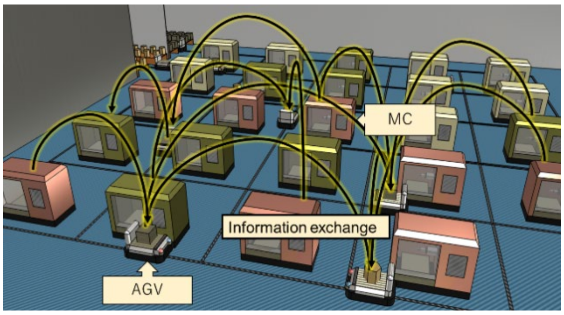


Fig.1 Example of AD-FMS

3. AGV behaviors and mind

AGVs autonomously decide their actions in the AD-FMS of this study. AGVs sometimes meat path interference in AD-FMS. To avoid the interferences, we adopted if-then rules in our past study. However, it was not possible flexibly to cope in the case of increasing the number of AGVs and changing the layouts of the factory. In order to solve this, we insert the MUM to AGV and realize autonomous movements without AGVs interferences.

3.1. *Mind model and its problem*

The AGVs mind expresses two kinds of mind, the arrogant mind that takes action forcebly to approach the destination and the modest mind that takes action to give ways to other AGVs. To express the mind, we use MUM in Fig.2. A1 and A2 are units, X are the weights and arrows are called stimulus vectors. Units have thresholds and internal values. A unit whose internal value reaches the threshold is called excited. A unit whose internal value does not reach the threshold is called normal. By the amount of the signal input to the unit, MUM changes the mind to arrogant or modest frequently. We describe the internal functions of MUM. When the arrogant AGV has a path interference, A1 is increased by 1. Keeping the situation of the interference and being increased by 1, the A1 value becomes the threshold, and the AGV is changed to a modest one. When AGVs with a modest mind keep giving a way, the value of A2 is increased by 1. When the situation is repeated at optional times, A2 becomes excited and a signal is sent to a load. The received load decreases the values of units A1 and A2 by optional values. Owing to this, A1 and A2 are returned to normal and a modest mind is changed to an arrogant mind. In this way, AGV avoids the path interference by the change of mind.

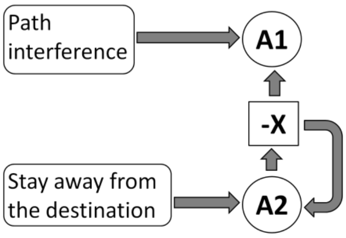


Fig.2 MUM model

The problem with MUM is that there are various patterns of path interferences in a large number of AGVs that move in AD-FMS. Because of it, it takes time to solve route interferences.

3.2. *MUMN*

MUMN is a model of the heart that was developed in order to achieve more efficient heart changes by using a neural network for the MUM and to change the heart according to the situation. The MUMN model is shown in Fig. 3.

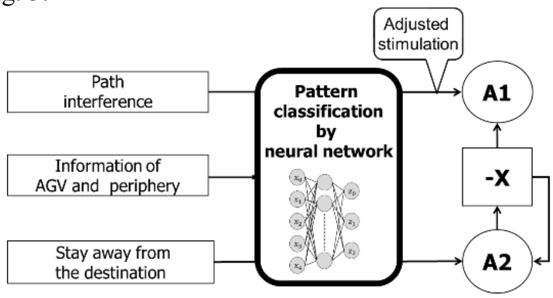


Fig.3 Model of MUMN

3.3. *MUMD*

We explain how MUMD affects the production in the factory by increasing the number of intermediate layers of MUMN explained in 3.2 to four layers. Fig.4 shows the MUMD that adjusts the amount of stimulus sent to A1 and A2 by a deep learning system after receiving an information such as how close or how far they are from the destinations and how many AGVs are in the surrounding areas. The model is shown in the diagram below.

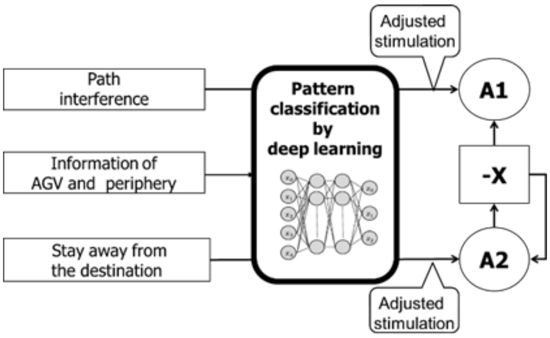


Fig.4 MUMD model

4. Inside information of MUMN

4.1. *Input information of deep learning*

The input information of the deep learning is the following five.
x_0: bias (always 1)
x_1: Destination
x_2: Distance from the destination
x_3: Number of other AGVs which are moving around the destination
x_4: Number of other AGVs numbers which go away from the destination

4.2. *How to create teacher signals*

x_1: The destination is the sum of the X and Y coordinates of the AGV destination.
x_2: The distance to the destination is the absolute value of the difference between the coordinates of the destination (X, Y) of the AGV and into current position (X ', Y'), as shown in equation (1).

Distance to destination = | X - X'| + | Y - Y' | (1)

x_3: The number of other AGVs within the range of bold line as shown in Fig. 5. We call the AGVs as neighbor AGVs.

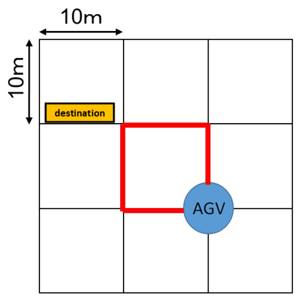


Fig.5 Other neighbor AGVs

x_4: The number of other AGVs which moves from the destinations. The AGVs take the bold line of Fig. 6 and we call the AGVs as not neighbor AGVs.

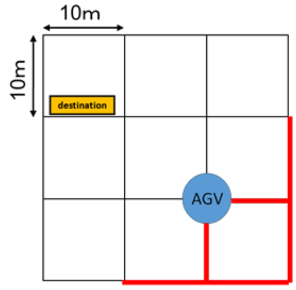


Fig.6 Not neighbor AGV

4.3. *Output information of deep learning*

The output information of the deep learning has the following three patterns. The input and output relationships are shown in Fig. 7.
z_0: Pattern to move to the destination with an arrogant mind ①
z_1: Pattern of giving way with a modest mind ②
z_2: neutral pattern ③

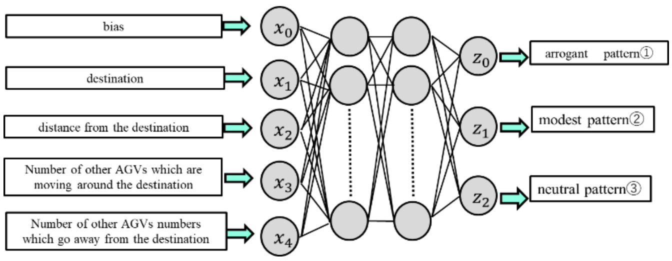


Fig.7 Internal information of MUMD

4.4. *Effect of MUMD*

The MUMD receives the output patterns and sends the stimulations, as shown in Table 1, to the units A1 and A2.

Table.1 Stimulation of each pattern

	A1	A2
arrogant pattern【1】	1	4
modest pattern【2】	4	1
neutral patter【3】	2	2

First, we explain the actions of each pattern in Table.1.
When the pattern 【1】 that moves with an arrogant mind is output, the stimulus to A1 gets reduced and the stimulus to A2 gets increased. Because of this, MUMD can keep an arrogant mind long.
Second, when the pattern 【2】 that moves with a modest mind is output, the stimulus to A1 gets increased and the stimulus to A2 gets reduced. Because of this, MUMD can keep a modest mind long.
Third, when pattern 【3】 is output, the amount of stimulus sent to A1 and A2 keeps equal. Because of this, MUMD can keep a neutral pattern between pattern 【1】 and pattern 【2】.

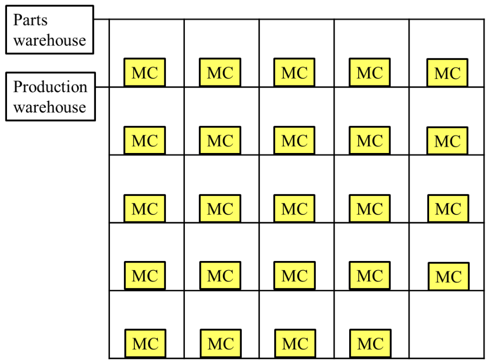


Fig.8 AD-FMS

5. Application Simulations and their results

By using the AD-FMS layout as shown in Fig. 8, the virtual production simulations with the three minds types, the conventional MUM, MUMN and MUMD, were carried out. Every simulation includes ten AGVs. The simulations for 8 hours AD-FMS operations were carried out 10 times each. As a result, Table 2 was acquired. It shows the comparisons of production outputs, the average operation rates of AGVs, the time of route interference for each AGV and the MC average operating rates.
From the table, the followings are found, MUMD improved the production outputs by 2.1%, AGV operation rate by 1.3% and MC operation rate by 1.9%, compared with MUM. MUMD decreased the production outputs by 0.9%, the AGV operation rate by 0.5% and the MC operation rate by 0.9%, compared with MUMN. MUMD reduced route interference time by 5.6%, compared with MUM. MUMD reduced route interference time by 3.0%, compared with MUMN. The reason MUMD is excellent is that MUMD can send appropriate stimulations to the AGVs minds and can give mind changes that strongly adapt the production situations. The production outputs and MC operating rate were not high. This is why the modest mind AGVs that need to avoid interferences increased and they had to choose detours for destinations.

6. Conclusions

From the simulation results, it was found that the route interference times of our proposed MUMD was shorter than those of MUMN. The study of our deep learning is to classify mind patterns by using 3 outputs and 4 inputs for each mind model. As our future study, we will develop AGVs that change their minds more flexibly by changing hyper parameters such as the numbers of input information, output information and the units of the middle layer.

References

1. Hidehiko YAMAMOTO, Takayoshi YAMADA and Shinsuke KATO, AGV Mind Model and its Usage for Decentralized Autonomous FMS by Change of Mind, Proceeding of Third KES International Symposium, KES-AMSTA 2009, Agent and Multi- Agent Systems : Technologies and Applications, Lecture Notes in Artificial Intelligence, Edited by Anne Hakansson, Ngoc Thanh Nguyen etc. , ISSN 0302-9743, Uppsala, Sweden, pp.744-753, June(2009), Springer.

Table.2 Simulation results

	outputs(num)	occupancy rate(%)	path interference time(sec)	occupancy rate of MC(%)
MUM	264.1	62.024	10880.2	35.581
MUMN	272.3	63.146	10613.9	36.586
MUMD	269.8	62.831	10306.8	36.244

Secondary School Robotics Education in Camarin High School: Developments and Challenges for Improvement

Jeffrey Rivera Galino
Camarin High School, Address Cadena de Amor Street, Area B, Camarin Caloocan City, Philippines
Graduate School of Humanity and Social Science, Hiroshima University, Address
1-1-1 Kagamiyama, Higashi-Hiroshima city, Hiroshima, 739-8524 Japan

Hideyuki Tanaka
Graduate School of Humanity and Social Science, Hiroshima University, Address
1-1-1 Kagamiyama, Higashi-Hiroshima city, Hiroshima, 739-8524 Japan

E-mail: tanakalpha@hiroshima-u.ac.jp

Abstract

Camarin High School is one of the schools in the Philippines that is making an initiative in improving STEM education through robotics education. In this paper, the authors report the development of a robotics program for students in junior high school. Through the Teacher Training Program of the Japanese Government for international teachers, localized challenges were identified and advances in technology education were examined for consideration. This report is hoped to provide useful insights to both practitioners of robotics education and initiators in a similar situation.

Keywords: Robotics Education, STEM Education, Robotics Club,

1. Introduction

Various research has substantiated positive effects of robotics in Science, Technology Engineering and Mathematics (STEM) education. According to Mosley, et al.[1] robotic cooperative learning metho-dologies promote STEM interest and enhance critical thinking of students. Kanlhofer and Steinbauer revealed the significant impact of educational robotics on pupil's mathematics and scientific investigation, teamwork and social skills. [2] A systematic review of Bareto and Benitti showed that the most common result on the topic is that the use of robotics helps the understanding of concepts related to the STEM areas.[3]

Robotics activities can be beneficial to students in various ways. Participation of the youth in robotics summer camps, academic year clubs and compe-titions increase their STEM content knowledge, their perceived problem solving skills, and their interest in engineering careers.[4] Robotics clubs positively change students perception of robots, humans and society while they increase skills in scientific creativity and science process skills.[5]

Various robotics competitions reported positive impacts on students. Participants of FIRST LEGO League, FIRST Tech Challenge, and FIRST Robo-tics Competition show significantly greater gains in STEM related measures such as interest, careers, identity, involvement and knowledge.[6] RoboCup Junior has presented success on enhancing learning of STEM contents and skills for innovation and creativity among students partaking the com-petition.[7] Using Test of Science Related Attitudes (TOSRA), it has been shown that students who participated in a robotics competition had positive attitude toward science and areas such as social implication of science, normality of scientists, attitude toward scientific inquiry and adoption of scientific attitudes.[8]

Robotics programs are seen to yield positive outcomes. First LEGO league robotics program conveyed prospects for learning 21st century skills such as systems thinking, decision making, problem solving, teamwork, conflict resolution, flexibility, perseverance and self-management.[9] Students acknowledged that learning skills in cooperation and communication as well as collaboration as one of the best results of an educational robotics course.[10]

Robotics programs can be done with various approaches. Instructional strategies such as modeling, coaching, scaffolding, examples and case studies provide fruitful experience to children.[9] A project-based robotics program can be beneficial to students in implementing informal instruction in science, technology and problem solving.[11] Challenge-based learning approaches in robotics are able to enhance student's course achievement and motivation.[12]

This paper focuses on the initiative and development of a school robotics program for a public general high school primarily enthused by the participation in science and technology fairs. Localized challenges and advances in technology education are explored for consideration.

2. Robotics in Camarin High School

Camarin High School is the largest public junior high school in Caloocan City and one of the largest schools in the Philippines in terms of population size. The school caters four grade levels – grade seven, grade eight, grade nine and grade ten with ages 13, 14, 15, and 16 respectively. As of 2019, it has 347 staff members and 9923 students divided into 224 sections.

Camarin High School has Special class sections comprised of students selected based on their elementary school grade and screening examination scores. The program started in 2017 with the first section in grade 7 rolling out as the years pass. As of 2019, there are 3 sections of special class, one for grade 7, grade 8 and grade 9. Aside from the regular core-subjects, additional Mathematics, English and Science classes are taken by the students of Special class. Students of the program are also required to do research projects during their grades 9 and 10.

National Science and Technology Fair (NSTF) is an annual science fair organized by the Department of Education of the Philippines that aims to promote Science and Technology and a culture of innovation among the youth. It also aims to identify the most creative and innovative student researchers from junior and senior high school who shall represent the country in the international science fairs. There are four categories, one of which is Robotics and Intelligent Machines category.[13]

In 2018, a robotics team was formed to satisfy the need of producing entries for science and technology fair competitions. Robotics team members were selected from grade 10 students based on their Science and Mathematics performance as well as their dedication to give time to the endeavor. Members received basic training on micro-controllers and proceeded to develop their own projects. Figure 1 shows the first robotics team.

Fig. 1. Training of robotics team members

Robotics club was also offered as an after-school activity to students interested in robotics. Since training was only given to members of the robotics team, a 3-day workshop was conducted to teach the students the fundamentals of robotics and micro-controllers.

In 2019, robotics class was offered as a non-credit elective subject for grade 7 and grade 8 students of Special class and an option for research for grade 9 special class. Figure 2 shows the robotics class.

Fig. 2. Robotics Class

The teacher-in-charge of the robotics team, robotics class and robotics club was selected among the existing faculty members based on exposure to related seminars and training. The teacher-in-charge (the first author) also attended available training in robotics prior to formation of the robotics team and start of the robotics class.

3. Developments in Robotics Education in Camarin High School

This section explains initial developments in robo-tics education at Camarin High school as well as the problems encountered during the development.

3.1 *Use of Arduino*.
Arduino was used to introduce the students to micro-controllers. Arduino is an open-source electronics platform based on easy-to-use hardware and software.[14] It is a low-cost alternative to popular robotics kits. An Arduino set was purchased for the robotics team. Sensors and other materials were bought according to need.

3.2 *Use of Simulations*
To cater more students in the workshop, Tinkercad was used. It is a free and easy to use application for 3D design, electronics and coding.[15] Students who can borrow and bring laptops to school were strategically grouped to form teams. Simulation is a way to reduce the need for more Arduino kits and to prevent unnecessary breakage of devices.

3.3 *Flipped Learning*
With a huge amount of resources available in the internet and video streaming platforms such as YouTube, robotics team was advised to utilize differentiated resources. Flipped learning is an approach in which students watch recorded videos for homework and tasks are completed when the teacher is present.[16] The use of flipped learning with the available online resources is a way for the teacher-in-charge to save time on instruction and focus more on solving problems, monitoring progress and giving feedback.

3.4 *Project-Based Learning*
Since the initial purpose of the robotics initiative is to provide entries to science and technology fairs, students studied robotics based on the needs in the project the groups are working on. Project-based learning is a teaching method in which students are engaged in complex real-world tasks that result in a product or a presentation (e.g. [17]). The use of project-based learning enabled the students to learn tools and techniques required for developing their projects in a short amount of time.

3.5 *Problems Encountered*
With the establishment of a local robotics program, several problems were encountered such as (1) topics and projects to be included in robotics program (class, team and club) and effective practices in teaching to students (2) further training of teacher-in-charge project development and supervision and (3) low-cost devices and instructional materials.

4. Challenges for Improvement

With the participation of the teacher-in-charge on Teacher Training Program of the Japanese Govern-ment for foreign teachers, several activities have been executed and expected to bring improvement in the robotics program of Camarin High school.

4.1 *Seminars*
Seminars on microcontrollers and mechanisms for the teacher-in-charge provided technical training and pedagogical examples in teaching robotics as well as the outline of topics for a classroom-based robotics program. Since the teacher-in-charge had no formal training on technology education, the second author provided several seminars in Hiroshima University. Topics are listed as follows:

(i) Software and Programming:
  - (a) C coding
  - (b) state transition diagrams.

(ii) Electronics:
  - (a) pull-up and pull-down registers
  - (b) transistors
  - (c) oscilloscope
  - (d) relay logic circuit

(iii) Hardware and microcomputer:
  - (a) digital input and output
  - (b) AD converter
  - (c) PWM
  - (d) real time interrupt
  - (e) DA converter
  - (f) motor driver
  - (g) bypass capacitor
  - (h) sensors

(iv) Mechanism
  - (a) linkages
  - (b) pulleys, gears and cams

(v) Teaching:
  - (a) connecting the electrical devices on the breadboard
  - (b) wiring layout diagram

The topics cover programming, microcomputers, electronics, mechanics and robotics and are selected based on practicality.

4.2 *Robotics Projects*
The first author has built various low-cost robotics projects during the teacher training to be used as common projects for project-based-learning in Camarin High School. The building process was also a coaching example in guiding students in creating their projects. The following projects were created:

(i) Tank type robot
(ii) Remote controlled robot
(iii) Edge avoidance and obstacle avoidance robot

Creating robotics projects integrate connected robotics topics. The tank-type robot project involves the basics of making robot bodies and mechanics. The remote-controlled robot project deals robotic control practices without using microcontrollers. The edge avoidance and obstacle avoidance robot project includes topics on sensors, actuators and micro-controllers.

4.3 Instruction Materials

Guidance on the use of Raspberry Pi is valuable to the teacher-in-charge in utilizing it as a low-cost computer for programming and project development. The authors explored the use of Raspberry Pi as an alternative to laptop computers in teaching Arduino. The teacher-in-charge is going to prepare teaching materials for robotics class.

5. Conclusions

This paper reported that the robotics program of Camarin High School consists of the robotics team, robotics class and robotics club, primarily enthused by the participation in science and technology fairs. In its initial phase, flipped learning and project-based learning was utilized while using an Arduino kit. Simulation was used to cater many students. It also reported that the teacher-in-charge (the first author) is in the teacher training program of the Japanese Government and that he has received seminars in Hiroshima University and created robotics projects for enhancing teaching skills in robot education. The teacher-in-charge is going to prepare instruction materials for robotics class.

Acknowledgements

The teacher training program of the first author was sponsored by the Japanese Government through the Ministry of Education, Culture, Sports, Science, and Technology.

References

1. P. Mosley, G. Ardito, and L. Scollins, Robotic Cooperative Learning Promotes Student STEM Interest. *American Journal of Engineering Edu-cation,* **7**(2), 2016, pp. 117-128.
2. M. Kandlhofer, and G. Steinbauer, Evaluating the impact of educational robotics on pupils' technical and social-skills and science related attitudes, *Robo-tics and Autonomous Systems*, **75**, 2016, pp. 679-685.
3. F. Bareto and V. Benitti, Exploring the educational potential of robotics in schools: A systematic review, *Computers and Education*, **58**(3), 2012, pp. 978-988
4. G. Nugent, B. Barker, N. Grandgennett and G. Welch, Robotics camps, clubs and competitions: Results from a US robotics project. *Robotics and Autonomous Systems,* **75** (Part B), 2016, pp. 686-691
5. B. Cavas, T. Kesercioglu, J. Holbrook, M. Rannikmae, E. Ozdogru1 and F. Gokler, The Effects of Robotics Club on the Students' Per-formance on Science Process & Scientific Creativity Skills and Perceptions on Robots, Human and Society, *Proceedings of 3rd International Workshop Teaching Robotics, Teaching with Robotics Integrating Robotics in School Curriculum*, 2012
6. A. Melchior, C. Burack, M. Hoover and Z. Haque, FIRST® Longitudinal Study: Findings at 72 Month Follow-Up., *Heller School for Social Policy and Management Brandeis University*, 2020
7. A. Eguchi, RoboCupJunior for promoting STEM education, 21st century skills, and technological advancement through robotics competition, *Robo-tics and Autonomous Systems,* **75** (Part B), 2016 pp. 692-699
8. A. G. Welch, Using the TOSRA to Assess High School Students' Attitudes toward Science after Competing In the FIRST Robotics Competition: An Exploratory Study, *Eurasia Journal of Mathematics, Science & Technology Education*, **6**(3), 2010, pp. 187-197
9. Y. Ma and D. C. Williams, The Potential of a First LEGO League Robotics Program in Teaching 21st Century Skills: An Exploratory Study, *Journal of Educational Technology Development and Ex-change*, **6**(2), 2013, pp 13-28
10. A. Eguchi, Educational Robotics for Promoting 21st Century Skills, *Journal of Automation, Mobile Robotics & Intelligent Systems,* **8**(1), 2014, pp. 5-11
11. M. Barak and Y. Zadok, Robotics projects and learning concepts in science, technology and problem solving, *International Journal of Techno-logy and Design Education* **19**, 2009, pp. 289-307.
12. M. Jou, C.-K. Hung, & S.-H. Lai, Application of Challenge Based Learning Approaches in Robotics Education, *International Journal of Technology and Engineering Education,* **7**(2), 2010, pp.17-20
13. DepEd Memorandum No. 113, s. 2019, retrieved from https://www.deped.gov.ph/wp-content/uploads/2019/09/DM_s2019_113.pdf
14. Arduino, https://www.arduino.cc/
15. Autodesk, https://www.tinkercad.com/
16. J. Bergmann and S. Sams, *Flip Your Classroom: Reach Every Student in Every Class Every Day*, ISTE: Washington DC, USA, 2012
17. C.-H. Chena, Y.-C. Yangb, Revisiting the effects of project-based learning on students' academic achievement: A meta-analysis investigating mode-rators, *Educational Research Review,* **26***,* 2019, pp. 71-81.

Deep-Learning Based Segmentation Algorithm for Defect Detection in Magnetic Particle Testing Images *

Akira Ueda, Huimin Lu, Tohru Kamiya†
Department of Mechanical and Control Engineering, Kyushu Institute of Technology, 1-1 Sensui-Cho, Tobata-ku, Kitakyushu-shi, Fukuoka, 804-8550, Japan‡

Abstract

Magnetic Particle Testing (MPT), also referred to as magnetic particle inspection, is a nondestructive examination (NDE) technique used to detect surface and slightly subsurface flaws in most ferromagnetic materials such as iron, nickel, and cobalt, and some of their alloys. In a bad environment, the procedure is complicated, and automation of MPT is strongly desired. To find defects in the formed magnetic powder pattern, it is required to be highly skilled and automation has been considered difficult. In recent years, many defect detection methods based on deep learning have been proposed, and the effectiveness of deep learning has been shown in the task of automatically detecting various types of defects having different shapes and sizes. In this paper, we describe the development of deep learning based segmentation algorithm for defect detection in MPT images. We have achieved a F2 score of 84.04% by using U-Net as the segmentation model and by utilizing a strong backbone network and an optimal loss function.

Keywords: Magnetic Particle Testing, Nondestructive Examination, Defect Detection, Segmentation.

1. Introduction

Magnetic Particle Testing (MPT) is used the leakage flux generated in cracks, holes, and internal defects when magnetizing a ferromagnetic material. By scattering the fluorescent magnetic powder while the test target is magnetized, the magnetic powder is attracted to the leakage magnetic flux generated in the defect. In that way, defects are highlighted to help visual confirmation.

Defects in MPT have various shapes such as circular, linear, etc. In addition, magnetic particles are also adsorbed on the irregularities other than defects, the residual magnetism, and the boundaries of the metal structure, and they have the property of forming "pseudo patterns". MPT provides very good defect resolution and is used extensively on: welded fabrications in magnetic material, castings, locating fatigue cracks in items subject to cyclical stress. However, cleaning and demagnetization are required when performing inspection, and development of automation technology is required to reduce work load and improve efficiency.

Regarding the identification of fatigue cracks of the target item, it is easy to check whether it is normal or abnormal, and it is possible to inspect it on the line. Therefore, automation has been realized by high-resolution cameras and image processing technologies such as binarization and thinning. However, in the field of welding processing of magnetic materials such as buildings and facilities, the surface condition is poor, and therefore many pseudo patterns occur. Moreover, since the shooting environment cannot be completely fixed, the difference in images is large. For the above reasons, image analysis has been difficult and automation has not been realized.

In recent years, research on anomaly detection areas by deep learning has made remarkable progress. There are many approaches such as binary classification of abnormal and normal, classification of abnormal data,

.

GAN[1] and VAE[2] that learns only normal data and detects abnormal data, detection of abnormal areas by object detection method and segmentation. They are highly accurate and have already been put to practical use in many fields. As described above, anomaly detection by deep learning is highly reliable. We believe that the effective use of deep learning can realize the automation of MPT, which has been considered difficult to detect defects.

In this paper, we discuss a deep learning based segmentation algorithm for the MPT automation. Based on U-Net, we apply an effective backbone network and loss function, and describe its effectiveness.

2. Datasets

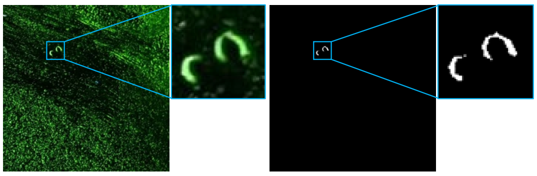
2.1. *Dataset*

In this article, we use images taken in the field inspection of the welding fabrications of magnetic materials. The image size is 4608 x 3456 pixels (138 images) and 4000 x 2256 pixels (35 images). However, the defect area is very small for the image, which is too large to apply the deep learning model. Therefore, the area including the defect is cropped, and the training and testing are executed. The procedure for creating a data set is as follows.

1. The original image is cropped at the size of 512 x 512 pixels including defects (total 173 images). An example of the generated images is shown in Fig.1 (a). When cropping, we consider not to have all the defects in the same position.
2. In the cropped image, the label images were created by annotating the defect area in detail using labeling technique. An example of annotation of the defect images is shown in Fig.1.

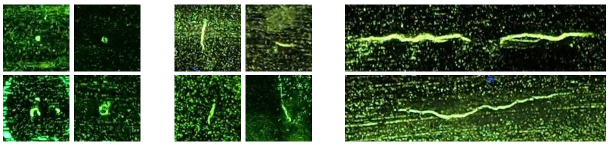
2.2. *Type of defects*

The dataset includes defects of various sizes and shapes, from 5 mm to 150 mm with circular, linear, and so on. Circular and near-circular defects with the size of 5 mm to 10 mm are the most common, followed by linear defects with the size of 20 mm to 50 mm. It also contains a small number of linear defects of about 150 mm and defects that do not fit any of them. Fig.2 shows an example of defects.



(a) (b)

Fig. 1. Examples of annotation of defect image. (a) An image including defects and (b) a label image are shown.



(a) (b) (c)

Fig. 2. Examples of defect types. (a) Circular and near circular defects, (b) linear defects, and (c) large linear defects.

3. Methods

We adopted U-Net[3] as the segmentation model. The detection score is improved by changing the encoder and applying the loss function considering the balance between classes. This section gives an overview of our approach.

3.1. *Models*

U-Net is a state-of-the-art semantic segmentation method with an encoder/decoder architecture. These encoder/decoder architectures use skip connections to preserve the location information lost due to convolutions, allowing for more precise output. Also, instead of using the original U-Net encoder, we use EfficientNet[4], which claims to be balanced between network depth, width, and resolution.

EfficientNet is available with different versions, starts from B0 at 5.3 million parameters to B7 at 66 million. Subsequent results show that a deeper encoder is not needed, so the subsequent experiments will be performed with a smaller encoder (EfficientNet B1).

We speed up the training process with pre-trained weights. Although it was trained on ImageNet, which is a database of natural images, the pre-trained weights do improve the training and local validation scores.

3.2. *The loss function*

Since the defects in MPT images are very small, training is greatly affected by the balance between the

Table. 1. Quantitative result of using different backbone network for U-Net encoder. In addition to the original encoder and EfficientNet B1 to B4, these show the results of experiments with ResNeXt50[5].

Models	Precision	Recall	F1	F2
U-Net	0.7147	0.7390	0.6980	0.7156
U-Net + EfficientNetB1	**0.8166**	**0.7929**	**0.7933**	**0.7904**
U-Net + EfficientNetB2	0.8085	0.7718	0.7777	0.7716
U-Net + EfficientNetB3	0.7886	0.7898	0.7712	0.7784
U-Net + EfficientNetB4	0.8143	0.7817	0.7865	0.7811
U-Net + ResNeXt50	0.7490	0.7500	0.7227	0.7362

background and defect classes. Therefore, it is desirable to use a loss function suitable for the target.

In the task of segmentation, Dice score coefficient (DSC) is commonly used. The 2-class DSC variant for class c is expressed in Equation 1, where g_ic ∈ {0,1} and p_ic ∈ {0,1} represent the ground truth label and the predicted label, respectively. The total number of pixels in an image is denoted by N. The ϵ provides numerical stability to prevent division by zero.

$$DSC_c = \frac{\sum_{i=1}^{N} p_{ic} g_{ic} + \epsilon}{\sum_{i=1}^{N} p_{ic} + g_{ic} + \epsilon} \tag{1}$$

The linear Dice loss (DL) is therefore defined as a minimization of the overlap between the prediction and ground truth:

$$DL_c = \sum_{c} 1 - DSC_c \tag{2}$$

One of the limitations of the Dice loss function is that it equally weighs false positive (FP) and false negative (FN) detections. In practice, this results in segmentation maps with high precision but low recall. With highly imbalanced data and small ROIs such as small defects, FN detections need to be weighted higher than FPs to improve recall rate. The Tversky similarity index is a generalization of the Dice score which allows for flexibility in balancing FP and FNs:

$$TI_c = \frac{\sum_{i=1}^{N} p_{ic} g_{ic} + \epsilon}{\sum_{i=1}^{N} p_{ic} g_{ic} + \alpha \sum_{i=1}^{N} p_{i\bar{c}} g_{ic} + \beta \sum_{i=1}^{N} p_{ic} g_{i\bar{c}} + \epsilon} \tag{3}$$

where, p_{ic} is the probability that pixel i is of the lesion class c and $p_{i\bar{c}}$ is the probability pixel i is of the non-lesion class, c¯. The same is true for gic and gic¯, respectively. Hyperparameters α and β can be tuned to shift the emphasis to improve recall in the case of large class imbalance. The Tversky index is adapted to a loss function (TL) in [6] by minimizing $\sum_c 1 - TI_c$.

4. Experiments and results

To detect a defect region on a sequential image, we mainly conducted two experiments.

1. Experiment to verify the effectiveness when using EfficientNet as U-Net encoder and to determine the optimal EfficientNet version.
2. Experiment to verify the effectiveness of Tversky loss compared to Dice loss. For Tversky loss, we search for optimal α and β values.

We used 80% of the dataset for training and 20% for testing. When partitioning the dataset, we took care to ensure that the defect types (circular, linear, large defects, etc.) are even. We trained for 500 epochs with batch size of 2 and evaluated using the weight that gave the highest score. Also, we used Adam as an optimization algorithm and trained with an initial learning rate of 0.001. We used precision and recall, F1 score, which is the harmonic mean of them, and F2 score, which emphasizes recall, as evaluation functions.

The results of each experiment are shown in Table.1 and Table.2. From the Table.1, by using EfficientNet as the encoder of U-Net, both the precision and recall improved. In addition, among the models used, EfficientNet B1 achieved the highest score in both precision and recall, and we did not see any improvement by increasing the

Table. 2. Quantitative results of using different loss functions (Dice loss, Tversky loss). We use EfficientNet B1 for Unet encoder (Efficient U-Net)

Models	Parameters	Precision	Recall	F1	F2
Efficient U-Net + DL	$\alpha = 0.5, \beta = 0.5$	**0.8166**	0.7929	**0.7933**	0.7904
Efficient U-Net + TL	$\alpha = 0.6, \beta = 0.4$	0.7099	0.8760	0.7725	0.8281
Efficient U-Net + TL	$\alpha = 0.7, \beta = 0.3$	0.7112	0.8922	0.7812	**0.8408**
Efficient U-Net + TL	$\alpha = 0.8, \beta = 0.2$	0.7086	0.8765	0.7720	0.8276
Efficient U-Net + TL	$\alpha = 0.9, \beta = 0.1$	0.6112	**0.9258**	0.7275	0.8305

version. It shows that deeper encoders are not needed in the segmentation of MPT defects. We also found that using EfficientNet as the backbone is more effective than ResNeXt50.

From Table.2, it was found that using Tversky loss instead of Dice loss improves Recall and F2 score. It is considered that the weighting of FN detection higher than that of FP led to improvement in recall. The highest F2 score was obtained when $\alpha = 0.7$ and $\beta = 0.3$. Fig.2 shows a comparison of output images when using Dice loss and Tversky loss. We obtained better output results when using Tversky loss.

5. Discussion and conclusions

We examined the defect segmentation method using the MPT defect dataset. It was verified that the EfficienNet used as the backbone of U-Net is effective and does not require a large model holding a large number of parameters. It was also shown that Tversky loss is effective in advancing the training of MPT defect dataset.

In this paper, we focused only on the areas containing defects, but we need to consider an approach to normal areas. Therefore, we created normal area images dataset from the original images described in Section 2. We prepared about 3000 normal images of 512 x 512 pixels. For those normal images, we tested using the segmentation model built in Section 3. Images that could be recognized as normal were not over-detected, and there was no false detection of pseudo-patterns or other noise that did not resemble defects. However, there were several cases in which the pseudo patterns similar to defects were erroneously detected. We need to improve them.

As the countermeasure, it is possible to include normal images in the training dataset. However, it is considered that the inclusion of normal images that has not been erroneously detected causes a decrease in detection accuracy. Therefore, we are mainly considering the following approaches.

i) Include normal images with remarkable pseudo patterns in the training dataset.

ii) Build a model to classify normal and abnormal and ensemble with our segmentation model. The approach to the normal area is important, and we would like to work on it in the future.

We developed a deep learning based segmentation algorithm for MPT defects. We achieved an F2 score of 84.08% by utilizing the backbone network and the loss function that facilitates optimization. In the future, we would like to consider the approach to the normal area.

References

1. Houssam Zenati et al., "Efficient GAN-Based Anomaly Detection", arXiv preprint arXiv: 1802.06222, 2018.
2. Yuchen Lu et al., "Anomaly Detection for Skin Disease Images Using Variational Autoencoder", arXiv preprint arXiv: 1807.01349, 2018.
3. Ronneberger et al., "U-Net: Convolutional networks for biomedical image segmentation", International Conference on Medical Image Computing and Computer-assisted Intervention, pp.234-241, 2015.
4. Mingxing Tan et al., "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", arXiv preprint arXiv: 1905.11946, 2019.
5. Saining Xie et al., "Residual Transformations for Deep Neural Networks", arXiv preprint arXiv: 1611.05431, 2016.
6. Seyed Sadegh Mohseri Salehi et al., "Tversky loss function for image segmentation using 3D fully convolutional deep networks", arXiv preprint arXiv: 1706.05721, 2017.
7. Nabila Abraham et al., "A Novel Focal Tversky loss function with improved Attention U-Net for lesion segmentation", arXiv preprint arXiv: 1810.07842, 2018.

A Method for Improving Recognition of Lying Postures Using a Measured Signal Intensity of Respiration and Heartbeat by Flexible Tactile Sensor Sheet

Kazuya Matsuo
Department of Mechanical and Control Engineering, Kyushu Institute of Technology, 1-1, Sensui, Tobata, Kitakyushu, Fukuoka, 804-8550, Japan

Toshiharu Mukai
Meijo University, Japan

Shijie Guo
Hebei University of Technology, China

E-mail: matsuo@cntl.kyutech.ac.jp

Abstract

We describe a method for improving recognition of lying postures using a measured signal ntensity of respiration and heartbeat. We can obtain respiration and heartbeat by means of using the body pressure measured at the suitable location determined by lying posture in the conventional method. Therefore, a recognition rate of lying postures and a measured signal intensity of respiration and heartbeat have positive correlation. In the experiments, we show that recognition of lying postures is improved by our method.

Keywords: Unconstrained measurement, Lying posture, Respiration, Heartbeat, ubber-based flexible tactile sensor sheet.

1. Introduction

Japan's population of people aged 65 and above increased by 7 million, that is, from 29 million to 36 million in the ten years between 2010 and 2020. Further, the aging rate is expected to increase from 23.0% to 29.1% [1], [2]. Seemingly, the number of people that require long-term care will substantially increase. However, the burden on individual care staff is expected to increase as there will be no corresponding increase in the number of care staff [3]. There is therefore an increasing need for the introduction of care robots in order to reduce the burden on care staff and further solve this problem.

The main areas of focus that have been chosen for the use of robot technology in care are transfer assistance, mobility support, toilet support, watching over dementia patients and bathing assistance [4]. Out of these, the ability to use sensors to watch over dementia patients would significantly help to reduce the burden on care staff. Moreover, there are many benefits of the ability to identify a patient's position in bed that are not limited to dementia patients. The patient positions
that must be identified include the patient being in or out of bed, sitting up or lying down, any danger of the patient falling out of bed or a patient lying on their back or on their front. Additionally, the ability to measure the depth or state of sleep is useful too. Increasing the quality of sleep by observing the state of sleep is effective for maintaining the patient's healthy sleep state.

The standard and reliable method for measuring sleep states is polysomnography (PSG) [5]. This technique measures brain waves, ECG, eye movement, chin muscle activity and pneumogram. Further, based on the R&K method [6], the technique classifies sleep state into six categories namely wake, REM sleep and non-REM sleep stages 14. The technique is beneficial as it measures sleep states with high accuracy. However, since measuring biosignals requires the attachment of large numbers of sensors to the body, the technique brings a sense of constriction and may impede natural sleep. Moreover, apart from being difficult to automate, recognition of sleep states from biosignals requires judgment by a laboratory technician. Consequently, there is need to develop a method that allows for easy estimation of sleep state.

Evidently, sleep state can be estimated from biosignals such as respiratory rate, heart rate, number of body movements and lying posture. These biosignals can be measured relatively easily. For instance, a strong correlation has been found between heart rate and sleep state [7]. The relationship between the heartbeat and the number of body movements is also strong [8], [9]. Additionally, respiratory disorders such as sleep apnea syndrome are known to be unlikely to manifest when sleeping in a lateral position [10]. Watanabe *et al.* used air pressure mattress-sensors [11] to measure heart rate, respiration, body motion, snoring and coughing and further developed a method for estimating sleep state based on rhythm irregularities such as increased respiration and heartbeat during REM sleep [12]. Other studies use infrared motion sensors [13] or cameras [14] and focus on the extent and frequency of the measured body motion to assess the depth of sleep.

We propose a method that uses a flexible tactile sensor sheet to measure lying posture [15], respiration and heartbeat [16]. This involves the use of rubber, flexible capacitive tactile sensors that we developed (Smart Rubber (SR) sensors [17]) to measure lying posture, respiration and the heartbeat of people lying on their side in bed and in a non-restrictive manner. First, we identify lying posture based on the tactile image formed by the body pressure distribution pattern. We then obtain the rate respiration and heartbeat by measuring the pressure of the body locations in line with the recognition results. Further, lying posture, respiration and heartbeat can be measured simultaneously by using high spatial and pressure resolution SR sensors and the measured pressure information appropriately. The method has additional benefits in that, at a few mm, it is a great deal thinner than an air mattress. Further, measurements can be performed while the subject's privacy is taken into consideration, which is different from when cameras are used. SR sensors are easy to use on a daily basis as they are cheaper than existing measuring devices and require no effort to install.

We want to estimate sleep state based on lying posture, respiration and heartbeat that have been measured using SR sensors. However, the main challenge is that these biosignals can occasionally be inaccurately measured. For instance, heart rate causes minimal pressure fluctuations. Therefore, it is likely to be hidden by respiratory and body motion fluctuations and noise, and simultaneously difficult to measure. Moreover, errors in identifying a lying posture may also occur following the differences in physique and ways of sleeping.

Errors are highly likely in recognizing lying postures as well as in measuring the respiration and heart rates our method uses information on pressure in the locations that conform to the type of lying posture. On the other hand, if the signal intensity for the measured respiration and heartbeat is high, then it can be assumed that the recognition of the lying posture is successful. In the current paper, we propose a method that can help improve the recognition of lying posture using measured respiration and heartbeat signal intensity. First, we confirm that the success or failure of the recognition of a lying posture in the method we previously developed are correlated with the success or failure of the measurement of a heartbeat. In addition to the probability obtained through machine learning from tactile images, we use signal intensity for respiration and heartbeat to propose a method to improve the performance of the recognition of a lying posture. We demonstrate the usefulness of the proposed method by conducting an experiment using real data. In conclusion, we summarize our findings.

2. Lying Posture Recognition Method Using Tactile Images and Respiration and Heartbeat Signal Intensity

2.1. *Respiration and heartbeat measuring method*

First, we describe the method to measure a person's respiration and heartbeat based on the pressure data that have been measured using a flexible tactile sensor sheet (Fig. 1) [16]. We first measure the body pressure of a person lying on their side using tactile sensors laid out on the bed. We then conduct a frequency analysis for the time series data for the body pressure measured by the sensor cells in appropriate locations such as the chest etc. We used a fast Fourier transform (FFT) for the frequency analysis [18]. We obtained the person's respiration and heartbeat by extracting the appropriate frequency band.

The person whose measurements were taken was resting on a bed. While resting, respiration and heartbeat are kept virtually constant. The frequency bands for respiration and heartbeat can be separated using a band pass filter as they do not overlap in rest. In a preliminary experiment, the extracted respiration and heartbeat frequency band are determined as follows. (1) The respiration frequency band was set at 0.1[Hz] - 0.4[Hz]. (2) The heart rate frequency band was set at 0.8[Hz] 1.6[Hz]. Respiration and heartbeat wave forms can be obtained by returning the frequency band to the signal of the time band. Moreover, the dominant frequency in the corresponding band after FFT should be obtained in order to obtain the respiration and heart rates alone.

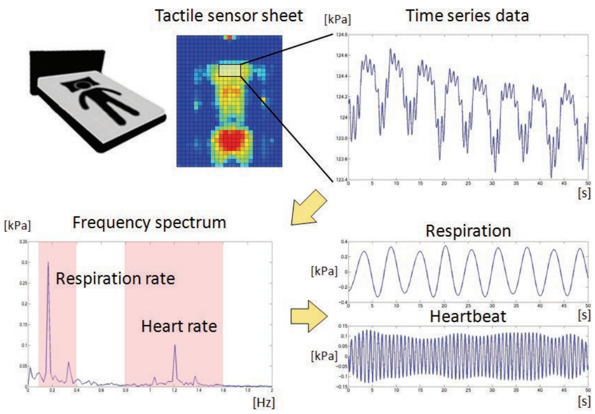


Fig. 1. Measurement method of respiration and heartbeat using a flexible tactile sensor sheet.

Furthermore, body pressure at the appropriate locations such as the chest must be measured in order to accurately measure respiration and heartbeat. A lying posture is detected using machine learning AdaBoost [19] while body pressure is measured in locations in accordance to the results (Fig. 2). This achieved robust measurements of respiration and heart rate in relation to changes in lying posture [15]. Fig. 2 shows the appropriate measuring locations where heartbeat peaks were seen in all of the 11 subjects that we took actual measurements for. For the current study, we used body pressure data of the locations where pressure was the highest, in the range shown in the ellipse in Fig. 2.

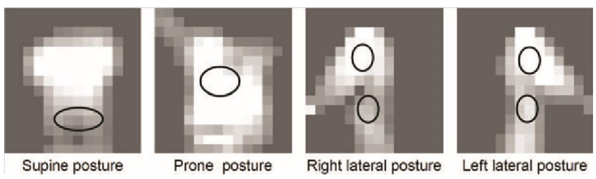


Fig. 2. Suitable locations for measurement of heartbeat. The circles show the suitable locations for different body postures. The suitable locations depend on not only the location but also the lying posture of the person [17].

Pressure fluctuations are higher for respiration than for heartbeat. Therefore, respiration can often be measured anywhere on the body and without any challenges. However, if lying posture recognition fails substantially and the pressure measuring location is placed outside the body, then it is patent that respiration cannot be measured. Fig. 3 illustrates such a case. The actual left side lying posture was erroneously detected as a supine posture. Therefore, the pressure measuring location is placed outside the body and the respiration measurement failed. The time series data shown in the top right corner of Fig. 3 reveal that the fluctuation range is extremely small and represents noise. Therefore, no peak respiration can be obtained by applying frequency analysis.

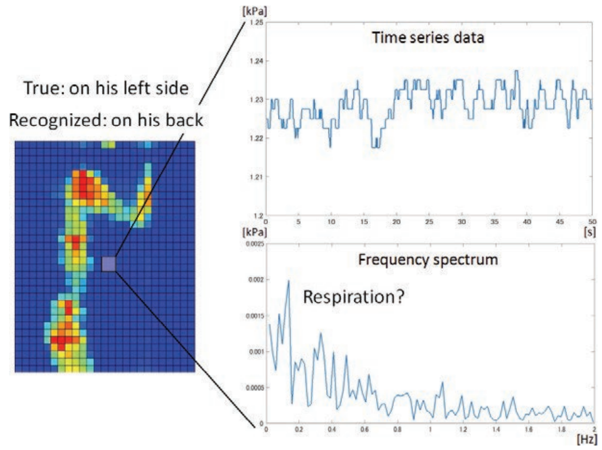


Fig. 3. Failure example of respiration measurement. We cannot obtain respiration by means of using the time series data measured at cells without body pressure.

2.2. *Improving Lying Posture Recognition Using Respiration and Heartbeat Signal Intensity*

Measuring respiration and heartbeat is more likely to fail if lying posture recognition fails. On the other hand, lying posture recognition is highly likely to be successful if signal intensity for the measured respiration and heartbeat is high. We therefore propose a method for the performance of improving lying posture recognition by using respiration and heartbeat signal intensity in addition to probability obtained from the tactile images. In specific terms, lying posture is obtained based on probability $q_i(\boldsymbol{x})$ in addition to probability $p_i(\boldsymbol{x})$ for each lying posture that is obtained through machine
learning from time series data $\boldsymbol{x}$ for the tactile images

$$q_i(\boldsymbol{x}) = \alpha \cdot r_i(\boldsymbol{x}) + \beta \cdot h_i(\boldsymbol{x}) + \gamma \cdot p_i(\boldsymbol{x}) \quad (1)$$

using respiration signal intensity $r_i(\boldsymbol{x})$ and heartbeat signal intensity $h_i(\boldsymbol{x})$ measured for each assumed lying posture. However, α, β, γ are weights. Eventually, they shall be $\alpha + \beta + \gamma = 1$ and $0 \leq \alpha,\beta,\gamma \leq 1$. Moreover, i is the number of lying posture. Further, when the lying posture is category N, probability from $q_1(\boldsymbol{x})$ to $q_N(\boldsymbol{x})$ can be obtained. In the current study, these were the maximum lying posture recognition results. Seemingly, $r_i(\boldsymbol{x})$ and $h_i(\boldsymbol{x})$, which represent signal intensity for respiration and heartbeat, are the SN ratio. This study's experiment used, SN ratios for respiration and heartbeat, respectively.

SN ratio for respiration $S\,N_{ri}$ and SN ratio for heartbeat $S\,N_{hi}$ are obtained based on the frequency analysis of time series data for body pressure locations in line with each of the assumed lying postures. These SN ratios were defined as:

$$S\,N_{ri} = 20\log_{10}\frac{S\,ri}{N_i}, \quad (2)$$

$$S\,N_{hi} = 20\log_{10}\frac{S\,hi}{N_i}. \quad (3)$$

S_{ri}, S_{hi}, and N_i represent the respiration, heartbeat, and noise signals respectively. The maximum spectra in the frequency bands for respiration (0.1[Hz]-0.4[Hz]) and heartbeat (0.8[Hz]-1.6[Hz]) were set to S_{ri} and S_{hi}, respectively. Moreover, N_i was set to the RMS of the signal in the part remaining after deducting the respiration, heartbeat and body movement signal bands from the overall frequency band. The respiration signal band is the range within 0.05[Hz] from frequency f_{ri} to S_{ri} and the second harmonic $2f_{ri}$ and the third harmonic $3f_{ri}$ for respiration. The heartbeat signal band is the range within 0.05 [Hz] from frequency f_{hi} to S_{hi} and the second harmonic $2f_{hi}$ for heartbeat. Further, we used a range of ± 0.05[Hz] from the peak as the signal band as the peak width of the signals for respiration and heartbeat in the preliminary experiment was 0.1 [Hz]. Moreover, we set the body movement band as 0[Hz] - 0.1[Hz].

3. Flexible Tactile Sensors

Sensors that are capable of high-resolution measurements of the body pressure of a person lying on a bed, in spatial terms as well as in terms of pressure are required in order to implement the proposed method. We used SR sensors [19], which are tactile sheet sensors. SR sensors, including the wiring, are entirely made of rubber. This makes them flexible and pliable and allow for measurements that avoid any sense of discomfort. Further, covering them with a sheet for instance does not result in any issues as it is possible to take measurements without the need to directly touch them.

This means that using SR sensors that are laid out on a bed allow for measurements that are non-restrictive, non-invasive and that can be conducted in circumstances that are similar to natural sleep.

SR sensors consist of a thin dielectric layer between two conductive rubber sheets of electrodes (Fig. 4). The black part of the rubber sheet are electrodes made from conductive rubber that was made using carbon filler, and are further laid out perpendicular to each other and on two layers. The total number consists of $m \times n$ cells when the number of electrodes on both sides is m and n. This is because capacitive sensors occur where sensors on both sides intersect. Spatially high-resolution measurements are achieved by narrowing the electrodes. Further, high resolution measurements in terms of pressure can be achieved by increasing the cell surface area.

When pressure is applied, the dielectric layer deforms hence reducing the distance between the electrodes. This increases the capacitance of the said area. Scanning the capacitance of the respective cells using the sensor controller allows for the recognition of the 2D pressure pattern. In specific terms, a sinusoid voltage is applied to each cell. Impedance is further obtained from the current going through thus obtaining capacitance. The effect of different wire resistance values for each cell is eliminated because for equivalent circuits at the time of scanning, wire resistance is connected in a series of cell condensers. This is a sensor composition that has been proposed in the past. Further, sensors with metal electrodes are commercially available. However, for applications in care aids such as sleep monitoring systems, issues such as lack of pliability and high cost remained. We therefore made use of rubber printing technology to develop flexible and low cost SR sensors. These sensors are sold by Sumitomo Riko Co., Ltd under the name SR Soft Vision.

Table 1 outlines the specification of the SR sensors used in this study. This electrode width allows for the measurement of body pressure with sufficient spatial resolution to enable the recognition of different lying postures. On the other hand, reducing the cell area would reduce the resolution in terms of body pressure. These sensors can measure respiration without any challenges as changes in body pressure are substantial. However, because these changes are small in relation to heartbeat, respiration cannot be measured as is, due to noise.

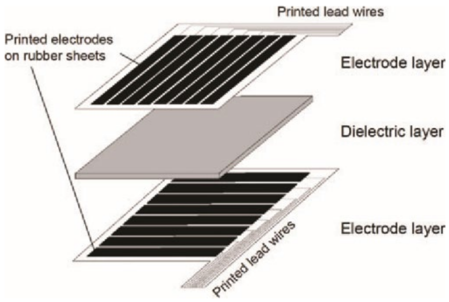


Fig. 4. Schematic structure of SR sensor.

Noise must therefore be suppressed and resolution in terms of pressure increased so as to enable the measurement of heartbeat.

Table 1. Specifications of the SR sensor.

Number of cells	800	
	Length	Width
	32	25
Size of sensor sheet	882 [mm]	686 [mm]
Size of a sensor cell	14 [mm]	14 [mm]
Size of a gap between cells	14 [mm]	14 [mm]
Thickness of sensor sheet	3.5 [mm]	
Sampling rate of all the 800 cells	5 [Hz]	
Sampling rate of the 4 precision cells	20 [Hz]	

4. Lying Posture Recognition Experiment

4.1. Experimental method

To verify the usefulness of the proposed method, we conducted an experiment to detect the lying posture of a person lying on their side on top of SR sensors laid out on a bed. The subject's positions were divided into 4 namely facing up, facing down, lying on their right side and lying on their left side. Each of these lying postures were taken 10 times. Further, time series data for body pressure were obtained for a total of 40 times. For each of these 40 data sets, probability $q_i(\boldsymbol{x})$, as defined in Section II-B, was obtained and lying postures detected.

First, we used the method by Mukai *et al.* [17] to obtain probability $p_i(\boldsymbol{x})$ for each lying posture based on the body pressure data. The length of one body pressure data set was set to 51.2 seconds (51.2 x 20 [Hz] = 1024) because conducting FFT requires that the number of data sets be a power of 2. The subject maintains the same lying posture for the full 51.2 seconds. Therefore, the obtained probability is virtually constant. The average for the 51.2 seconds was therefore set as $p_i(\boldsymbol{x})$. The 4 high-accuracy cells were then placed in the respective suitable locations for measuring the heartbeat for all lying postures, as shown in Fig. 2. Further, the signal intensity for respiration and heartbeat ($r_i(\boldsymbol{x})$ and $h_i(\boldsymbol{x})$: i = 1,2,3,4) was obtained from the time series data for body pressure. We used the SN ratio as the signal intensity for this study. Finally, $q_i(\boldsymbol{x})$ is calculated from these and lying posture recognition is performed. The present study was approved by the Research Ethics Committee of the Kyushu Institute of Technology. Further, Informed Consent was obtained in a writing from each subject taking part in the experiment.

If the signal intensity for respiration and heartbeat during the true lying posture is stronger than others, the final recognition result can be accurately corrected based on $q_i(\boldsymbol{x})$ by setting the appropriate weight. This can be done despite an error existing in the posture recognition using probability $p_i(\boldsymbol{x})$ from the tactile image only. We further conducted the following experiment to evaluate the performance of the proposed method.

Forty data sets were obtained for one subject, a 34-yearold male. These data sets were labeled A1. Recognition rates were examined for various weights (α,β,γ) for data set A1. A confusion matrix of the 4 lying postures was also studied.

4.2. Experimental results and discussion

Fig. 6 shows the recognition rates with weights varying between 0 and 1 at 0.1 intervals. The horizontal and vertical axes represent weight α and β respectively. The colors represent the recognition rates corresponding to the weight indicated by the bottom left coordinates. In conditions where $\alpha + \beta + \gamma = 1$ and $0 \le \alpha, \beta, \gamma \le 1$, values with the respective weights are $\alpha = 0, 0.1, \cdots, 1$, $\beta = 0, 0.1, \cdots, 1-\alpha$, $\gamma = 1 - \alpha - \beta$. Fig. 6 does not display weight γ as γ is determined when α and β are decided upon.

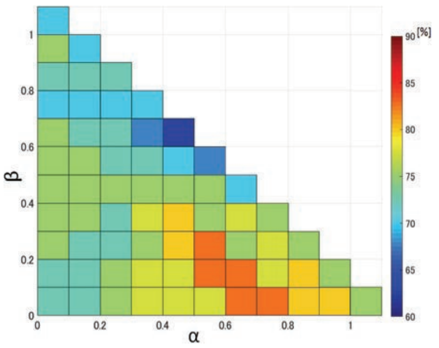


Fig. 6. Recognition rates of the lying postures (Dataset A1). Recognition of lying postures is improved by means of using a measured signal intensity of respiration and heartbeat. For example, the recognition rate of (α,β,γ) = (0.6,0.1,0.3) is higher than that of (α,β,γ) = (0,0,1).

Fig. 6 shows the recognition rate is the highest at 82.5% when (α,β,γ) = (0.6,0.1,0.3) or close to that. In detail, weights are the highest for (α, β, γ) = (0.5,0.1,0.4), (0.5,0.2,0.3), (0.6,0,0.4), (0.6,0.1,0.3), (0.7,0,0.3).
The recognition rate is 72.5% when (α, β, γ) = (0,0,1). The recognition rate for the proposed method with weight set at e.g. (α,β,γ) = (0.6,0.1,0.3), is 10% higher than for conventional methods that use only tactile images.

Moreover, Table 2 shows the averages and standard errors for the parameters obtained from the 40 data sets rather referred to as A1. These are $p_i(\boldsymbol{x})$, $r_i(\boldsymbol{x})$, $h_i(\boldsymbol{x})$, $q_i(\boldsymbol{x})$ when weight is (α,β,γ) = (0.6,0.1,0.3), with the top and bottom rows showing the parameters when lying posture is true and false respectively. A false lying posture is the posture with the highest value out of the 3 false postures. All the parameters are higher for the true lying positions. Further, it is clear that they are appropriate to use for recognition. Furthermore, the ratios for $p_i(\boldsymbol{x})$ and $q_i(\boldsymbol{x})$ for true and false lying postures are shown in rows 7 and 8, respectively. The higher the ratio, the higher the probability for the true lying posture in relation to the probability for other lying postures. Furthermore, stable recognition is possible. We further performed a paired t-test with a significance level of 0.05 in order to examine and confirm whether there is a difference between these two ratios. The null hypothesis states that the ratio between $p_i(\boldsymbol{x})$ and $q_i(\boldsymbol{x})$ does not differ. Furthermore, the degree of freedom is 39. This revealed that t-value $t = -3.27$, $p = 2.3 \times 10^{-3}$, and that there was no significant difference. This confirmed that in addition to the probability obtained from the tactile image, lying posture recognition can be improved by the appropriate use of signal intensity for respiration and heartbeat.

Table 2. Parameters of dataset A1 with (α,β,γ) = (0.6,0.1,0.3): $p_i(\boldsymbol{x})$, $q_i(\boldsymbol{x})$, $p_T(\boldsymbol{x})/p_F(\boldsymbol{x})$, and $q_T(\boldsymbol{x})/q_F(\boldsymbol{x})$. Recognition performance using $q_i(\boldsymbol{x})$ is improved than that using $pi(\boldsymbol{x})$.

Lying posture	$p_i(\boldsymbol{x})$	$q_i(\boldsymbol{x})$	$p_T(\boldsymbol{x})/p_F(\boldsymbol{x})$	$q_T(\boldsymbol{x})/q_F(\boldsymbol{x})$
True	50.1 ± 15.7	40.2 ± 5.8	2.09 ± 1.30	2.42 ± 1.04
False	29.7 19.(	19.8 11.(		

5. Conclusion

We proposed a method to improve the performance of lying posture recognition by using measured respiration and heartbeat signal intensity. We conducted experiments using actual data and demonstrated the fact that recognition performance can be improved by using respiration and heartbeat signal intensity along with the probability obtained from the tactile images. Moreover, we used data from different trials with the same subject and data for different subjects to confirm the versatility of the proposed method.

6. References

[1] Statistics Bureau, Ministry of Internal Affairs and Communications, Japan. (2011, Oct.). The 2010 Population Census of Japan. Public announcement on 26 Oct. 2011. [Online]. Available: <http://www.stat.go.jp/english/data/kokusei/index.htm>, (accessed on 3rd Dec. 2020).

[2] National Institute of Population and Social Security Research. (2012, Mar.). Population Projections for Japan (January 2012) 2011 to 2060-. Public announcement on 30 Mar. 2012. [Online]. Available: <http://www.ipss.go.jp/site-ad/index-english/ esuikei/gh2401e.asp>, (accessed on 3rd Dec. 2020).

[3] Ministry of Health, Labour and Welfare, Japan. (2011, June). General Welfare and Labour. Public announcement on 2 June 2011. [Online]. Available: <http://www.mhlw.go.jp/english/policy/ other/social-security/index.html>, (accessed on 3rd Dec.2020).

[4] Ministry of Economy, Trade and Industry and Ministry of Health, Labor and Welfare, Japan. (2012, Nov.). Priority Areas to Which Robot
Technology Is to Be Introduced in Nursing Care of the Elderly. Public announcement on 22 Nov. 2012. [Online]. Available:
<http://robotcare.jp/en/home/index.php?\&lang=en\ &PHPS

ESSID=q34fa48pe9plr3go0oo9vndmon>, (accessed on 3rd Dec. 2020).

[5] C. A. Kushida, M. R. Littner, T. Morgenthaler, C. A. Alessi, D. Bailey,
J. Coleman Jr, L. Friedman, M. Hirshkowitz, S. Kapen, M. Kramer,
T. Lee-Chiong, D. L. Loube, J. Owens, J. P. Pancer, and M. Wise, "Practice parameters for the indications for polysomnography and related procedures: An update for 2005," *Journal of Sleep*, vol. 28, no. 4, pp. 499–521, 2005.

[6] A. Rechtschaffen and A. Kales, *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Bethesda, MD. : U.S. Dept. of Health, Education, and Welfare, Public Health Services-National Institutes of Health, National Institute of Neurological Diseases and Blindness, Neurological Information Network, 1968.

[7] T. Watanabe and K. Watanabe, "Estimation of the Sleep Stages by the Non-restrictive Air Mattress Sensor: Relation bitween the Change in the Heart Rate and Sleep Stages," *Trans. of the Society of Instrument and Control Engineers*, vol. 37, no. 9, pp. 821–828, 2001.

[8] A. Deguchi and H. Hagiwara, "Evaluation of Heart Rate Variability and Respiratory Variability during Sleep Using a Lorenz Plot," *Journal of Sleep and Biological Rhythms*, vol. 9, no. 4, p. 366, 2011.

[9] K. Otsuka, Y. Ichimaru, T. Yanaga, and Y. Sato, "Studies of Arrhythmias by 24-hour Polygraphic Recordings: Relationship between Atrioventricular Block and Sleep States", *American Heart Journal*, vol. 105, no. 6, pp. 934–940, 1983.

[10] Z. Beattie, C. Hagen, and T. Hayes, "Classification of Lying Position Using Load Cells under the Bed", in *Proc. 33rd Annual Int'l Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 474–477, 2011.

[11] Y. Kurihara and K. Watanabe, "Sleep-Stage Decision Algorithm by Using Heartbeat and Body-Movement Signals," *IEEE Trans. on Systems, Man, and Cybernetics, Part A*, vol. 42, no. 6, pp. 1450–1459, 2012.

[12] T. Watanabe and K. Watanabe, "Noncontact Method for Sleep Stage Estimation," *IEEE Trans. on Biomedical Engineering*, vol. 51, no. 10, pp. 1735–1748, 2004.

[13] Y. Kambayashi and H. Hagiwara, "Estimating Sleep Cycle Using Body Movement Density," in *Proc. 2012 5th Int'l Conf. on Biomedical Engineering and Informatics*, pp. 1029–1033, 2012.

[14] S. Okada, Y. Ohno, Goyahan, K. Kato-Nishimura, I. Mohri, and M. Taniike, "Examination of Non-restrictive and Non-invasive Sleep Evaluation Technique for Children Using Difference Images," in *Proc. 30th Annual Int'l Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 2483–2487, 2008.

[15] T. Mukai, K. Matsuo, Y. Kato, A. Shimizu, and S. Guo, "Lying Posture Detection for Unconstrained Measurement of Respiration and Heartbeat of a Person on a Bed," in *Proc. 11th IASTED Int'l Conf. on Biomedical Engineering*, pp. 46–53, DOI: 10.2316/P.2014.818-014, 2014.

[16] K. Matsuo, T. Mukai, Y. Kato, A. Shimizu, and S. Guo, "Measurement of Respiration and Heartbeat Using a Flexible Tactile Sensor Sheet on a Bed," in *Proc. 11th IASTED Int'l Conf. on Biomedical Engineering*, pp. 59–64, DOI: 10.2316/P.2014.818-031, 2014.

[17] S. Guo, Y. Kato, H. Ito, and T. Mukai, "Development of Rubber-Based Flexible Sensor Sheet for Care-Related Apparatus," *Sumitomo Electric Industries Technical Review*, no. 75, pp. 125–131, 2012.

[18] J. W. Cooley and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series," *Mathematics of Computation*, vol. 19, no. 90, pp. 297–301, 1965.

[19] Y. Freund and R. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting", *Journal of Computer and System Science*, vol. 55, pp. 119–139, 1997.

[20]

Basic research for the realization of online MEG using SSD

Kazuhiro Yagi
Interdisciplinary Graduate School of Agriculture and Engineering, University of Miyazaki,
1-1, Gakuen Kibanadai-Nishi, Miyazaki, 889-2192, Japan

Yuta Shibahara
Graduate School of Engineering, University of Miyazaki,
1-1, Gakuen Kibanadai-Nishi, Miyazaki, 889-2192, Japan

Lindsey Tate
Faculty of Engineering, University of Miyazaki,
1-1, Gakuen Kibanadai-Nishi, Miyazaki, 889-2192, Japan

Keiko Sakurai
Faculty of 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: kazuhiro_yagi@junwakai.com, hi16019@student.miyazaki-u.ac.jp, teitorinzeirini.c4@cc.miyazaki-u.ac.jp, sakurai.keiko.u6@cc.miyazaki-u.ac.jp, htamura@cc.miyazaki-u.ac.jp

Abstract

Neurofeedback systems have been found to be effective in the clinical rehabilitation of paralysis. However, most systems exist only for use with EEG, which is cumbersome to apply to patients and has lower spatial resolution than MEG. Furthermore, the best practices for neural data feature extraction and feature selection are not well established. The inclusion of the best performing feature extraction algorithms is critical to the development of clinical neurofeedback systems. Using simultaneously collected MEG and accelerometer data before and during 10 spontaneous finger movements, we performed an in-depth comparison of independent components analysis (ICA) and spatio-spectral decomposition (SSD) algorithms for their individual abilities to isolate movement-relevant features in brain activity. Having restricted raw data to that from sensorimotor rhythm (SMR) frequencies in select MEG sensors over sensorimotor cortex, we compared ICA and SSD components using: (1) 2D topographies, (2) activations over time, (3) and correlations with accelerometer data at both 0ms and 60ms time delays. SSD performed more quickly and produced components that were more highly correlated with the behavioral data than ICA. We will discuss these results and suggestions for application to neurofeedback systems. In particular, we will present detailed visualizations of SSD results and discuss potential strategies and pitfalls for feature selection.

Keywords: Magnetoencephalography, Spatio-spectral decomposition, Morlet wavelet transform, Neurofeedback

1. Introduction

For diseases that affect brain function, such as strokes, immediate treatment with medication and surgery is important, but post-onset rehabilitation also plays a critical role in the wellbeing of patients.

One of the techniques used for non-invasive brain function evaluation is magnetoencephalography (MEG)[1-

[2]. MEG has high temporal resolution as well as high spatial resolution, and it is commonly used clinically for epilepsy diagnosis and rehabilitation[10]. Without the ability to monitor a patient's relevant brain activity in real time during rehabilitative exercises, efficient rehabilitation cannot occur.

The purpose of this study was to evaluate the use of spatio-spectral decomposition (SSD) of real-time MEG data during spontaneous movement[3-6]. Performance was evaluated by comparing SSD results to the results from a standard analysis technique, independent components analysis (ICA). SSD completed decomposition faster than did ICA, with the SSD analysis completed about 270 times faster in the preliminary experiment. In addition, as shown in Table 1, the correlation with the accelerometer data (shifted by 60[ms] to account for the time difference between neural motor planning and actual motor execution) was also stronger for the most highly correlated SSD component as compared to the most highly correlated ICA component. Our results indicated that SSD outperforms ICA in the context of feature extraction for online, real-time MEG analysis; therefore, we present our investigation of the SSD components and conclude with suggestions for feature selection.

Table 1. The correlation with the accelerometer data and ICA or SSD.

	60[ms]
1st ICA	0.71
2nd ICA	-0.18
3rd ICA	-0.55
4th ICA	-0.07
5th ICA	0.35
1st SSD	0.15
2nd SSD	-0.89
3rd SSD	-0.13
4th SSD	0.47
5th SSD	-0.45

2. Experiment

The data were collected using a full-head 306-channel magnetoencephalograph (Vectorview, Elekta-Neuromag, Helsinki, Finland) at a sampling frequency of 1000 [Hz]. In order to reduce power supply noise and other interference, measurements were taken inside a magnetic field shield room (1 [kHz] shielding rate 55.2 [dB]). The participant was a healthy person who attached an acceleration sensor to the middle finger of his left hand and performed ten spontaneous flexion (i.e., bending) and extension (i.e., relaxing) movements of the indicated finger as shown in Fig.1 (hereinafter referred to as “the task”). The accelerometer data indicated the start and duration of flexion and extension.

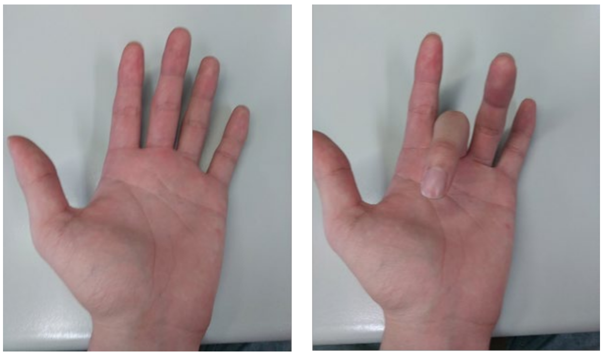


Fig.1: The spontaneous flexion (bending) and extension (relaxing) movements.

3. Analysis Method

After collecting the raw MEG data from all 306 channels, we sub-selected 26 gradiometers corresponding to the right sensorimotor cortex (SMC) and performed ICA and SSD analyses[7-8], which each identified components within the largest cited frequency band for sensorimotor rhythm (SMR)[9], 8-30 [Hz]. As presented in Table 1, we calculated Pearson correlations between the time-adjusted accelerometer data and each of the top five components from ICA and SSD. We calculated the SSD topographies about 20 seconds and then performed a Morlet wavelet transform to examine frequency power over time (Fig.2). In addition, in order to increase the signal-to-noise ratio (SNR), the wavelet analysis was performed on task-locked averaged components to indicate average brain activity in the final 2 seconds before flexion start. There were 10 tasks completed over 20 seconds at irregular intervals.

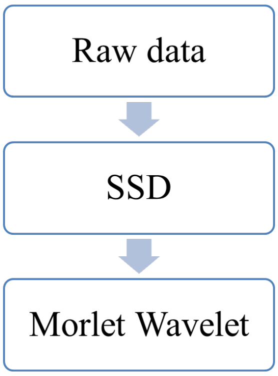


Fig.2: The flow of analysis method

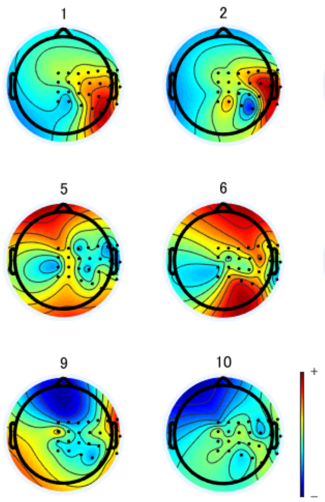
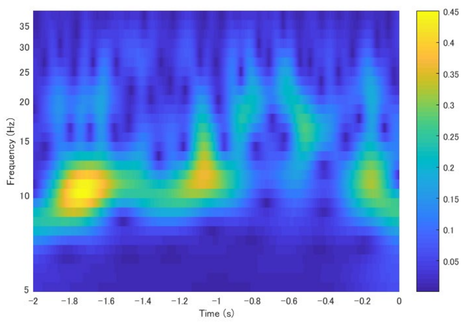
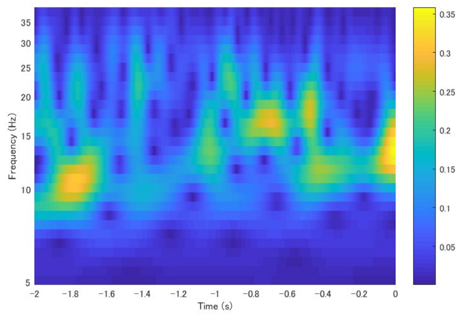


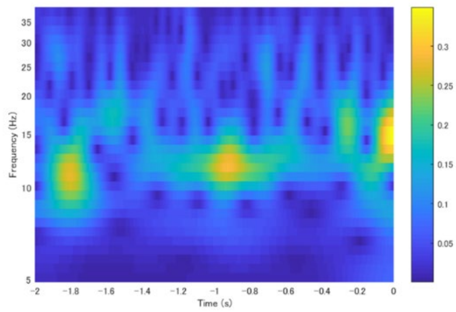
Fig.3: The results of Topography using SSD



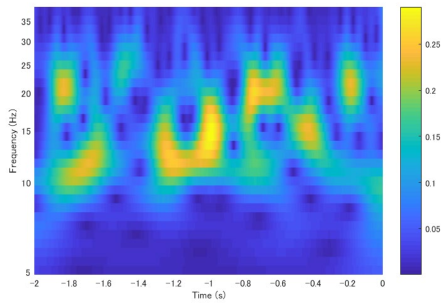
2nd SSD



3rd SSD



4th SSD



5th SSD

Fig.4: The results of the across-task averaged second, third, fourth and fifth SSD components

4. Results

From the results of topography (Fig. 3), the first eight SSD components showed activity localized around the right SMC. As shown in Fig. 4, the across-task averaged third and fourth SSD components had higher SMR band power immediately preceding the task. The pattern of activity observed in the SMR frequency band across task instances indicated that the first four SSD components captured artifactual activity (e.g., movement activity). In the averaged component as well as across task instances, steady activity during the two seconds preceding flexion was observed at 12-13 [Hz] in the fifth SSD component.

5. Conclusion

In this paper, we performed offline MEG data analysis during flexion and extension of the middle finger of the left hand in order to investigate the use of SSD in online, real-time MEG analysis.

There are several key points to take away regarding the automation of feature selection in this context. First, the SSD component with the highest correlation to accelerometer data (comp. 2, $r = -0.89$) was, in this case, indicative of muscular noise reaching the MEG sensors rather than indicating neural activity as desired. Second, the component most likely to indicate relevant neural activity (comp. 5) was, in addition to not being the most highly correlated with accelerometer data ($r = -0.45$), also not the first component (i.e., it was not the component with the strongest eigenvalue or largest SNR). Finally, when only considering a few trials (i.e., at the beginning of neurofeedback rehabilitation) or particularly noisy data, using averaging techniques can lead to incorrect feature selection due to the extreme amplitude of artifacts and the components that capture them. Therefore, a component to use for patient feedback cannot be selected based solely on having the highest average SMR power, the highest correlation with behavioral data, or the highest eigenvalue.

Future study will include the same analysis for SSD components 6-8, which show promising topographies. Furthermore, we will investigate the usefulness of baseline correction in component selection and of logarithmically scaled components as features in unsupervised learning.

References

1. Franca Tecchio, Filippo Zappasodi, Mario Tombini, Antomio Oliviero, Patrizio Pasqualetti, Fabrizio Vernieri, Matilde Ercolani, Vittorio Pizzella, and Paolo Maria Rossini, Brain plasticity in recovery from stroke: An MEG assessment, NeuroImage 32 (2006) 1326 – 1334.
2. Andrea Paggiaro , Niels Birbaumer , Marianna Cavinato , Cristina Turco, Emanuela Formaggio, Alessandra Del Felice, Stefano Masiero and Francesco Piccione, Magnetoencephalography in Stroke Recovery and Rehabilitation, Frontiers in Neurology March 2016 Volume7 Article 35.
3. Vadim V. Nikulin, Guido Nolte, Gabriel Curio, A novel method for reliable and fast extraction of neuronal EEG/MEG oscillations on the basis of spatio-spectral decomposition, NeuroImage 55 (2011) 1528–1535.
4. Natalie Schaworonkow, Pedro Caldana Gordon, Paolo Belardinelli, Ulf Ziemann, Til Ole Bergmann and Christoph Zrenner, μ-Rhythm Extracted With Personalized EEG Filters Correlates With Corticospinal Excitability in Real-Time Phase-Triggered EEG-TMS, Frontiers in Neuroscience December 2019 Volume12 Article954.
5. Stefan Haufe, Sven Dähne, Vadim V. Nikulin, Dimensionality reduction for the analysis of brain oscillations, NeuroImage 101 (2014) 583–597
6. Hanna-Leena Halme, Lauri Parkkonen, Comparing Features for Classification of MEG Responses to Motor Imagery, PLOS ONE December 16, 2016.
7. Franc¸ois Tadel,1 Sylvain Baillet,2 John C.Mosher,3 Dimitrios Pantazis,4 and Richard M. Leahy1, Brainstorm: A User-Friendly Application for MEG/EEG Analysis, Hindawi Publishing Corporation Computational Intelligence and Neuroscience Volume 2011, Article ID 879716, 13 pages.
8. Maren Stropahl, Anna-Katharina R. Bauer, Stefan Debener and Martin G. Bleichner, Source-Modeling Auditory Processes of EEG Data Using EEGLAB and Brainstorm, Frontiers in Neuroscience May 2018 Volume 12 Article 309.
9. Benjamin Blankertz, Claudia Sannelli, Sebastian Halder, Eva M. Hammer, Andrea Kübler, Klaus-Robert Müller, Gabriel Curio, Thorsten Dickhaus, Neurophysiological predictor of SMR-based BCI performance, NeuroImage 51 (2010) 1303–1309.
10. Stephen T. Foldes, Douglas J. Weber and Jennifer L. Collinger, MEG-based neurofeedback for hand rehabilitation, Journal of NeuroEngineering and Rehabilitation (2015) 12:85.

Motion control of a cable-restricted underwater vehicle for long-term spot observation

Yoshiki Tanaka
Dept. of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0196, Japan

Yuya Nishida
Dept. of Human Intelligence Systems, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0196, Japan

Kazuo Ishii
Dept. of Human Intelligence Systems, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0196, Japan
E-mail: tanaka.yoshiki732@mail.kyutech.jp, y-nishida@brain.kyutech.ac.jp, ishii@brain.kyutech.ac.jp

Abstract

In order to acquire a time fluctuation data of the resources which needs for a marine resource development, we developed an observation device with low operational risk and a wide observable area. The observation device consists of an underwater station and an underwater vehicle, and underwater vehicle is tethered with a cable. By using the restraint condition of the cable, our device was able to navigate the trajectory planned only by thrust control with an error of up to 0.14m.

Keywords: Underwater vehicle, Long-term observation, Cable-tethered, Self-localization

1. Introduction

In Japan's exclusive economic zone, there are many marine resources including mineral resources and energy resources [1]. These are very important basic resources for the development of an industrial society, but they have not yet reached the economic use of marine resources. The reason is that these resources are widly distributed on the seabed at a depth of 50m or more where general divers cannot dive, it is difficult for divers to search and observation. Un-tethered underwater vehicles, AUVs (automonous underwater vehicles) are used as the practical tools to investigate these marine resources [2][3]. However, the AUVs have the risk of their losing itself on a hardware or software failures, the AUVs cannot be used often for the resource investigation. Thus, a time fluctuation data of the resources which needs for a marine resource development cannot be obtained by an existing survey. This research aims to develop a long-term observation device with low operation risk and wide observable area to acquire a time fluctuation data. In this paper, we explain the proposed long-term observation device and show the experimental results for evaluating the position accuracy using the underwater vehicle.

2. Trajectory model and Motion control

2.1. *Overview*

As shown in Fig.1, the proposed system consists of an underwater station for supplying electric power and an underwater vehicle, the underwater vehicle is tethered to the underwater station with a cable. Underwater vehicle can observe around the underwater station for a long-term by keeping the altitude constant and navigating while applying tension to the cable. By using the restraint condition by tethering, an underwater vehicle can navigate the same trajectory with only simple motion control without using self-localization based on DVL or INS or using waypoints. Therefore, it is possible to observe a wide area with lower risk than existing underwater vehicles.

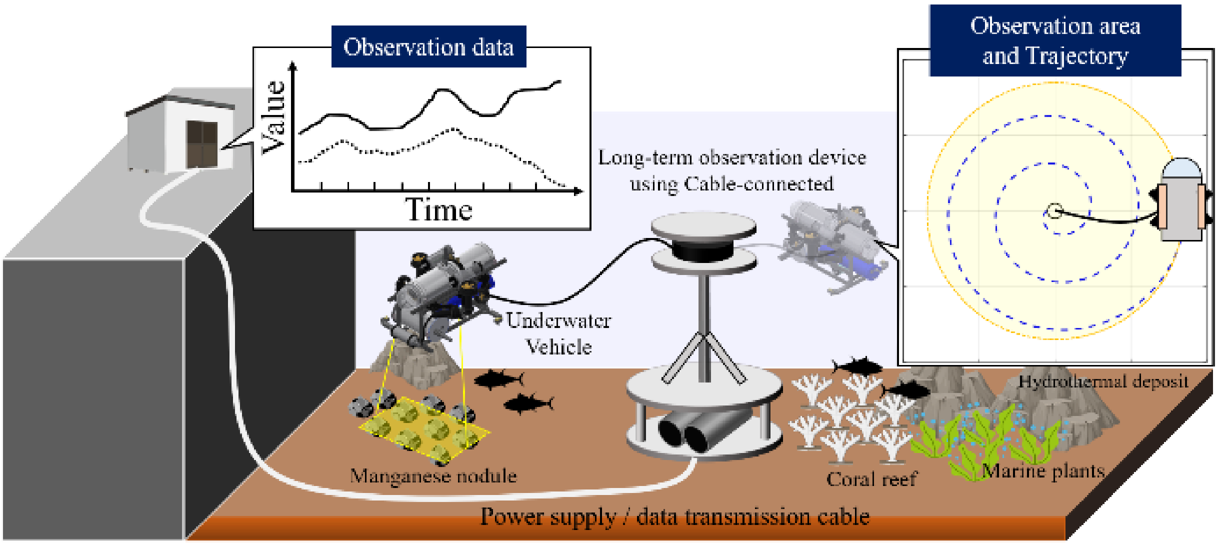


Fig. 1 Overview of proposal method

2.2. *Trajectory model*

To derive the relationship between the structure of the underwater station and the trajectory of the underwater vehicle, we constructed a theoritical formula based on the involute curve. The involute curve is a curved line by the end point $p_i = [x_i\ y_i]^T$ of the cable when the cable wound around the basic circular radius a centered on the origin us always pulled and unwound, x_i and y_i are expressed by the following equations:

$$x_i = a(\cos\theta + \theta\sin\theta)$$
$$y_i = a(\sin\theta - \theta\cos\theta) \qquad (1)$$

Here, θ express the angle from the initial position. In this research, we consider that the basic circular radius a changes by the cable thickness d_{ca} depending on the number of times n of winding and unwinding of the cable. Thus, the basic circular radius a_i is defined by the following equations:

$$a_{wi} = n \cdot d_{ca} + r_b \qquad (2)$$
$$a_{uwi} = (N - n)d_{ca} + r_b \qquad (3)$$

Here, a_{wi} is the basic radius when winding the cable, a_{uwi} is the basic radius when unwinding the cable, and n is the maximum number of times the cable can be winding and unwinding. r_b is the radius of the device, and we adopted a parameter that considers the allowable bending radius of the cable used. Considering that the length of the cable when winding or unwinding changes depending on the number of times n and the angle θ, it is defined by the following equations:

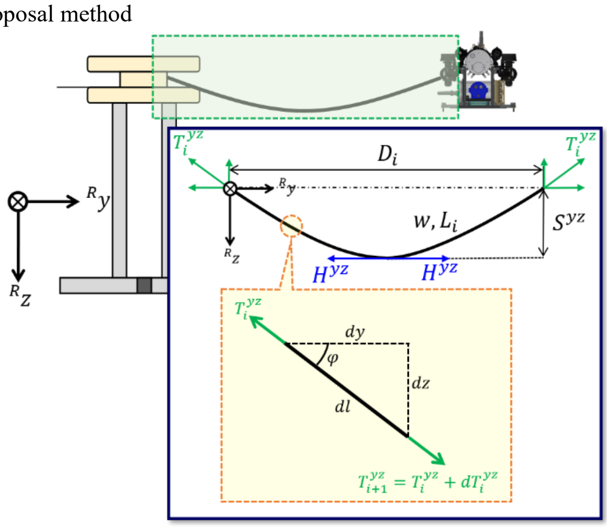


Fig. 2 Cable model

$$L_{wi} = L_m - 2\pi\left\{n \cdot r_b + \frac{1}{2}n(n-1)d_{ca}\right\} \qquad (4)$$
$$L_{uwi} = 2\pi n\left[r_b + \left\{N - \frac{1}{2}n(n-1)\right\}d_{ca}\right] \qquad (5)$$

Here, L_{wi} is the cable length when winding, L_{uwi} is the cable length when unwinding, and L_m is the maximum length of the cable. When applied to Eq. (1) with the changes in the radius of the device and cable length calculated from Eq. (2) and Eq. (4) as constraints, the trajectory $p_{wi} = [x_{wi}, y_{wi}]^T$ at the time of winding is expressed by the following equation:

$$\begin{aligned} x_{wi} &= a_{wi}\cos\theta - (L_{wi} - a_{wi}\theta)\sin\theta \\ y_{wi} &= a_{wi}\sin\theta + (L_{wi} - a_{wi}\theta)\cos\theta \end{aligned} \quad (6)$$

Similarly, the trajectory $p_{uwi} = [x_{uwi}, y_{uwi}]^T$ at the time of unwinding is expressed by the following equation:

$$\begin{aligned} x_{uwi} &= a_{uwi}\cos\theta - (L_{uwi} + a_{uwi}\theta)\sin\theta \\ y_{uwi} &= a_{uwi}\sin\theta + (L_{uwi} + a_{uwi}\theta)\cos\theta \end{aligned} \quad (7)$$

When the cable length calculated from Eq. (4) or Eq. (5) reaches L_m, the trajectory of the underwater vehicle becomes a circle centered on the device. Therefore, it is expressed by the following equation.

$$\begin{aligned} x_i &= L_m\cos\theta \\ y_i &= L_m\sin\theta \end{aligned} \quad (8)$$

Here, the cable length in the trajectory calculated fromEq. (8) is symmetrical around the device. Therefore, the range of the angle θ is $0 \leq \theta \leq 180$ [deg].

2.3. *Motion control*

In the water, the weight and fluid resistance of the cable increase depending on the length of the unwound cable, so the sag of the cable increases. As the sag increases, it becomes more difficult for underwater vehicle to draw the planned trajectory and the positioning accuracy decreases. Therefore, we calculated the thrust required in the direction of travel and the thrust required in the direction of extension. At first, we explain the thrust force F_{SRG} required in the direction of travel. In the future, we plan to create a map from seabed images by a camera mounted on an underwater vehicle. We are considering the creation of a seafloor map for long-term observation of marine resources. In order to create a map using images taken by the equipped camera in underwater vehicle, it is necessary to input velocity parameters to the underwater vehicle so that the captured images overlap. The velocityV_x required to satisfy a specific overlap rate is expressed by the following equation:

$$V_x \leq \frac{d_S(1 - \alpha_{OL})}{T_S} \quad (9)$$

Here, d_S is the shooting range of the camera, α_{OL} is the overlap rate, and T_S is the shooting cycle. By substituting the velocity calculated from Eq. (9) into the following equation, the thrustF_{SRG} in the direction of travel can be obtained.

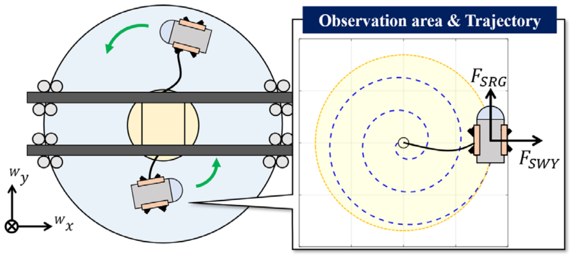


Fig.3 Experimental setup

Table1 Thrust value F_{SRG} and F_{SWY}

F_{SRG}[N]	F_{SWY}[N]			
4.0	5.0	10.0	15.0	20.0
9.0	5.0	10.0	15.0	20.0
16.0	5.0	10.0	15.0	20.0

$$F_{SRG} = D_x|V_x|V_x \quad (10)$$

Here, D_x is the fluid drag force. This parameter was approximated by experiment. Next, we explain the thrust force F_{SWY} in the direction of extension. Fig.2 shows the cable model of the underwater vehicle when it is tethered to the underwater station. If the cable has ideal sag without considering the stiffness, and the underwater vehicle moves slowly, the cable can be considered to have a catenary shape [4]. Therefore, the sag z_{ci} generated by the cable's own weight and the tensionT_i^{yz} applied to the cable can be expressed by the following eqations:

$$z_{ci} = \frac{H^{yz}}{w}\left\{\cosh\frac{wD_m}{2H^{yz}} - \cosh\frac{w}{H^{yz}}\left(\frac{D_m}{2} - D_i\right)\right\} \quad (11)$$

$$T_i^{yz} \geq H^{yz}\cosh\frac{w}{H^{yz}}\left(\frac{D_m}{2} - D_i\right) \quad (12)$$

Here, H^{yz} is the horizontal tension at the lowest point of the cable, w is the weight in water per unit cable length, D_i is the distance between center of the device and underwater vehicle, D_m is the maximum distance between center of the device and underwater vehicle. Therefore, the thrust force in the direction of extension required for the underwater vehicle to suppress the sag caused by the weight of the cable is expressed by the following equation:

$$F_{SWY} \geq T_i^{yz}\cos\varphi \quad (13)$$

In this paper, we calculated the minimum thrust F_{SWY} in the direction of extension by assuming that the trajectory of the underwater vehicle obtained from Eq. (6) to Eq. (9) contains an error of at most 1% due to cable's sag.

3. Wet test and evaluation of trajectory

To evaluate the trajectory of the cable-tethered underwater vehicle described in section 2.2, we

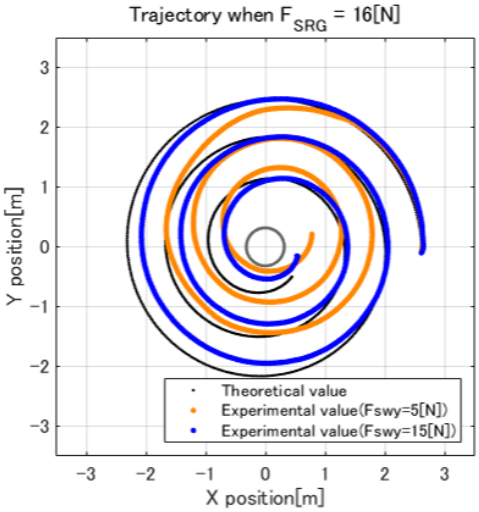


Fig.4 Trajectory of AUV KYUBIC
(Theoretical value vs. Experimental value)

experimented using the hovering type AUV KYUBIC. This AUV is equipped with IMU, DVL and depth sensors as navigation sensors. We used the velocity data from DVL and the heading data from IMU to evaluate the self-localization. Fig.3 shows an overview of this experiment. The AUV is tethered to the device by a cable (thickness is 12mm and length is 2.2m). The device resembling an underwater station is set in the center of the water tank (diameter is 6m and depth is 1.2m). In this experiment, we used the parameters shown in Table1 as the thrust force applied to the AUV and its self-localization was measured when winding the cable. Fig.4 shows the trajectory calculated using Eq. (6) and the self-localization of the AUV at $F_{SRG} = 16$[N]. The positional accuracy is improved by increasing the thrust force F_{SWY}. However, a different trajectory from the theoretical value can be confirmed immediately after the AUV starts moving. This cause is considered that the azimuth during navigation contained an error with the teoretical value because the AUV moved only by thrust control. Table2 to Table4 show the mean square error (MSE) of the self-localization when the thrust applied to the AUV is changed. It can be confirmed that the distance error between the center of the device and AUV is reduced by increasing the thrust force F_{SWY}. However, the MSE of distance at $F_{SWY} = 20$[N] is larger than at $F_{SWY} = 15$[N]. This cause is considered that the velocity of the AUV is reduced by increasing the thrust force F_{SWY}, and the position data calculated by the DVL and IMU contain accumulated error.

Table2 Position error when $F_{SRG} = 4$[N]

F_{SWY} [N]	MSE X position[m]	MSE Y position[m]	MSE Distance [m]
5.0	0.06	0.03	0.07
10.0	0.03	0.03	0.03
15.0	0.02	0.04	0.03
20.0	0.01	0.10	0.06

Table3 Position error when $F_{SRG} = 9$[N]

F_{SWY} [N]	MSE X position[m]	MSE Y position[m]	MSE Distance [m]
5.0	0.06	0.05	0.07
10.0	0.05	0.03	0.04
15.0	0.03	0.04	0.04
20.0	0.03	0.11	0.07

Table4 Position error when $F_{SRG} = 16$[N]

F_{SWY} [N]	MSE X position[m]	MSE Y position[m]	MSE Distance [m]
5.0	0.12	0.14	0.14
10.0	0.05	0.07	0.05
15.0	0.03	0.02	0.03
20.0	0.03	0.09	0.06

4. Conclusions

We proposed a new method that can observe a wide area with lower risk than exisiting method. Our instrument was able to move only by thrust control, with an average error of up to 0.14m relative to the theoreticel value. In future work, we plan to evaluate the trajectory when unwinding the cable and add the heading control.

References

1. C.E. Fitzgerald, et al., Hydrothrmal Manganese Oxide Deposits from Baby Bare Seamount in the Northeast Pacific Ocean, Marine Geology, Vol.225, Issue 1-4, pp.145-156,2006
2. Y. Nishida, et al., Development of an autonomous underwater vehicle for survey of cobalt-rich manganese crust, OCEANS2015 MTS/IEEE Washington, pp.1-5, 2015
3. Toshihiro Maki, et al., Docking Method for Hovering-Type AUVs Based on Acoustic and Optical Landmarks, Journal of Robotics and Mechatronics, Vol.30, No.1, pp.55-64, 2018
4. Atsushi Imadu, et al., Control Method for Helicopters Tethered to Ground Station with Compensation for Disturbance Caused by Cable Tension, Jornal of Robotics and Mechatronics, Vol.29, No.4, pp.737-745, 2017

Development of current sensors for digitizing expert knowledge in fish feeding towards sustainable aquaculture

Dominic B. Solpico
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka 808-0135, Japan

Yuya Nishida
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka 808-0135, Japan

Kazuo Ishii
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka 808-0135, Japan
E-mail: solpico.dominic-bautista806@mail.kyutech.jp, y-nishida@brain.kyutech.ac.jp, ishii@brain.kyutech.ac.jp
www.lsse.kyutech.ac.jp

Abstract

Improving the efficiency of fish feeding contributes to achieving sustainable expansion of the aquaculture industry. However, expert knowledge on feeding remains reliant on experience. This paper presents a new approach of digitizing such knowledge by measuring underwater currents induced by fishes as indicator of their behavior and appetite. A prototype current sensor suite was constructed to measure the current around the fish cage, especially during feeding.

Keywords: aquaculture, fish feeding, underwater currents, current sensors

1. Introduction

While seafood consumption has been increasing every year, capture fisheries production, facing issues of overfishing, has remained static with slight fluctuations in the past decades. Since early 1990s, the aquaculture industry has been expanding consistently to meet the rising demand worldwide, making up 46% of the global production in 2018.[1]

The problem arises in the industry's sustainability. Unsustainable management of fish farms not only poses a threat to their surrounding aquatic environment but also to the health of the fish stocks.[2,3,4] One crucial issue is the need for efficient feeding decision-making. Poorly timed and excessive feeding leads to poor cost-efficiency in raising fish.[5] Worse, it produces uneaten feeds that decompose at the bottom, contributing to the decline of water quality in the surrounding environment.[6] Such threat leads to slower growth rates, poorer quality of harvest, or at worst, massive fish kills.[6,7,8] Fish farmers incur losses in their operations as a result.[3,5,6]

Efficient feeding has usually been achieved with a decision made by an expert farmer. There is a significant difference in the quality of the harvested fish fed based on expert and non-expert decisions. Such decision making remains to be an "art," where prediction is still intuitive, subject to the expert's experience, and unquantifiable by a unified standard.[9,10]

Digitization of experts' knowledge may help non-experts improve their feeding decisions, which will not only improve the amount and quality of their harvest, which will not only consequently increase their income, but will also increase the supply of high-quality seafood, lower their prices, and at the same time minimize pollution in the farm environment.

Current research efforts on this problem focused on the measurement of fish behavior through machine vision and water quality changes using sensors as input parameters into machine learning models to make a feeding decision.[5,10,11] So far, experiments in these works were performed in intensive farms where environmental factors were controlled. No research work performed in open environments, i.e. extensive farms, have been found so far.
This paper explores a new approach on digitizing expert feeding knowledge in tuna farms in aquatic environments. In this approach, velocities of underwater currents induced by the movement of fishes inside a cage will be measured and their relationship to the behavior and appetite of the fish inside will be investigated. While presenting the sensor system's overall architecture, this paper focuses on the design of a prototype sensor for initial current measurements and on its ongoing development.

2. Sensor System Architecture

In this system, a network of sensor nodes is placed around a 40-meter diameter fish cage with depth of 20 meters. Each node consists of multiple current sensors measuring current velocities for every 1-meter depth. In

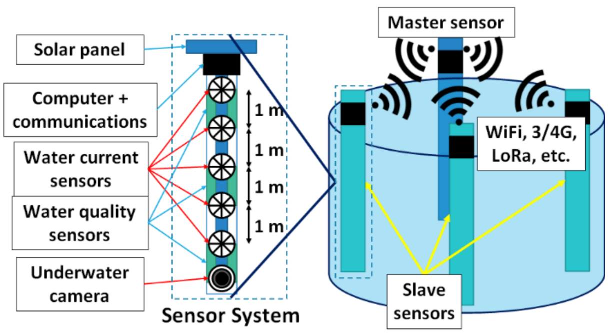


Fig. 1. System architecture for monitoring fish appetite and water quality in a fish cage

addition to measuring currents, it also has water quality sensors (dissolved oxygen, temperature, conductivity, pH, etc.) at multiple depths. An underwater camera is placed to observe the fish movements, especially during feeding. Each node is designed to have a capacity for energy-harvesting– solar, wind, tidal, or other sources – so that it can operate continuously off-grid. A computer above surface performs corresponding calculations on the sensor readings to obtain the measurements, timestamps them, and stores them internally.
These sensor nodes communicate in a star network, where one node is designated as the master and the rest as slaves. Slaves send their measurement data to the master, which also collects its own measurements. Data aggregated by the master node may either be collected by the farmer onsite or be transmitted directly to a data center. Communication technology to be used – WiFi, 3G, 4G, LoRa, etc. – will be configured accordingly to implement this network architecture.
Initial data collection is to be made in a fish farm, where underwater currents are to be measured near cages at depths with fish presence, especially during feeding times. This is to provide insight for validating or in improving the proposed design.

3. Sensor Suite for Initial Measurements

3.1. *Sensor suite system design*

For the initial measurements, two custom-built current speed loggers are mounted on a metal frame at two different depths at which fishes are observed before, during, and after feeding. These sensors measure only

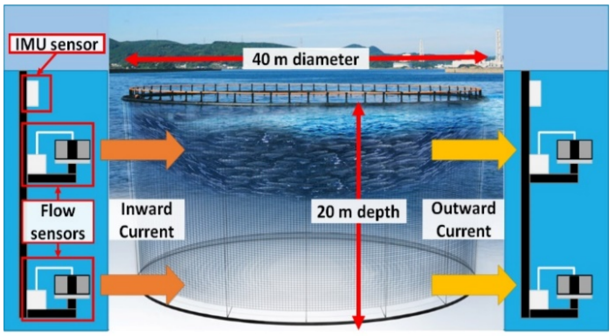


Fig. 2. Design of the initial farm experiment using the prototype sensor suite consisting of flow and IMU sensors

one flow axis given their fixed mounting. An inertial measurement unit (IMU) sensor is mounted closest to the current sensor of interest. This will measure movement of the frame caused by fish-induced currents. These sensors are remotely triggered by Bluetooth to start logging measurements before they are placed underwater for hours.
Ideally, at least two sets of sensor units should be deployed so that the other will measure the current going towards the fish cage, which will be cancelled from the

outgoing current. Due to time constraints, however, one set will be used for the initial measurement. Two or more sets will be deployed for the next measurement campaigns.

3.2. *Custom-built current sensor*

3.2.1. *Components and operation*

Since initial measurements will be performed, it was decided to use a low-cost current sensor. The sensor developed for this experiment is a modified propeller-type flow sensor intended for measuring water flow through water pipes. Its Hall-effect sensor generates

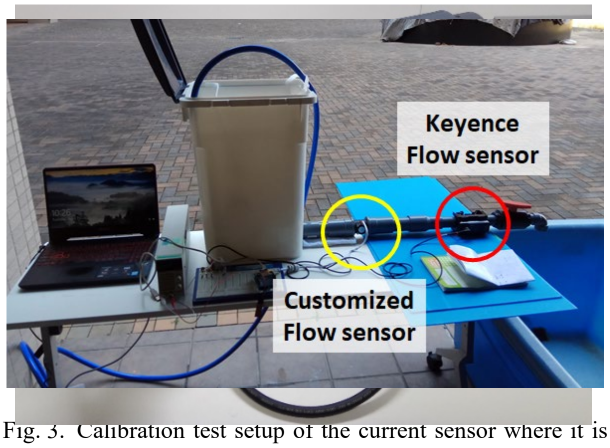
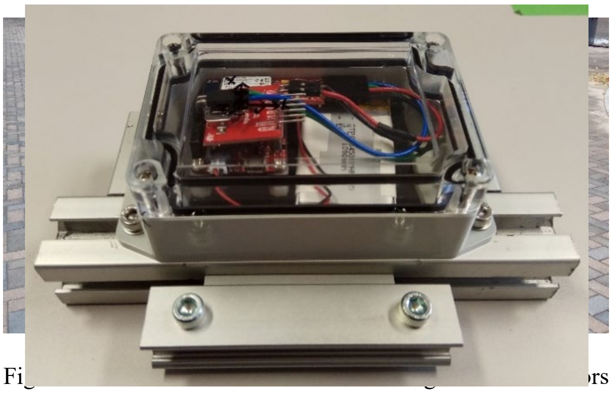


Fig. 3. Calibration test setup of the current sensor where it is connected to a digital flow sensor as they both measure water flow from the container

Fig. 5. Constructed current sensor for initial measurements with the reinforced flow sensor and the electronics box on its aluminum mounting

pulses proportional to the magnetic propeller's rotation. An Arduino microcontroller counts these pulses for a given period, obtains the average frequency throughout the period, and then calculates the current speed using a calibration coefficient. It then adds timestamping to the reading and stores it in a microSD card with its datalogging shield with SD and RTC capability. A Bluetooth module is also connected so it can be triggered remotely by a computer to start and stop measurements.
The current sensor is powered by a 9-volt supply (six AA batteries in series) with a capacity of 2700 mAh. A power endurance test was made by allowing it to operate continuously while being powered by the batteries. Result showed that it can collect measurements reliably for around 22 hours.
The flow sensor was not originally designed to be waterproof and reinforcements were therefore made by permanently sealing its electronics enclosure and by replacing its original cable with a waterproof rubber molded cable. Other unit components are housed in an IP68 enclosure.
The mounting is a 12-meter aluminum structure with four legs to which the sensors are attached to. The sensors depth can be adjusted by sliding them through the legs. Each leg is made of three four-meter frames. Cross-like reinforcements are attached at the leg joints to minimize



Fi... ...rs without reinforcement at the segments which explains the bending

Fig. 3. IMU sensor enclosed in its waterproof box with its aluminum attachment to the frame

bending and to enable the mounting to withstand the underwater currents.

3.2.2. *Flow sensor cross calibration*

For accurate measurements, the developed flow sensor was cross-calibrated with a digital clamp-on type flow sensor (Keyence FD-Q32C). Both sensors were connected to an elevated water source where flow was partially controlled, as velocity and flow rate were dependent on the height, and subsequently the volume, of the water in the container.
Calibration is done by correlating the pulse frequency to the speed of water through the custom sensor, which is calculated by measuring the flow rate through the digital sensor. This is calculated using the relationship between the flow rate and speed of a fluid through a pipe and the continuity equation, which given in the following equation:

$$v_F = \frac{v_{DF} A_{DF}}{A_F} = \left(\frac{q_{DF}}{A_F}\right)\left(\frac{1000}{60}\right) \tag{1}$$

In this equation, v is the water speed, A is the cross-section area of the sensor pipe, and q is the flow rate. Subscripts F and DF denote the custom and the digital sensors, respectively. Regardless of the difference of the cross-section areas of the two sensors, that of the digital sensor is cancelled out in the equation. The constant at the right converts the units from flow rate (L/min) to speed (cm/s).

To select the best averaging period to be used in measurements, three periods were selected for calibration – 1, 5, and 10 seconds. Readings at 1 second were found to be discrete as the microcontroller count discrete number of pulses per second. Range of readings from the digital sensor are classified as the discrete readings from the custom sensor. Readings using the 10-second configuration were more continuous. However, there are higher chances of averaging high and low sample values, which may not properly represent the actual measurement. Using the 5-second period seems to be a favorable configuration as there are smaller chances of samples with large differences, while its readings are still continuous. This is therefore the selected configuration for the upcoming experiments.

3.3. *IMU water movement sensor*

The core component of this sensor is the Sparkfun 9DoF Razor IMU M0, a very compact microcontroller with an MPU-9250 IMU and a µSD card slot onboard. Its IMU consists of accelerometer, gyroscope, and magnetometer sensors, and is therefore capable of measuring linear acceleration, angular rotation velocity, and magnetic field vectors. With a Real-Time Clock (RTC) attached, this board can timestamp its measurements before writing them to an µSD card. Powered by a 1000-mAh lithium-ion polymer (Li-Po) battery, it is enclosed in an IP68 enclosure. Power endurance test result showed that this sensor can collect data for around 22 hours as well.

4. Trial river measurements

While waiting for the plans of the initial fish farm experiment to be finalized, we decided to perform current measurements with the constructed sensors along two rivers in the northern part of Fukuoka. The first measurement was along Onga River (遠賀川), one of the longest rivers in Kyushu island. Its width and depth at the point of measurement is around 290 and 1.6 meters, respectively. The second measurement was along Nishi River (西川), a small river connected to the northwestern end of Onga River, where the width and depth at the measurement point is around 75 and 1.3 meters, respectively. Due to the shallow depths, only one 4-meter frame segment was used. Currents at the bottom and near surface of both rivers were measured, with the heights of the upper sensor were adjusted accordingly. The IMU sensor was placed in proximity to the bottom current sensor. Measurements at each river were taken for around 30 minutes.

Save for two data points for each depth ranging from 0.75 to 2.26 cm/s, the currents sensor readings along Onga River were at 0 cm/s. More non-zero sensor readings were collected at Nishi River, with the measurement at the surface peaking to 18.8 cm/s. On the other hand, almost all readings at the bottom of this river is at zero, except for a few instances peaking at 3.76 cm/s. No significant changes were observed in the accelerometer and gyroscope readings since the frame was settled at the riverbed. Any changes observed was attributed to lack of calibration as well as human intervention.

It was expected for Onga river's current to be slow because of its large width. However, visual observations suggest that there still were small movements by the water, as seen from the waves at the surface. This may indicate that the flow sensor could not detect those small currents, reading them as 0 cm/s.

For better results in future measurements, we will calibrate the sensors' pulse readings for each measurement to the readings of a calibrated digital sensor. Doing this can verify whether the custom sensors can detect small currents or not. In addition, we are considering attaching a funnel at the inlet of the current sensors. This will amplify small currents entering the sensor, which will then use its measured current to calculate for the current at the funnel inlet.

5. Conclusion and future work

This paper presents the ongoing development of a prototype sensor suite, consisting of modified flow for measuring water currents induced by fish movement. Flow sensors were modified and recalibrated to measure underwater currents. The IMU sensor was also developed for measuring movement of the mounting frame. Both the current sensors and the IMU sensor can record current measurements reliably for at least 22 hours.

As the mounting frame has been completed, the sensor suite is ready to be deployed in the fish farm for the initial measurements, which will be performed within the next few weeks. Trial measurements have been made along two rivers in Fukuoka, showing A second sensor suite will also be constructed after the first experiment. The

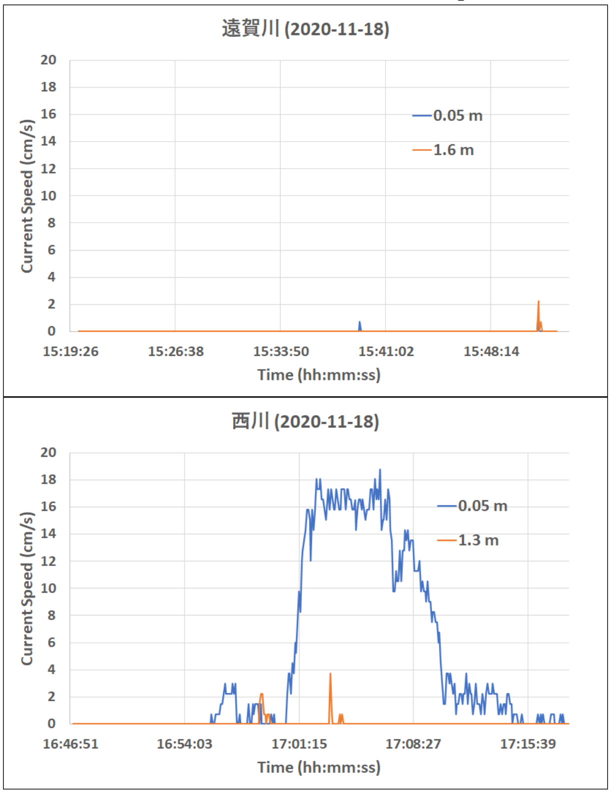


Fig. 3. Current sensor readings at both depths at Onga River were very low, while larger readings were measured at the near surface of Nishi River.

relation of current measurements from the experiments with the fish behavior during feeding will be analyzed. Future work also includes development of a networks of sensor nodes as described in the system architecture using more robust current sensors. These are future research tasks towards digitizing expert knowledge in fish feeding in aquaculture farms.

References

1. FAO: The State of World Fisheries and Aquaculture 2020. In brief. Sustainability in action, FAO, 2020.
2. R. S. S. Wu, "The environmental impact of marine fish culture: towards a sustainable future," *Marine Pollution Bulletin*, vol. 31, no. 4–12, pp. 159–166, 1995.
3. A.R. Brown et al., "Assessing risks and mitigating impacts of harmful algal blooms on mariculture and marine fisheries," *Review in Aquaculture*, vol. 12, no. 3, pp. 1663-1688, 2020.
4. M. Martinez-Porchas, L. R. Martinez-Cordova, "World Aquaculture: Environmental Impacts and Troubleshooting Alternatives", The Scientific World Journal, vol. 2012, Article ID 389623, 9 pages, 2012.
5. T. Wu, Y. Huang, J. Chen, "Development of an adaptive neural-based fuzzy inference system for feeding decision-making assessment in silver perch (Bidyanus bidyanus) culture," Aquacultural Engineering, vol. 66, pp. 41-51, 2015.
6. M. L. San Diego-McGlone, R. V. Azanza, C. L. Villanoy and G. S. Jacinto, "Eutrophic waters, algal bloom and fish kill in fish farming areas in Bolinao, Pangasinan, Philippines," *Marine Pollution Bulletin*, vol. 57, no. 6-12, pp. 295-301, 2008.
7. G. A. Santos, J. W. Scharma, R. E. P. Mamauag, J. H. W. M. Rombout and J. A. J. Verreth, "Chronic stress impairs performance, energy metabolism and welfare indicators in European seabass (*Dicentrarchus labrax*): The combined effects of fish crowding and water quality deterioration," *Aquaculture,* vol. 299, no. 1-4, pp. 73-80, 2010.
8. H. Sun, J. Li, L. Tang and Z. Yang, "Responses of crucian carp Carassius auratus to long-term exposure to nitrite and low dissolved oxygen levels," *Biochemical Systematics and Ecology*, vol. 44, no. 1, pp. 224-232, 2012.
9. M. Sun, S. G. Hassan, D. Li, "Models for estimating feed intake in aquaculture: A review," *Computers and Electronics in Agriculture*, vol. 127, pp. 425-438, 2016.
10. C. Zhou et al., "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.
11. C. Zhou, et al., "Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision," *Aquaculture*, vol. 507, pp. 457-465, 2019.

Underwater image reconstruction using convolutional auto-encoder

Shinsuke Yasukawa
Dept. of Human Intelligence Systems, Kyushu Institute of Technology,
2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka, 808-0196, Japan
E-mail: s-yasukawa@edu.brain.kyutech.ac.jp

Sreeraman Srinivasa Raghura
Dept. of Human Intelligence Systems, Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka, 808-0196, Japan
E-mail: srinivasa.raghuraman-sreeraman344@mail.kyutech.jp

Yuya Nishida
Dept. of Human Intelligence Systems, Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka, 808-0196, Japan
E-mail: y-nishida@brain.kyutech.ac.jp.

Kazuo Ishii
Dept. 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

One of the main tasks of AUVs is to capture deep-sea images like fishes, crabs, other living organisms and resources for information leading to research on deep-sea ecosystems. Acoustic transmission are used to establish wireless underwater communications between the AUV and the ship. However, there are some limitations in the communication channels due to limited bandwidth, multi-path, temperature distribution and change in the direction of transmitting source and receiving sensor which results in losses in data being transmitted. Initially, the captured images are enhanced to reduce the effect of light attenuation and then compressed for transmission through acoustic modems. Only an important part of image is being transmitted through set of data packets. The received data packets in the ship will be reconstructed to predict the presence of living organisms. The loss in data during transmission creates a difficulty for the operators to predict the exact information. In this research, to compensate this transmission loss, an efficient compression and reconstruction technique using convolutional autoencoder with minimal distortion is proposed. Finally, for evaluation of the proposed image compression technique, the quality of reconstruction of images with and without data loss will be compared using the quality metrics signal to noise ratio (PSNR), structural similarity index(SSIM) and perceptual quality of image.

Keywords: sampling-AUV, acoustic communication, image compression, convolutional auto-encoder

1. Introduction

Presently numerous research and development activities have been carried out in the field of underwater robotics in ocean engineering. Research studies, especially in the area of autonomous underwater vehicles (AUV) which are being used as tools for deep sea-floor surveying and observation. AUVs are mainly employed for performing tasks like underwater imaging, sea-floor mapping, exploiting resources and minerals which are of great interest to geologists and marine biologists to understand the sea flora and fauna.

One of the main tasks of AUV is optical imaging of parts of the sea-floor that contains marine habitat. These kinds of images contribute to research studies on deep-sea biology. The further task involves collecting samples using end-effectors. It is difficult for the AUV to perform the above task autonomously as it does not have any prior knowledge on the targets of interest. To overcome this difficulty, it operates semi-autonomously by acquiring assistance from biologists and geologists. Fig. 1 shows the flow of the mission to sampling objects on sea-floor. The acquired image is then transmitted to the vessel. Images are transmitted underwater using acoustic waves through hydrophones. The received image in the ship is examined by the operator and if it shows any existence of marine habitat then behavior commands are sent to the AUV. Finally, AUV is sent back to the specified location for collecting samples. Our group developed the sampling-AUV for the sampling mission[1,2,3] and succeeded three times in a row in sampling experiments in Suruga Bay in March 2018[3].

The process of image transmission involves the following steps:

(i) First, the AUV acquires raw images of seafloor that are of interest.
(ii) These images are then enhanced to reduce the effect of light attenuation and unbalanced illumination.
(iii) Next, images compressed using the required compression technique for transmission through the acoustic medium.
(iv) Finally, the data received in the ship reconstructed and the operator examines the reconstructed image.

During the process of image transmission, there is a loss of data(information). The data loss is mainly due to the following factors: multipath, temperature distribution, Signal attenuation over long ranges, and directional changes with the transmitting source and the receiving sensor. Underwater communication has low data rates when compared to terrestrial communication since it uses acoustic waves instead of electromagnetic waves.

It is known that some loss occurs during the transmission of the compressed data through the acoustic communication. This data loss has to be minimized to obtain a better-quality image reconstruction at the receiving end. Therefore, we propose a lossy based image compression method that utilizes the merits of convolutional autoencoder network. The purpose of this research is developing an efficient image compression model that can reconstruct the image with minimal data loss. To achieve the above, the deep-learning approach is used to simulate the compression/ reconstruction model with the generation of data loss in the hidden codes that would result in improved Peak Signal to Noise Ratio (PSNR) and better Structural Similarity Index (SSIM).

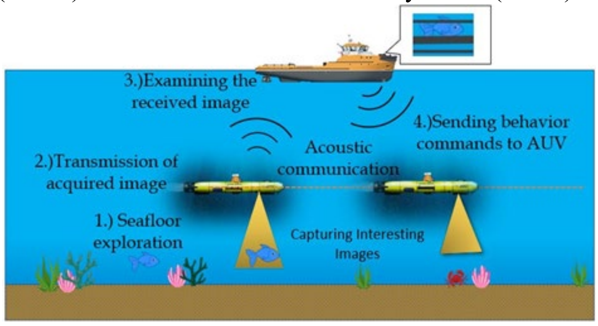


Fig. 1. Underwater Image Transmission.

2. Related works

Ahn et. al.[2] have developed an image enhancement and a compression technique for underwater acoustic communication with limited data information density. Their proposed method enhances the brightness of the sea floor images and reduces the image depth from 24-bit to 4-bit. They have used a data compression method that is based on indexed color technique which expresses images with a fixed number of color indices using a suitable color palette. The color palette consists of set of indexed colors to express the enhanced images depending on the targeted creatures. As shown in Fig. 2, the problem of data loss that occurred during transmission could not be minimized which affected the quality of image reconstruction at the receiving end. The image is transmitted in the form of packets of information.

Hoag et.al.[4] have developed an underwater image compression technique using wavelet transforms (WT). They have employed a wavelet decomposition method in combination with vector quantization for compressing still underwater images. Compared to JPEG compression their proposed method tends to perform well at low bit rates. However, over a wide range of directionality the wavelet transforms are generally inefficient in representing geometric structures. The texture and details of reconstructed images are not efficiently retained using wavelet transforms.

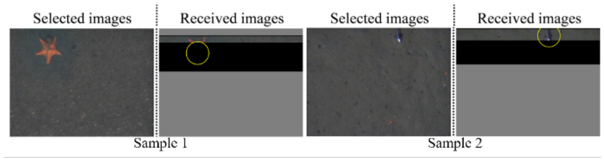


Fig. 2. Selected and received images with data loss during transmission[2].

In the research work carried out by Cheng et.al.[5] a lossy image compression method using deep convolutional autoencoders (CAE) was proposed. They have designed a convolutional autoencoder network that replaces conventional transforms and performed training using rate-distortion loss function. Compared to JPEG2000 and other traditional algorithms their method performs well on the images from kodak database.

3. Proposed Method

The Fig. 3 shows an example of overview of the proposed image compression method utilized for underwater image transmission process. Initially the acquired image data is compressed to codes (assuming the encoder in the AUV). In the compressed representation we generate a data loss by discarding a limited amount data from the codes (i.e. eliminating few rows or columns). This corrupted data will be sent to the decoder for reconstruction. The decoder extracts relevant information from this data and attempts to reconstruct the image similar to that of the original. The reconstruction depends on the robustness of the decoder.

Fig. 4 shows CAE Network with data loss. As an initial step, the dataset consisting of interesting images was created for training and testing from a database consisting of enhanced underwater images. The selected images from the database were resized from a dimension of 320×240×3(width, height) to 224×224×3(width, height) to match the overall receptive field of the standard convolutional layers. As a preprocessing step, the input images of dimension 224×224×3 are initially normalized to [0,1]. The dataset consists of a total 100 images with 70 images for training and 30 images for testing. Using convolution and deconvolution filters a symmetric convolutional auto-encoder network is designed. Then, training is performed using mean squared error loss function and Adam optimizer.

Here, based on the standard CAE, an additional pseudo layer was introduced which takes the input data from the encoded layer. This layer attempts to slice the compressed data (i.e. from a dimension of 14×14×16 to 8×8×16, where 16 refers to the number of feature maps) there by creating some loss of data randomly. Therefore, the decoder attempts to reconstruct the image from this representation. Depending on the size of data loss, the decoder needs to be modified to match the size of the original image. Due to the above sperate decoder networks are required for reconstruction.

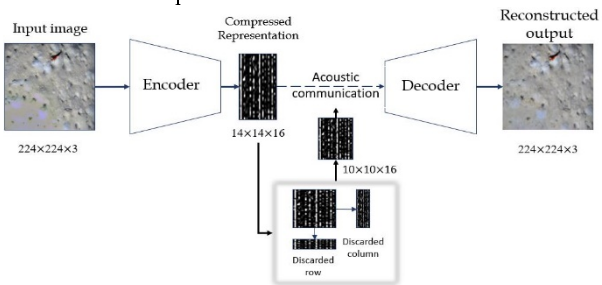


Fig. 3. Overview of the proposed image compression method.

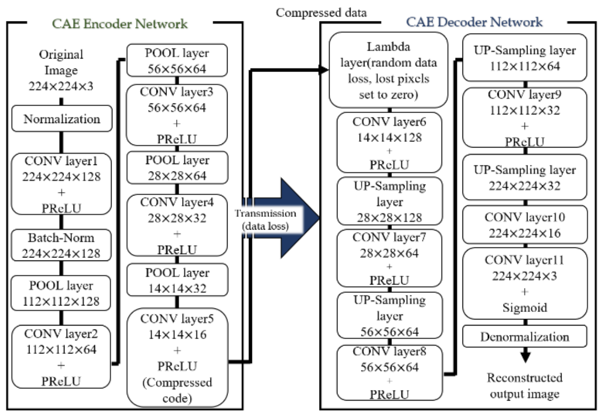


Fig. 4. CAE Network with data loss.

4. Experiments and Result

4.1 Experimental setup

A dataset consisting of 100 images was used for training and testing the proposed model. These images were obtained from previously conducted seafloor exploration experiments using Sampling-AUV: TUNA-SAND2. The number of filters in the convolutional layers were set as {128,64,64,32,16} and the decoder part mirrors the encoder. Mean square error loss function was used during training process to compute the distortion between original and reconstructed images. The parametric ReLU activation was employed in the network to improve the quality of reconstruction at high bit rate. For measuring the efficiency of the proposed image compression technique, the quality of reconstruction of images with and without data loss were compared using the quality metrics signal to noise ratio (PSNR) and structural similarity index(SSIM).

The simulations were performed on a PC with the following specifications Intel Core i7-8750H CPU at 2.40 GHz, 8GB RAM and GeForce GTX 1060Ti GPU. The CAE model was trained for 150 epochs i.e. 150 iterations over all the samples in mini-batches of 2. The running time i.e. the time taken for one complete encoding and decoding process for RGB images with resolution of 224 × 224 was approximately 3.60s/step.

4.2 Results

The Fig. 5 shows the comparison of PSNR at various compression ratios of original and reconstructed images. The proposed model was able to achieve a high signal to noise ratio which results in better quality of image. At higher compression ratios there is slight reduction in quality along with decrease in PSNR value.

Since the PSNR evaluation considers only numerical comparison of images, SSIM is also taken as a measure for evaluating the quality of the image. From the graph shown in Fig. 6 we can see that a maximum SSIM value of 0.78 for sample image 1 at a compression ratio of 18:1, the decoder is able to reconstruct the image with satisfactory image quality even with data loss in the compressed codes. The quality of reconstruction decreases at higher compression ratios. This is because when trying to reconstruct the discarded portion the autoencoder tries to add black and white color pixel intensities in place of lost pixels.

While evaluating the quality metrics in terms of luminance, contrast and structure of the reconstructed image with the original image considering the data loss, the results indicate that the proposed CAE model is able to provide better reconstruction quality and achieve an accuracy of 55 to 70%.

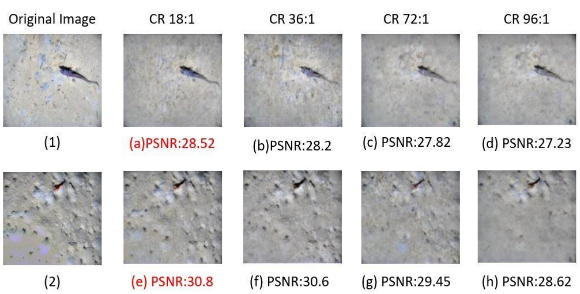


Fig. 5. Comparison of PSNR at different compression ratios with data loss.

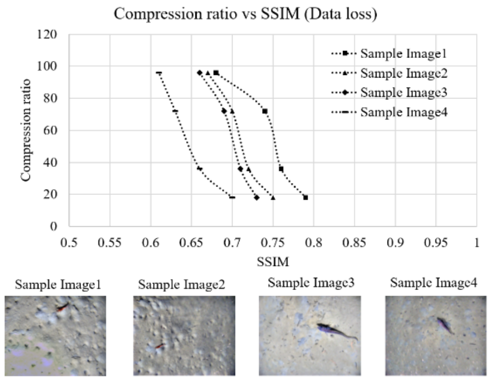


Fig. 6. Comparison of SSIM with data loss.

5. Conclusion

In this research, an efficient lossy image compression/reconstruction architecture using convolutional autoencoder was developed. The proposed CAE model tends to achieve better image reconstruction and perceptual quality with minimal distortion. The performance of the image compression algorithm was evaluated using quality metrics like PSNR and SSIM. It was observed from the results that the proposed autoencoder model was able to achieve fairly high signal to noise ratio with data loss providing a good reconstruction quality image.

References

1. Nishida, Y., Sonoda, T., Yasukawa, S., Nagano, K., Minami, M., Ishii, K., & Ura, T., Underwater Platform for Intelligent Robotics and its Application in Two Visual Tracking Systems, *Journal of Robotics and Mechatronics,* 30(2), 238-247, 2018.
2. Ahn, J., Yasukawa, S., Sonoda, T., Nishida, Y., Ishii, K., & Ura, T. An optical image transmission system for deep sea creature sampling missions using autonomous underwater vehicle, *IEEE Journal of Oceanic Engineering*, 2018.
3. Yasukawa, S., Nishida, Y., Ahn, J., Sonoda, T., Watanabe, K., & Ishii, K., Development and Sea Trials of Vision-Based Control for Sampling-AUV, *In 2019 IEEE Underwater Technology*, pp. 1-4, 2019.
4. Hoag, D. F., & Ingle, V. K., Underwater image compression using the wavelet transform. *In Proceedings of OCEANS'94*, Vol. 2, pp. II-533, 1994.
5. Cheng, Z., Sun, H., Takeuchi, M., & Katto, J., Deep convolutional autoencoder-based lossy image compression. *In 2018 Picture Coding Symposium (PCS)* , pp. 253-257, 2018.

The Spherical Robot Transfer Problem with Minimal Total Kinetic Energy*

Kenji Kimura
Engineering course, Fukuoka Daiichi High School
22-1 Tamagawamati, Minamiku, 815-0037, Fukuoka, Japan
E-mail: kimuken1977_2058@yahoo.co.jp†

Kazuo Ishii
Graduate School of Life Science and engineering, Kyusyu Institute of Technology
2-4 Hibikino,Wakamatsu-ku, Kitakyushu 808-0196, Fukuoka, Japan
E-mail: ishii@brain.kyutech.ac.jp

Abstract

Previous spherical mobile robots were driven by two rollers with a fixed rotational axis, which restricts the angular velocity vector of the sphere to two dimensions. Three-dimensional freedom is expected to improve the rotational diversity of the sphere. This study proposes a spherical mobile robot with a variable roller-rotational axis that allows three-dimensional freedom of movement. Furthermore, the kinetic energy of transporting the sphere by the rollers is minimized by an optimization procedure.

Keywords: Angular velocity vector of the sphere, Angle of sphere rotational axis, Total kinetic energy

1. Introduction

Many types of robot, such as omnidirectional mobile robots and sphere-transported robots, are based on spherical motions. Therefore, various roller arrangements have been proposed for mobile robot applications.

Table 1 shows the usage statuses of single spherical robots operated by different mechanisms, and the dimensions of the existence spaces of their angular velocity vectors. **Figure 1** shows the numbers of rollers N_w arranged per wheel in each mechanism, and their contact types.

In the ACROBAT-S [1] mechanism with $N_w = 2$, the caster of each sphere is driven by two roller drives (see **Figure 1** (a)). The wheel chair mechanism [2] has three spheres (**Figure 1** (b)). The rollers are arranged on the equator, generate an angular velocity vector on the horizontal plane, and can move in all directions. The angular velocity vector of the sphere has two-dimensional freedom. This situation is theoretically considered in [3]. In the ball dribbling mechanism [4], the rollers are arranged in the upper hemisphere, where they hold the balls by friction (See **Figure 1** (c)).

Among the three-roller cases ($N_w = 3$), OWMPs [5] deployed in highway maintenance move the spheres within three constrained rollers (See **Figure 1** (d)).The rollers are arranged on the equator parallel to the horizontal plane, and the sphere can be rotated in all directions by generating an angular velocity vector on the

.

plane (note that the existence space of the angular velocity vector is two-dimensional).

A ball balanced robot [6] has three unconstrained rollers (omni-rollers) arranged in a regular triangular configuration (See **Figure 1** (e)). CPU-Ball Bot [7] has four unconstrained rollers fixed in a square configuration (See **Figure 1** (f)). Such three-dimensional freedom of the angular velocity vector of a sphere will improve the motion diversity of the sphere. However, a mechanism adapting two constrained rollers is suitable in a spherical object conveyance. It is desired to transport with high kinetic energy efficiency.

Table 1 Existence space of angular velocity vector for sphere mobile robot

Mechanism	statuses	Rotational Dimension
ACROBAT-S[1]	Caster	2
Wheel chair [2]	Wheel	2
Ball dribbling mechanism [4]	Sphere conveyance	2
OWMPs [5]	Wheel	2
Ball Balanced robot [6]	Wheel	3
CPU-Ball Bot [7]	Wheel	3

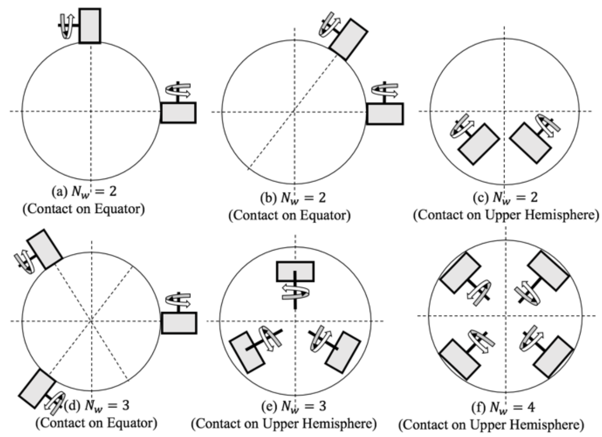


Figure 1 Type of roller arrangement for sphere mobile robot

In this study, we optimize the total kinetic energy of two rollers. Section 2 calculates the angle of the rotational axis of the sphere, and theoretical formula of minimizes the sum of the kinetic energies of the two rollers. Section 3 presents a simulation of total kinetic energy, and Section 4 summarizes the results and suggests ideas for future work.

2. Total kinetic energy

2. 1 Angular velocity vector of the sphere

The center $\boldsymbol{O}$ of a sphere with radius r is fixed as the origin of the coordinate system $\Sigma - xyz$. The i^{th} constraint roller (i = 1 or 2) is in point contact with the sphere at a position vector $\boldsymbol{P_i}$ and is arranged such that the center of mass of the roller $\boldsymbol{P_i}$ and $\boldsymbol{O}$ are on the same line. $\boldsymbol{\omega}$ denotes the angular velocity vector of the sphere. $\boldsymbol{\eta_i}$ denotes the unit vector along the rotational axis of constraint roller. sphere direction φ ($0° \le \varphi < 360°$) is the angle from x-axis.

Now, given the sphere mobile velocity $\boldsymbol{V}$ (the center velocity of sphere)

$$\boldsymbol{V} = \|\boldsymbol{V}\|[\cos\varphi \quad \sin\varphi \quad 0]^T \tag{1}$$

$\acute{\boldsymbol{\omega}}$ which is perpendicular to $\boldsymbol{V}$ is determined as follow.

$$\acute{\boldsymbol{\omega}} = \frac{\|\boldsymbol{V}\|}{r}[-\sin\varphi \quad \cos\varphi \quad 0]^T \tag{2}$$

And. $\acute{\boldsymbol{\omega}}$ is orthogonal projection of $\boldsymbol{\omega}$ with respect to $xy-$ plane.The angle of sphere rotational axis ρ ($-90° \le \rho \le 90°$) is the angle between $\boldsymbol{\omega}$ and $\acute{\boldsymbol{\omega}}$. Therefore, $\boldsymbol{\omega} = [\omega_x, \omega_y, \omega_z]^T$ has one-dimensional of freedom with parameter ρ ($\omega_z = \|\boldsymbol{V}\| \tan\rho / r$).

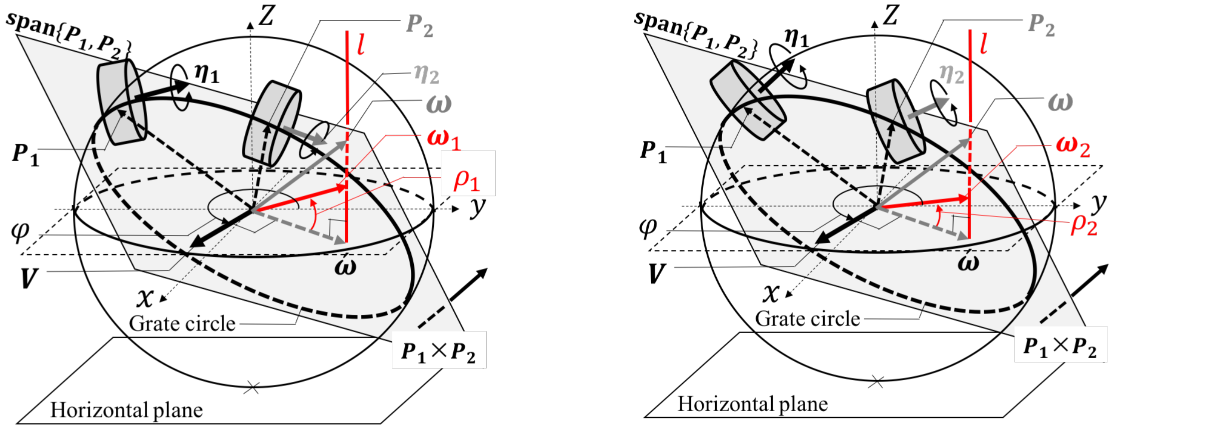
$$\boldsymbol{\omega} = \frac{\|\boldsymbol{V}\|}{r}[-\sin\varphi \quad \cos\varphi \quad \tan\rho]^T \tag{3}$$

2. 2 Kinetic energy of the two rollers

Consider two rollers with radius R, moment I, and roller's angular velocity ω_i at $\boldsymbol{P_i}$. The summed kinetic energy of the rollers is given by Eqs.(4).

$$\begin{aligned} E &= I(\omega_1^2 + \omega_2^2) \\ &= \frac{I}{R^2}(\|\boldsymbol{\omega} \times \boldsymbol{P}_1\|^2 + \|\boldsymbol{\omega} \times \boldsymbol{P}_2\|^2) \end{aligned} \tag{4}$$

Using the angular velocity vector of a sphere, we now derive the kinetic energy of **Type-I** (with both rotational



(a) (b)

Figure. 2 Existence of the sphere angular velocity vector with respect to sphere movile velocity.(a) $\boldsymbol{\omega_1}$ is determined ($\boldsymbol{\omega_1}$ is lay on intersection $\mathbf{span}\{\boldsymbol{P_1},\boldsymbol{P_2}\}$ and line $\boldsymbol{l}$. (b) $\boldsymbol{\omega_2}$ is determined as minimizes the summed kinetic energies of the two rollers.

axes fixed on the same plane) and **Type-II** (with variable rotational axes) configurations(indicated in **Figure 2** (a)(b)).

(A) Case of Type- I

As shown in **Figure 2** (a), the end point of $\boldsymbol{\omega_1}$ can be determined as the intersection of l (the line perpendicular to the horizontal plane passing through end point $\acute{\boldsymbol{\omega}}$) and $\mathbf{span}\{\boldsymbol{P_1},\boldsymbol{P_2}\}$. Therefore, The angle of sphere rotational axis is determined as $\rho = \rho_1$.

$$\rho_1 = \tan^{-1}\left[\frac{(\boldsymbol{P_1}\times\boldsymbol{P_2})_x \sin\varphi - (\boldsymbol{P_1}\times\boldsymbol{P_2})_y \cos\varphi}{(\boldsymbol{P_1}\times\boldsymbol{P_2})_z}\right] \tag{5}$$

Substituting Eqs.(5) into Eqs.(3), we obtain $\boldsymbol{\omega} = \boldsymbol{\omega_1}$ as Eqs.(6).

$$\boldsymbol{\omega_1} = \tag{6}$$

$$\frac{\|\boldsymbol{V}\|}{r}\left[-\sin\varphi, \cos\varphi, \frac{(\boldsymbol{P_1}\times\boldsymbol{P_2})_x \sin\varphi - (\boldsymbol{P_1}\times\boldsymbol{P_2})_y \cos\varphi}{(\boldsymbol{P_1}\times\boldsymbol{P_2})_z}\right]^T$$

Substituting Eqs.(6) into Eqs.(4), we obtain $E = E_1$ as Eqs.(7).

$$E_1 = \frac{I}{R^2}\left(\|\boldsymbol{\omega_1}\times\boldsymbol{P}_1\|^2 + \|\boldsymbol{\omega_1}\times\boldsymbol{P}_2\|^2\right) \tag{7}$$

(B) Case of Type-II

As shown in **Figure 2** (b), we determine $\rho = \rho_1$, $\boldsymbol{\omega} = \boldsymbol{\omega_2}$ that minimizes the summed kinetic energies of the two rollers, and calculate the minimum energy ($E = E_2$).

To this end, we first express $\boldsymbol{\omega}$ as the sum of $\acute{\boldsymbol{\omega}}$ and $\omega_z \boldsymbol{e_3}$.

$$\boldsymbol{\omega} = \acute{\boldsymbol{\omega}} + \omega_z \boldsymbol{e_3} \tag{8}$$

where

$$\boldsymbol{e_3} = [0, 0, 1]^T,\ \ \acute{\boldsymbol{\omega}} = [\omega_x, \omega_x, 0]^T \tag{9}$$

And.

$$\begin{aligned}\boldsymbol{\omega}\times\boldsymbol{P}_i &= (\acute{\boldsymbol{\omega}} + \omega_z \boldsymbol{e_3})\times\boldsymbol{P}_i \\ &= \acute{\boldsymbol{\omega}}\times\boldsymbol{P}_i + \omega_z(\boldsymbol{e_3}\times\boldsymbol{P}_i)\end{aligned} \tag{10}$$

$$\begin{aligned}&\|\boldsymbol{\omega}\times\boldsymbol{P}_i\|^2 \\ &= \langle \acute{\boldsymbol{\omega}}\times\boldsymbol{P}_i + \omega_z(\boldsymbol{e_3}\times\boldsymbol{P}_i), \acute{\boldsymbol{\omega}}\times\boldsymbol{P}_i + \omega_z(\boldsymbol{e_3}\times\boldsymbol{P}_i)\rangle \\ &= \omega_z^2\|\boldsymbol{e_3}\times\boldsymbol{P}_i\|^2 + 2\omega_z\langle\acute{\boldsymbol{\omega}}\times\boldsymbol{P}_i, \boldsymbol{e_3}\times\boldsymbol{P}_i\rangle + \|\acute{\boldsymbol{\omega}}\times\boldsymbol{P}_i\|^2\end{aligned} \tag{11}$$

Using, Eqs.(11), it is represented as quadratic function with respect to ω_z as Eqs.(12)

$$\begin{aligned}&\|\boldsymbol{\omega}\times\boldsymbol{P}_i\|^2 + \|\boldsymbol{\omega}\times\boldsymbol{P}_i\|^2 = \\ &(\|\boldsymbol{e_3}\times\boldsymbol{P}_1\|^2 + \|\boldsymbol{e_3}\times\boldsymbol{P}_2\|^2)\omega_z^2 \\ &\quad + 2(\langle\acute{\boldsymbol{\omega}}\times\boldsymbol{P}_1, \boldsymbol{e_3}\times\boldsymbol{P}_1\rangle + \langle\acute{\boldsymbol{\omega}}\times\boldsymbol{P}_2, \boldsymbol{e_3}\times\boldsymbol{P}_2\rangle)\omega_z \\ &\quad + \|\acute{\boldsymbol{\omega}}\times\boldsymbol{P}_1\|^2 + \|\acute{\boldsymbol{\omega}}\times\boldsymbol{P}_2\|^2\end{aligned} \tag{12}$$

Hence, focusing the coefficients of Eqs.(12), E takes minimal value E_2 as Eqs.(13).

$$E_2 = (\|\acute{\boldsymbol{\omega}} \times \boldsymbol{P}_1\|^2 + \|\acute{\boldsymbol{\omega}} \times \boldsymbol{P}_2\|^2)\, \frac{I}{R^2} - \frac{(\langle \acute{\boldsymbol{\omega}} \times \boldsymbol{P}_1, \boldsymbol{e_3} \times \boldsymbol{P}_1\rangle + \langle \acute{\boldsymbol{\omega}} \times \boldsymbol{P}_2, \boldsymbol{e_3} \times \boldsymbol{P}_2\rangle)^2}{\|\boldsymbol{e_3} \times \boldsymbol{P}_1\|^2 + \|\boldsymbol{e_3} \times \boldsymbol{P}_2\|^2} \frac{I}{R^2} \quad (13)$$

where

$$\omega_z = -\frac{\langle \acute{\boldsymbol{\omega}} \times \boldsymbol{P}_1, \boldsymbol{e_3} \times \boldsymbol{P}_1\rangle + \langle \acute{\boldsymbol{\omega}} \times \boldsymbol{P}_2, \boldsymbol{e_3} \times \boldsymbol{P}_2\rangle}{\|\boldsymbol{e_3} \times \boldsymbol{P}_1\|^2 + \|\boldsymbol{e_3} \times \boldsymbol{P}_2\|^2} \quad (14)$$

$$\rho_2 = \tan^{-1}\left[-\frac{\langle \acute{\boldsymbol{\omega}} \times \boldsymbol{P}_1, \boldsymbol{e_3} \times \boldsymbol{P}_1\rangle + \langle \acute{\boldsymbol{\omega}} \times \boldsymbol{P}_2, \boldsymbol{e_3} \times \boldsymbol{P}_2\rangle}{\|\boldsymbol{e_3} \times \boldsymbol{P}_1\|^2 + \|\boldsymbol{e_3} \times \boldsymbol{P}_2\|^2}\right] \quad (15)$$

Substituting Eqs.(15) into Eqs.(3), $\boldsymbol{\omega_2}$ as Eqs.(16).

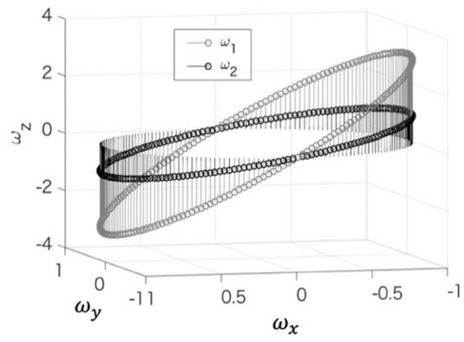
$$\boldsymbol{\omega_2} = \frac{\|\boldsymbol{V}\|}{r}\left[-\sin\varphi, \cos\varphi, -\frac{\langle \acute{\boldsymbol{\omega}} \times \boldsymbol{P}_1, \boldsymbol{e_3} \times \boldsymbol{P}_1\rangle + \langle \acute{\boldsymbol{\omega}} \times \boldsymbol{P}_2, \boldsymbol{e_3} \times \boldsymbol{P}_2\rangle}{\|\boldsymbol{e_3} \times \boldsymbol{P}_1\|^2 + \|\boldsymbol{e_3} \times \boldsymbol{P}_2\|^2}\right]^T \quad (16)$$

3. Simulation

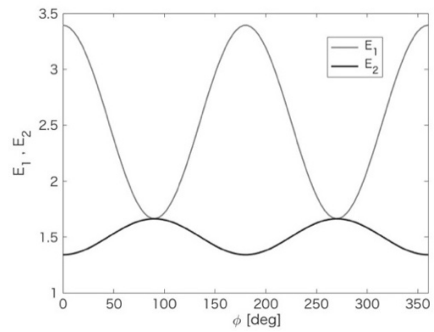
This section presents the simulation results $\boldsymbol{\omega}_1$ (See Eqs.(6)), $\boldsymbol{\omega}_2$ (See Eqs.(16)), E_1 (See Eqs.(7)) and E_2(See Eqs.(13)) with parameter φ ($0° \le \varphi < 360°$) in the given sphere mobile speed: $\|\boldsymbol{V}\| = 1$ [m/s]. The conditions are as follows: $I = 1$, $R = 00.1$[m], $\|\boldsymbol{V}\| = 1$ [m/s], $r = 0.1$[m], $\theta_{1,1} = 215°$, $\theta_{1,2} = 60°$, $\theta_{2,1} = 325°$, $\theta_{2,2} = 60$ °.

3. 1 Trajectory of the end point of the angler velocity vector and Totally kinetic energy

As shown in **Figure 3** (a), the ellipsoid trajectory of $\boldsymbol{\omega}_2$ lies nearer to the xy-plane than that of $\boldsymbol{\omega}_1$, and both trajectories cross a common line parallel to the y-axis.



(a)



(b)

Figure 3 Simulation result. (a) Trajectory of angular velocity vector of sphere. (b) Totally kinetic energy

As shown in **Figure 3** (b), E_1 is minimized at sphere direction angles $\varphi = 90°$ and $270°$, and maximized at $\varphi = 0°$ and $180°$. Meanwhile, E_2 is minimized at $\varphi = 0°$ and $180°$, and maximized at $\varphi = 90°$ and $270°$. And. E_1 and E_2 are same value at $\varphi = 90°$ and $270°$.

4. Conclusion

We optimized the total kinetic energy of the two rollers' movement and calculated the angle of the rotational axis of the sphere, and theoretical formula and we present a simulation of total kinetic energy.

In future work, we plan to evaluate the kinetic energy of the roller arrangement at an arbitrary contact point on the upper hemisphere as an evaluation function.

References

[1] M.Wada, K.Kato, "Kinematic modeling and simulation of active-caster robotic drive with a ball transmission (ACROBAT-S)". 2016 IEEE/RSJ International

Conference on Intelligent Robots and Systems. Daejeon, 2016-12-9/25, IEEE Robotics and Automation Society, pp. 4455-4460, 2016.

[2] S.Ishida, H.Miyamoto,“Holonomic Omnidirectional Vehicle with Ball Wheel Drive Mechanism, 2012The JapanSocietyofMechanicalEngineers.Vol.78,No.790, pp.2162-2170, 2012.

[3] K. Kimura, K. Ishii, Y. Takemura, M. Yamamoto, Mathematical Modeling and motion analysis of the wheel based ball retaining mechanism, *SCIS & ISIS*, pp.4106-4111, 2016.

[4] S. Chikushi, M. Kuwada, et al., Development of Next-Generation Soccer Robot”Musashi150”for RoboCup MSL,*30th Fussy System Symposium*, pp. 624-627,2014

[5] Lee,Y.C,Danny,Lee,D.V.,Chung,J.,andVelinky,S.A.,"Control of a redundant, reconfigurable ball wheel drive mechanism for an omunidirectional platform",Robotica, Cambridge University Press,Vol.25, pp.385-395, 2007.

[6] M. Kumagai, T. Ochiai: “Development of a robot balanced on a ball – Application of passive motion to transport”,Proc. ICRA IEEE(2009), pp. 4106-4111, 2009.

[7] Endo, Tatsuro. et al. "An omnidirectional vehicle on a basketball". 12th International Conference on Advanced Robotics. Seattle, 2005-07-18/20, IEEE Robotics and Automation Society, pp. 573-578, 2005.

Acoustic Impedance Measurement through the Modelling of Ultrasonic Wave Transmission*

Ryuugo Mochizuki
Center for Socio-Robotic Synthesis, Kyushu Institute of Technology, 2-4 Hibikino Wakamatsuku Kitakyushu, 808-0196/Fukuoka, Japan†

Yuya Nishida, Kazuo Ishii
Department of Human Intelligence Systems, Kyushu Institute of Technology, 2-4 Hibikino Wakamatsuku Kitakyushu,808-0196/Fukuoka, Japan
E-mail: mochizuki.ryuugo126@mail.kyutech.jp, y-nishida@brain.kyutech.ac.jp,ishii@brain.kyutech.ac.jp
www.lsse.kyutech.ac.jp

Abstract

In food industry, shortage of workers is a serious problem. Automation of food handling is a critical nowadays. To alleviate the damage during food picking by robotic hand, we propose non-contact acoustic impedance estimation with ultrasonic wave. We have the assumption of the correlation between hardness and acoustic impedance, and, built up ultrasonic transmission model considering attenuation by reflection and absorption, then, made an experiment to estimate the impedance. As the result, we succeeded in detecting acoustic impedance without contact.

Keywords: Acoustic Impedance, Ultrasonic Transmission, Hardness, Reflection, Absorption

1. Introduction

In Japan, the demand for prepared food has been increasing recently. The survey[1] says the expenditure for prepared food has been increasing since 2004 for all human ages. The rate of the increasing expenditure is 283.9[%] for the internet shopping, 28.0[%] for convenience stores. In domestic food industry, a few millions of lunch boxes are produced and consumed a day. Food packing stage requires human resources generally[2]. For alleviation of labor burden by human, automation of food packing is effective. however, picking food by solid hand causes of damages[3]. Hardness of food should be measured without contact, then, picking force should be optimized for the automation.

We assumed relation between acoustic impedance and hardness, then estimated the impedance through analysis of ultrasonic reflection. The impedance shows the difficulty of acoustic transmission in a material, and can be estimated from reflection coefficient[4]. Confirming the relation of acoustic impedance and hardness realizes non-contact hardness measurement system.

2. Related Works

J. Machando et.al introduced ultrasonic wave into Non-Destructive Analysis for evaluation of wood strength[5].

B. Cho and J. Irudayaraj measured cheese depth with ultrasonic wave and accuracy of the measurement was 99.98[%]. Moreover, they recorded more than 0.9 of correlation between cheese mechanic property and result of sonic velocity measurement[6]. S. Srivastava et.al introduced ultrasonic measurement for hardness estimation of tomatoes, and succeeded in detecting hardness changes during the growth[7].

However, to our knowledge, no report has been found as measuring hardness unevenness. The unevenness should be considered because uniqueness of heating food is not guaranteed under cooking.

3. Proposed Method

3.1. *Outline*

Fig.1 shows the process of our acoustic impedance estimation. In the system, receiver and transmitter are attached parallel, which enables observation of reflected wave. Here, the sample is assumed as a composition of multiple media with different acoustic impedances ζ_k[Pa s/m] (k shows index of media), therefore, reflection occurs at z_k where ζ_k changes. After the transmission, we calculate amplitudes of sound pressure $P_{r,k}(z_0)$ in the reflected wave by maxima detection (Here, z_0 denotes observation point.), to easily evaluate the attenuation between reflections. Then, reflection coefficient r_k, and attenuation coefficient α_k, finally ζ_k is estimated from the result of r_k estimation. Evaluation of the correlation between ζ_k and hardness is our future work.

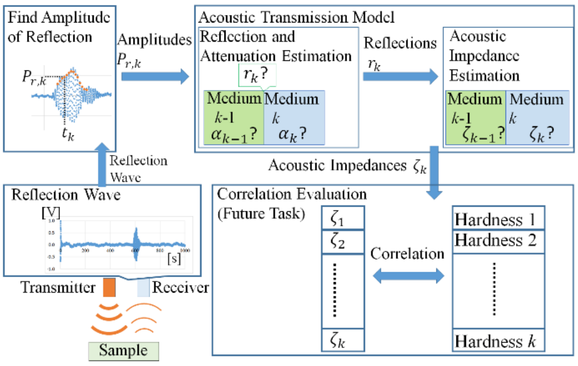


Fig. 1 Process of Acoustic Impedance Calculation

3.2. *Acoustic Transmission Model*

3.2.1. *Condition at media boundary*

In view of acoustic pressure sequence at media boundary z=z_k, the relation among incident wave (Amplitude: $P_{i,k}(z_k)$[Pa]), reflected wave (Amplitude: $P_{r,k}(z_k)$[Pa]), transmitted wave (Amplitude: $P_{t,k}(z_k)$[Pa]) is shown as Eq. (1). Here, wave frequency is ω[rad/s], sonic velocity is c_{k-1}[m/s], c_k[m/s] each other in media k-1 and k.

$$P_{i,k}(z_k)e^{j\omega\left(t-\frac{z-z_k}{c_{k-1}}\right)} + P_{r,k}(z_k)e^{j\left(t+\frac{z-z_k}{c_{k-1}}\right)} = P_{t,k}(z_k)e^{j\omega\left(t-\frac{z-z_k}{c_k}\right)}. \quad (1)$$

3.2.2. *Attenuation by Reflection*

As Fig. 2 shows, samples with uneven hardness is assumed to have multiple media of different ζ_k, thus, reflection occurs at a boundary where difference of ζ_k is observed. As Eq. (2) shows[4], small difference of ζ_k causes of the attenuation by reflection. The acoustic pressure is attenuated to r_k times of $P_{i,k}(z_k)$ after the reflection.

$$r_k = \frac{\zeta_{k-1}-\zeta_k}{\zeta_{k-1}+\zeta_k} = \frac{P_{r,k}(z_k)}{P_{i,k}(z_k)}. \quad (2)$$

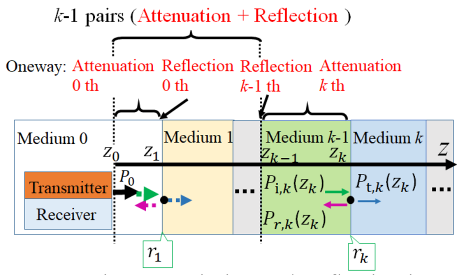


Fig. 2 Acoustic Transmission and Reflection in Media

The reflection coefficient of k-1 from k should be considered by changing the numerator of Eq. (2) into $\zeta_k - \zeta_{k-1}$. As the result, the coefficient is expressed in $-r_k$.

3.2.3. *Attenuation by Absorption*

The attenuation is caused by viscosity of media and heat transmission in the event of longitudinal wave penetration[8], whose magnitude is equal to attenuation coefficient α [Np/m][8-9]. Supposing P_0[Pa] is original acoustic pressure, and, the wave moves z [m], the pressure P diminishes as Eq. (3) shows.

$$P = P_0 e^{-\alpha z}. \quad (3)$$

We disregard wave diffusion[10] because the angle of wave expansion is enough small as our instrument.

3.2.4. *Integration of Reflection and Absorption*

As Fig. 2 shows, when P_0[Pa] is emitted at z=z_0[m] of media 0, penetration and reflection occurs at each boundary z_k. During the transmission in media k, absorption attenuation occurs in α_k[Np/m]. By the arrival at media k (z=z_k), totally, k absorptions and k-1 reflections occur. The state of incident wave (Amplitude: $P_{i,k}(z_k)$) into media k is expressed by Eq. (4).

$$P_{i,k}(z_k)e^{j\omega t} = \left\{P_0 e^{-\alpha_{k-1}(z_k - z_{k-1})} \prod_{k'=1}^{k-1} e^{-\alpha_{k'-1}(z_{k'} - z_{k'-1})}(r_{k'} + 1)\right\} e^{j\omega t} \quad (4)$$

Immediately after the reflection, the acoustic pressure of reflected wave is $P_{r,k}(z_k)$ [Pa]. By the arrival at the receiver (z=z_0), totally, k absorptions and k-1 reflections occur again. $P_{r,k}(z_k)$ is diminished to $P_{r,k}(z_0)$. This state is expressed in Eq. (5). All signs of all r_k s are opposite to Eq. (4). Eq. (4)(5) is summarized to Eq. (6)(7).

$$P_{r,k}(z_0)e^{j\omega t} = \{P_{r,k}(z_k) e^{-\alpha_{k-1}(z_k - z_{k-1})} \prod_{k'=1}^{k-1} e^{-\alpha_{k'-1}(z_{k'} - z_{k'-1})}(-r_{k'} + 1)\} e^{j\omega t}. \quad (5)$$

$$P_{r,k}(z_0)e^{j\omega t} = \{r_k P_0 e^{-2\alpha_{k-1}(z_k - z_{k-1})} \prod_{k'=1}^{k-1}(1 - {r_{k'}}^2) e^{-2\alpha_{k'-1}(z_{k'} - z_{k'-1})}\} e^{j\omega t}. \quad (6)$$

$$\frac{P_{r,k}(z_0)}{P_{r,k-1}(z_0)} = (1 - {r_{k-1}}^2)\frac{r_k}{r_{k-1}} e^{-2\alpha_{k-1}(z_k - z_{k-1})}. \quad (7)$$

3.3. *Solution*

3.3.1. *Amplitude of Reflected Wave $P_{r,k}(z_0)$*

To obtain the Amplitudes, we introduce Gradient method. From intersections with z-axis, the reflection climbs up then descends. In our condition, Large gradient (more than T_ε) should be observed at the intersections where maxima searching starts. the derivative of reflection should be T_ε or less, and the reflection should be T_p or more where candidates of $P_{r,k}(z_0)$ locate.

3.3.2. *Coefficient α_k and r_k*

Though all $P_{r,k}(z_0)$ are solved, unknown α_{k-1}, r_{k-1}, r_k are included in one equation (Eq. (7)). We introduce Monte Carlo Method to find the solutions. The relationship between α_{k-1}, r_{k-1}, r_k is expressed in Eq. (8)
From the Natural Logarithm of Eq. (7).

$$\alpha_{k-1} = \frac{1}{2(z_k - z_{k-1})} \ln\left|(1 - r_{k-1}^2)\frac{r_k}{r_{k-1}}\frac{P_{r,k-1}(z_0)}{P_{r,k}(z_0)}\right|. \quad (8)$$

Here, $\alpha_{k-1} \geq 0$ should be satisfied. $\alpha_{k-1} < 0$ means amplification, which leads a contradiction against attenuation. We treat only positive for r_k, then, necessary requirement of r_k is stated in Eq. (9) for $\alpha_{k-1} \geq 0$.

$$\frac{P_{r,k}(z_0)}{P_{r,k-1}(z_0)}\frac{r_{k-1}}{(1 - r_{k-1}^2)} \leq r_k < 1\,. \quad (9)$$

According to Eq. (9), left side of the inequality exceeds one depending on the combination of $P_{r,k}(z_0)$ and r_k. To solve the contradiction, Eq. (10) should be satisfied.

$$\frac{P_{r,k+1}(z_0)}{P_{r,k}(z_0)}\frac{r_k}{(1 - r_k^2)} < 1. \quad (10)$$

As the result of Eq. (10), (11) is obtained as a solution.

$$0 < r_k < \frac{-P_{r,k+1}(z_0) + \sqrt{P_{r,k+1}(z_0)^2 + 4P_{r,k}(z_0)^2}}{2P_{r,k}(z_0)}. \quad (11)$$

r_k=0, 1 should be excluded so that antilogarithm of Eq. (8) is one or more. The upper limit in Eq. (11) is considered so as not to make contradiction for r_{k+1}. Finally, necessary requirement of r_k is defined as the overlapped range of Eq. (9) and (11).

4. Experiment

4.1. *Method*

In this experiment, we confirmed possibility of finding ζ_k through reflected wave analysis. In the environment (Fig. 3), transmitter and receiver were attached parallel. Sponges were prepared to remove echoes. The distance between transmitter and sample surface is L_{ss}[mm]. As the setting, L_{ss}=100[mm], frequency ω=800π[krad], sample width W_s = 100[mm], depth H_s=10[mm] (We selected ABS). We emitted ultrasonic wave and analyzed reflected wave by Monte Carlo Method generating random values of r_k within the range Eq. (9), (11), and obtained α_{k-1} following Eq. (8). Then we averaged r_k and α_{k-1} for the same k. Acoustic impedance ζ_0 is known to 4.1x10^2[Pa s/m][11], and, true ζ_k of ABS is 2.4x10^6[Pa s/m][12]. N_k , N_{rd} (=500) is the number of amplitudes, r_k subsets (Each subset includes r_1, r_2, .., r_{Nk-1}) each other. Because H_s is sufficiently thin, we ignored the difference of sonic velocity c in each layer. We calculated distance from z_0 to z_k by Eq. (12). t_k is the time to receive $P_{r,k}(z_0)$.

$$z_k - z_0 = \frac{1}{2}c(t_k - t_0)\,. \quad (12)$$

4.2. *Result and Discussion*

Fig. 4 shows the results of measurement. (a) shows $v_r(z_0)$, that is correspondent of reflected wave in voltage, where amplitude is shown in orange. (b)(c)(d) shows average of r_k, α_{k-1} ,ζ_k, each other. In Fig. 5, the same averages ((a)r_k, (b)α_{k-1} , (c)ζ_k,) are expressed in blue bar and the standard deviations are red. ζ_k = 2.5x10^6[Pa s/m] (at z_k=107.5[mm]) was the closest from the truth.

According to Fig. 5, 6, $v_r(z)$, r_k and ζ_k rose, while $\alpha_{k\text{-}1}$ decreased in deeper layer of the sample. From the point, r_k also naturally increases as $v_r(z)$ rises. However, $\alpha_{k\text{-}1}$ is particularly large near the surface. The deviation of $\alpha_{k\text{-}1}$ is outstanding in Fig. 5(b). We estimated r_1 and α_1 without P_o, then, the accuracy of ζ_1 cannot be guaranteed. r_k and ζ_k may change if truly observed value of P_0 is used for the solution. Eq. (2) can be transformed into Eq. (13)

$$\zeta_k = \frac{1+r_k}{1-r_k}\zeta_{k-1} \ . \tag{13}$$

Thus, ζ_k dramatically changes if r_k is near 1.0. Enhancing the accuracy of r_k is essential for higher accuracy of ζ_k.

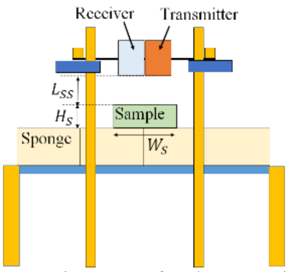
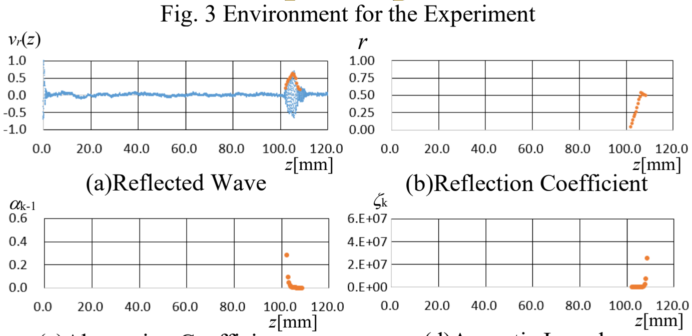


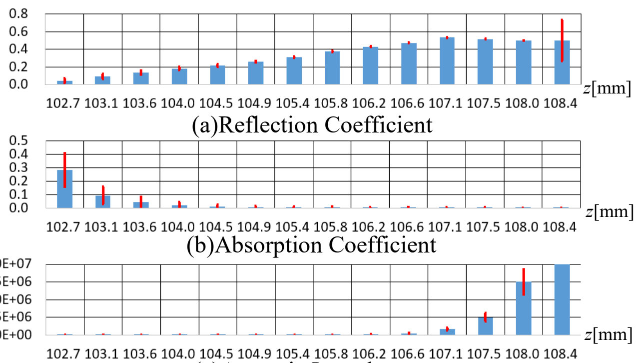
Fig. 3 Environment for the Experiment



(a)Reflected Wave (b)Reflection Coefficient

(c)Absorption Coefficient (d)Acoustic Impedance

Fig. 4 Result of Experiment



(a)Reflection Coefficient

(b)Absorption Coefficient

(c)Acoustic Impedance

Fig. 5 Averages(Blue) and Standard Deviation (Red)

5. Conclusion

In this research, we proposed acoustic impedance measurement method with ultrasonic wave. As the result, reasonable solutions of reflection coefficients were obtained in deeper layer. then accuracy of acoustic impedance estimation was 0.1[Pa s/m] error for 2.4x10^6[Pa s/m] of true ABS acoustic impedance. However, as near the surface, the solution may change according to setting of emitted acoustic pressure. Enhancing reliability in reflection coefficient estimation is necessary for better accuracy of acoustic impedance.

References

1. Yano Research Institute HP (In Japanese) https://www.yano.co.jp/press-release/show/press_id/2245
2. Ministry of Agriculture, Forestry and Fisheries, Summary of the Annual Report on Food, Agriculture and Rural Areas in Japan, MAFF, FY2017, 2017, pp. 17-18.
3. Z. Wang, M. Zhu, S. Kawamura and S. Hirai, Comparison of Different Soft Grippers for Lunch Box Packaging, *Robio*, 4:10, 2017, pp, 1-9.
4. Marine Acoustic Society of Japan, *Kaiyo Onkyo no Kiso to Oyo,* (in Japanese), *Seizando-Shoten* Publishing, 2004.
5. J. Machado, P. Palma, S. Simoes, Ultrasonic Indirect Method for Evaluating Clear Wood Strength and Stiffness, Non-Destructive Testing in Civil Engineering, 2009, pp. 1-6.
6. B. Cho and J. Irudayaraj, A Noncontact Ultrasound Approach for Mechanical Property Determination of Cheeses, *J. Food Sci.*, 68(7), 2003, pp. 2243-2247.
7. S. Srivastava, S. Vaddadi and S. Sadistap, Non-Contact Ultrasonic Based Stiffness Evaluation System for Tomatoes during Shelf-Life Storage, *Int. J. Food Sci. Nutr.*, 4(3), 2014, pp 1-6.
8. The Institute of Electronics, Information and Communication Engineers, *Chishiki* Base (In Japanese), 1(10), Chapter 2, 2014.
9. J. Carlson, J. Deventer, A. Scolan and C. Carlander, Frequency and Temperature Dependence of Acoustic Properties of Polymers Used in Pulse-Echo Systems, *Proceedings in IEEE Symposium on Ultrasonics*, 2003, pp. 885-888.
10. Y. Yokono, Measurement of Ultrasonic Attenuation in Materials, (In Japanese) Keisoku Series, Japan Welding Society, 62(7), 1993, pp. 522-527.
11. S. Takeuchi, *Hajimete no Suichu Cho-onpa* Tranceducer, (In Japanese) *J. Acoust. Soc.* of Japan, 72(5), 2016, pp. 264-272.
12. D. Serrano, A. Uris, S, Ibáñez, C. Rubio, MRI Compativle Planar Material Acoustic Lenses, *Appl. Sci.*, 8(12), 2018, pp. 1-9.

Human-Computer Communication Using Facial Expression

Yasunari Yoshitomi
Graduate School of Life and Environmental Sciences, Kyoto Prefectural University,
1-5 Nakaragi-cho, Shimogamo, Sakyo-ku, Kyoto 606-8522, Japan
E-mail: yoshitomi@kpu.ac.jp
http://www2.kpu.ac.jp/ningen/infsys/English_index.html

Abstract

To develop a complex computer system such as a robot that can communicate smoothly with humans, it is necessary to equip the system with a function for both understanding human emotions and expressing emotional signals. From both perspectives, facial expression is a promising research area. In our research, we have explored both aspects of facial expression using infrared-ray images and visible-ray images and have developed a personified agent for expressing emotional signals to humans.

Keywords: Emotion, Facial expression recognition, Infrared-ray image, Facial expression synthesis, Personified agent.

1. Introduction

The goal of our study is to present a paradigm whereby a complex computer system such as a robot can cooperate smoothly with humans. To do this, the computer system must have the ability to communicate with humans using some form(s) of information transmission. Such a system must be equipped with a function for both understanding human emotions and expressing emotional signals to its human counterparts. In this regard, facial expression is a promising target for research. Accordingly, we have been investigating both aspects of facial expression.

In this paper, we describe the challenges of reaching our goal. The remainder of the paper is organized as follows: Section 2 summarizes our studies on facial expression recognition; Section 3 briefly describes our studies on human-computer-human communication via the Internet; Section 4 outlines our studies on human-computer communication; Section 5 discusses our work on integration with speech; Section 6 concludes the paper.

2. Facial Expression Recognition

2.1. *Infrared-ray image utilization*

We have developed a method for recognizing facial expressions using thermal image processing.[1] In this study, infrared-ray was used. Figure 1 shows the influence of lighting at night on a facial image using both visible-ray ((a),(c)) and infrared-ray ((b),(d)). As is evident in the figure, the visible-ray image is strongly influenced by lighting conditions, while the thermal image is unaffected. With our method, neutral, happy, surprised, and sad facial expressions were recognized with 90% accuracy.[1] Figure 2 shows examples.

2.2. *Sensor fusion*

Sensor fusion is a promising way to improve the recognition accuracy of facial expression or emotion recognition. Several studies[2,3] to improve accuracy using sensor fusion have produced promising results.

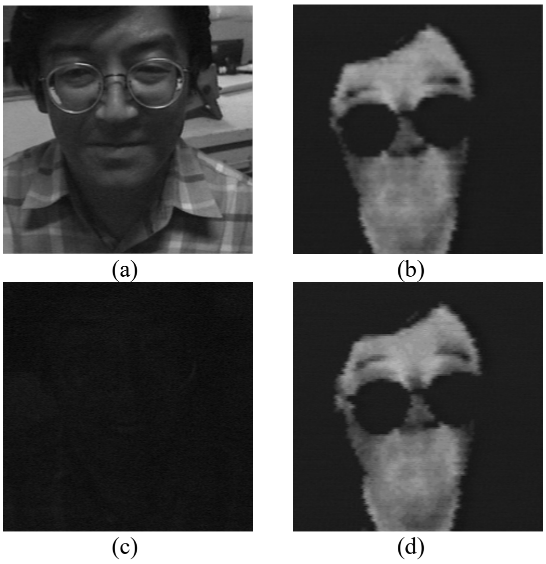


Fig. 1. Examples of face-image at night; (a) visible-ray image with lighting, (b) infrared-ray image with lighting, (c) visible-ray image without lighting, (b) infrared-ray image without lighting.[1]

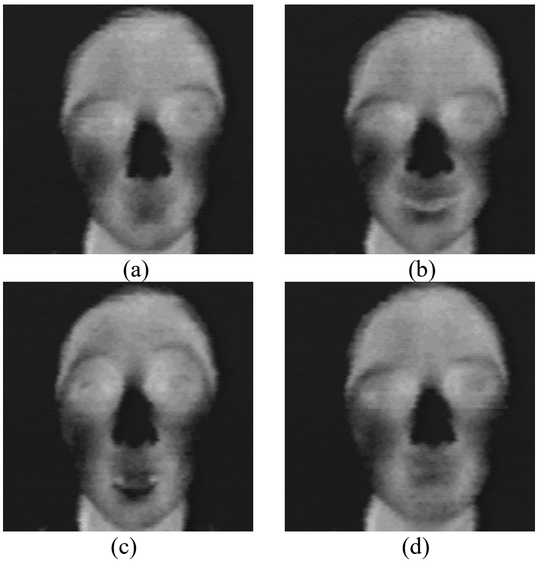


Fig.2. Examples of facial expressions; (a) neutral, (b) happy, (c) surprised, (d) sad.[1]

2.3. *Analysis for medical use*

Facial expression analysis using infrared-ray images[4] or visible-images[5-7] has been successfully conducted with the goal of identifying subjects suffering from pre-stage dementia or other medical problems.

3. Human-computer-human Communication

Social network services (SNSs) have become extremely popular as communication tools on the Internet. However, while it is possible to post a message, a static image, or a moving image on a platform such as Twitter, it is difficult to communicate the actual emotions felt when writing a message or posting an image. We believe that a support system is needed to facilitate smoother communication between humans in their use of SNSs. Not having immediate and direct contact with one another risks misunderstanding, especially from an emotional point of view.

One of our studies is aimed at expressing the real emotions of individuals writing messages for posting on an SNS site by analyzing their facial expressions and visualizing them as pictographs. To this end, we have developed a real-time system for expressing emotion as a pictograph selected according to the writer's facial expression while writing a message.[8,9] We applied the system to the posting on Twitter of both a message and a pictograph.[8,9]

The lower panels in Figure 3 show the output of our system in a situation where the subject was asked to intentionally show two types of emotions—neutral or smiling—when writing the message, '明日は情報伝達システム学サブゼミに参加します。時間は 5 時限目、場所は先生の部屋です。' (in Japanese), which means, "I will attend the discussion section held at the professor's room in the information communication system lab in fifth period tomorrow." [8]

Fig. 3. Snap-shots (upper) of posting on Twitter; messages and pictographs (lower) posted on Twitter (left: neutral; right: smiling).[8]

4. Human-computer Communication

4.1. *Personified agent*

The process of agent generation in our system[10] consists of six steps: (1) creating facial expression data, (2) recording vocal utterances, (3) automatic WAVE file division, (4) speech recognition by Julius[11], (5) insertion of expressionless data, and (6) the creation of facial expression motion.

Expressive motions are generated by combining the expression data of each vowel for each utterance motion. Then, the utterance contents are input as text and used by the MikuMikuDanceAgent (MMDAgent),[12] which is a freeware animation program that allows users to create and animate movies with agents, to output synthesized voice that is then recorded by a stereo mixer inside a PC and saved as a WAVE file. Speech is recognized using a speech recognition system called Julius,[11] followed by facial expression synthesis of the agent using preset parameters depending on each vowel. Facial expression data were created with MikuMikuDance[13]. In this study, in order to generate more human-like agent facial expressions, facial expression data were created for the vowels / a /, / i /, / u /, / e /, and / o / (Fig. 4).[10] In order to create more natural agent facial expressions, processing is then performed to insert a neutral facial expression when the same vowel, for example / a /, is continuous.[10]

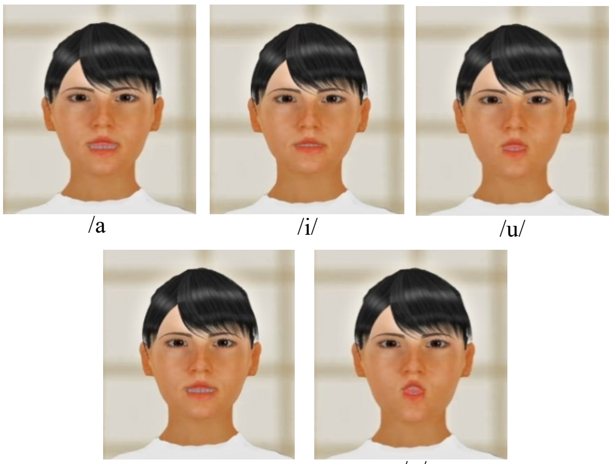


Fig. 4. Facial expression of the agent when uttering each vowel.[10]

4.2. *Human-computer communication in music recommendation*

The music recommendation module of the proposed system[14] is based on a previously proposed system[15] that uses collaborative filtering and impression words (see the paper[15] for details of the music recommendation module).

In the music-recommendation process, all user navigations are performed by the synthetic voice of the agent appearing on the PC screen facing the user. All dialogue spoken by the agent is situationally selected by the proposed system[14]. The user's answers to the questions generated by the agent are recognized using the voice recognition function of the system, and the agent motions, including facial expressions, are then generated.

Figure 5 shows two snapshots of the reaction of the agent after recognizing (a) a positive answer, i.e., the user wishes to listen to the recommended song again in the future, and (b) a negative answer, i.e., the user does not wish to listen to the recommended song in the future. In the case of (a), the agent nods twice and raises the corners of the mouth slightly, while in the case of (b), the agent also nods twice, but lowers the corners of the mouth slightly. Figure 6 shows a snapshot of the music recommendation being performed by the proposed system[14].

(a) (b)

Fig. 5. Snapshots of the reaction of the agent after recognizing (a) a positive answer, and (b) a negative answer. [14]

Fig. 6. Snapshot of performing song-recommendation by the proposed system.[14]

5. Integration with Speech

Utterance judgment is necessary for deciding the timing of facial expression recognition. Moreover, the mouth shape with or without an utterance influences facial expression. In our studies, the first and last vowels in an utterance such as a name were recognized for deciding the timing of the facial expression recognition.[16,17]

Speech recognition and synthesis are indispensable for human-computer communication. In particular, developing the function of emotional speech synthesis[18] offers a way to create a paradigm whereby a computer system such as a robot can work seamlessly with humans.

6. Conclusions

To develop a complex computer system such as a robot that can communicate smoothly with humans, it is necessary to equip the system with the ability to both understand human emotion and express emotional signals to humans. From both points of view, facial expression is a promising research field. In developing a method for recognizing facial expressions, we have used infrared-ray images as well as visible-ray images. For expressing emotional signals to humans, we have developed a personified agent. Developing the function of emotional speech synthesis is the next target of our studies.

Acknowledgements

We would like to thank Dr. J. Narumoto, Professor of the Kyoto Prefectural University of Medicine, Dr. M. Tabuse, Professor of the Kyoto Prefectural University, and Dr. T. Asada, Associate Professor of the Kyoto Prefectural University, for their valuable cooperation during the course of this research.

References

1. Y. Yoshitomi, N. Miyawaki, S. Tomita and S. Kimura, Facial expression recognition using thermal image processing and neural network, *in Proc. of 6th IEEE Int. Workshop on Robot and Human Communication*, (Japan, Sendai, 1997), pp. 380-385.
2. Y. Yoshitomi, S. Kim, T .Kawano and T. Kitazoe, Effect of sensor fusion for recognition of emotional states using voice, face Image and thermal Image of face, *in Proc. of of 9th IEEE Int. Workshop on Robot and Human Interactive Communication*, (Japan, Takamatsu, 2000), pp.178-183.
3. Y. Oka, Y. Yoshitomi, T. Asada, and M. Tabuse, Emotion recognition of a speaker using facial expression intensity of thermal image and utterance time, *J. Robotics, Networking and Artif. Life* **3**(3) (2016) 148-151.
4. Y. Yoshitomi, T. Asada, R. Kato, and M.T abuse, Method of facial expression analysis using video phone and thermal image, *J. Robotics, Networking and Artif. Life* **1**(1) (2014) 7-11.
5. T. Asada, Y. Yoshitomi, R. Kato, M. Tabuse, and J. Narumoto, Quantitative evaluation of facial expressions and movements of persons while using video phone, *J. Robotics, Networking and Artif. Life* **2**(2) (2015) 111-114.
6. T. Asada, Y. Yoshitomi, A. Tsuji, R. Kato, M. Tabuse, N. Kuwahara, and J.Narumoto, Facial expression analysis while using video phone, *J. Robotics, Networking and Artif. Life* **2**(4) (2016) 258-262.
7. R. Shimada, T. Asada, Y. Yoshitomi, and M. Tabuse, Real-time system for horizontal asymmetry analysis on facial expression and its visualization, *J. Robotics, Networking and Artif. Life* **6**(1) (2019) 7-11.
8. Y. Yoshitomi, T. Asada, K. Mori, R. Shimada, Y. Yano, and M. Tabuse, Facial expression analysis and its visualization while writing messages, *J. Robotics, Networking and Artif. Life* **5**(1) (2018) 37-40.
9. T. Asada, Y. Yano, Y. Yoshitomi, and M. Tabuse, A system for posting on SNS portrait selected using facial expression analysis while writing message, *J. Robotics, Networking and Artif. Life* **6**(3) (2019) 199-202.
10. T. Asada, R. Adachi, S. Takada, Y. Yoshitomi, and M. Tabuse, Facial expression synthesis using vowel recognition for synthesized speech, *in Proc. 2020 Int. Conf. on Artificial Life and Robotics*, ed. M. Sugisaka (Japan, Beppu, 2020), pp.398‐402.
11. Julius, http://Julius.osdn.jp/, Accessed 24 November 2020.
12. MMDAgent, http://www.mmdagent.jp/ Accessed 24 November 2020.
13. MikuMikuDance, https://sites.google.com/view/vpvp/, Accessed 29 November 2020.
14. A. Matsui, M. Sakurai, T. Asada, and M. Tabuse, Music recommendation system driven by interaction between user and personified agent using speech recognition, synthesized voice and facial expression, *in Proc. 2021 Int. Conf. on Artificial Life and Robotics*, ed. M. Sugisaka (Japan, 2021), in press.
15. S. Yoshizaki, Y. Yoshitomi, C. Koro, and T. Asada, Music recommendation hybrid system for improving recognition ability using collaborative filtering and impression words, *J. Artif. Life and Robotics* **18**(1-2) (2013) 109–116.
16. Y. Yoshitomi, T. Asada, K. Shimada, and M. Tabuse, Facial expression recognition of a speaker using vowel judgment and thermal image processing, *J. Artif. Life and Robotics* **16**(3) (2011) 318–323.
17. T. Fujimura, Y. Yoshitomi, T. Asada, and M. Tabuse, Facial expression recognition of a speaker using front-view face judgment, vowel judgment and thermal image processing, *J. Artif. Life and Robotics* **16**(3) (2011) 411-417.
18. R. Makino, Y. Yoshitomi, T. Asada, and M. Tabuse, Speech synthesis of emotions in a sentence using vowel features, *J. Robotics, Networking and Artif. Life* **7**(2) (2020) 107-110.

Detection of a Fallen Person and Estimation of the Head Position from UAV Images

Haruka Egawa, Seiji Ishiakwa, Joo Kooi Tan

Department of Mechanical and Control Engineering, Kyushu Institute of Technology, Sensui-cho Tobata-ku Kitakyushu, 804-8550, Japan
E-mail: {egawa-haruka , etheltan }@ss10.cntl.kyutech.ac.jp
http://lab.cntl.kyutech.ac.jp/~etheltan/

Abstract

In order to search for victims in the event of a disaster, we propose a method of detecting a fallen person from UAV images and estimating its head position. Rotation-invariant HOG and Rotation-invariant LBP are used, so that they may detected a fallen person regardless of its body orientation. In addition, the position of the head of a fallen person is estimated using the peak of the gradient histogram. Experimental results show satisfactory performance of the proposed method.

Keywords: *Aerial photographs,* Fallen persons, Rotation-invariant HOG, Rotation-invariant LBP, *Random Forest*, *Gradient histogram.*

1. Introduction

Japan is one of the world's leading countries of earthquakes [1] and has been hit by disasters caused by earthquakes. In the event of such a disaster, it is often very difficult to search for the victims manually. Therefore, recently, attention has been paid on the development of an automatic search method for those victims from aerial images taken by drones.

There are various human detection methods for automatic search [2,3,4], but most of them are based on the premise that the target person's body orientation is unified and the subject is upright [2,3]. Exists. However, there is no uniformity in the orientation of the person's body taken with the drone. Therefore, in this paper, we propose an automatic detection method for a fallen person that does not depend on the body orientation of the detection target. In this method, after detecting a fallen person [5], his/her head position is estimated and detected. This enables quick support of the fallen person for communication, medical care, food, etc.

The proposed method uses Rotation-invariant Histogram of Oriented Gradients (Ri-HOG) features [6] and Rotation-invariant Local Binary Pattern (Ri-LBP) as features that are robust to rotation in order to represent a fallen person.) Use the feature quantity [7]. In addition, Random Forest [8] employed as a classifier. In addition, the head position is estimated using the peak value of the gradient histogram in the circular region.

The proposed method is examined its performance by the experiment using bird-eye view videos.

2. Method

2.1. *Ri-HOG features*

The Ri-HOG feature is a feature of HOG [9] with rotational invariance. Unlike the original HOG, which uses rectangular cells, the cell arrangement obtained by dividing concentric circles is used in Ri-HOG. This section describes the Ri-HOG feature.

2.1.1. *Calculation of luminance gradient information*

The input color image is gray-scaled and the brightness gradient at each pixel is calculated. From the obtained luminance gradient, the luminance gradient intensity and the luminance gradient direction in the Cartesian coordinate representation are obtained.
Assuming that the origin O is at the center of an input image, the difference between the declination $\varphi(x, y)$ of a pixel position (x, y) provided by the polar coordinate representation and the obtained brightness gradient direction $\theta(x, y)$ of the pixel defines the brightness gradient direction θ' in the polar coordinate representation . $\theta'$is defined by

$$\theta'(x, y) = \theta(x, y) - \varphi(x, y) \qquad (1)$$

2.1.2. *Creating a 2-D histogram*

In the proposed method, a concentric circle consisting of three large, medium and small circles is divided into 36 in the angular direction, and each small area is defined as a cell. The cell numbers of the small concentric circle are given as 1,4,7···,106 clockwise from the start point, starting from the area in the 0 ° direction. Similarly, the cell numbers of the middle circle are 2,5,8, ···, 107, and the cell numbers of the largest circle are 3,6,9, ···, 108.
The offset regions $\mathrm{W}_1\ (p), \mathrm{W}_2\ (p), \cdots, \mathrm{W}_6\ (p)$ are further set by dividing the semicircular region existing in the radial direction of the pixel of interest $p = (x, y)$ into six parts. Let q be the pixel in the offset region, and, from pixels p and q, a 2-D histogram is calculated to use the co-occurrence information of their luminance gradient directions θ'. Six 2-D histograms are created with each cell as it has 6 offsets. The 2-D histogram is calculated using the following formula;

$$\mathrm{H}_{ij}(o_1, o_2) = \sum_{\boldsymbol{p}\in \mathrm{S}_i} \sum_{\boldsymbol{q}\in W_j(\boldsymbol{p})} \big(m(\boldsymbol{p}) + m(\boldsymbol{q})\big) K(\theta'(\boldsymbol{p}) \in o_1) K(\theta'(\boldsymbol{q}) \in o_2) \qquad (2)$$

Here, i is the cell number, j is the offset number, o_1 is the luminance gradient direction of pixel p after quantization, and o_2 is that of the offset pixel q after quantization. S_i is the cell of present concern, and $K()$ is the function whose value is 1, if the argument is true, and 0, otherwise.

2.1.3. *Normalization*

In order to reduce the influence of changes in local brightness and contrast, normalization is performed for each 2-D histogram obtained by Eq. (2).
Since the values of luminance gradient direction θ of the entire circular region in the Cartesian coordinate representation range from 0 ° to 360 °, it is quantized in 36 directions every 10 ° interval. A gradient histogram is created by adding the luminance gradient intensities in the circular region based on the luminance gradient direction in the quantized Cartesian coordinate representation. After that, the cells are rearranged in the clockwise order from the reference direction based on the angle at which the gradient histogram takes a peak, and the vector in which the 2-D histograms of each cell are connected is used as the Ri-HOG feature quantity. Here, the cell numbers of the smallest concentric circle are expressed as follows;

$$i_u = 1,4,7, \cdots, N \ (u = 1,2,3, \cdots 36) \qquad (3)$$

Using Eq.(3), the cells are connected as $\mathrm{i}_1, \mathrm{i}_1 + 1, \mathrm{i}_1 + 2, \mathrm{i}_2, \mathrm{i}_2 + 1, \mathrm{i}_2 + 2, \mathrm{i}_3 \cdots, \mathrm{i}_{36} + 2$, if the peak of the gradient histogram is 0 °, and they are connected as $\mathrm{i}_{36}, \mathrm{i}_{36} + 1, \mathrm{i}_{36} + 2, \mathrm{i}_1, \cdots, \mathrm{i}_{34} + 2, \mathrm{i}_{35}, \mathrm{i}_{35} + 1, \mathrm{i}_{35} + 2$, if the peak of the gradient histogram is 350 °.

2.2. *Ri-LBP features*

The Ri-LBP feature is a feature obtained by adding rotational invariance to the Local Binary Pattern (LBP) [10]. This section describes the Ri-LBP feature.

2.2.1. *Local binarization*

The input image is gray-scaled, and, as shown in **Fig. 1**, the brightness values of the pixel of interest and its 8 adjacent pixels are compared. If the gray value of an adjacent pixel is larger than or equal to that of the center pixel, the adjacent pixel is given 1, otherwise given 0. By aligning the obtained 0 or 1 in the direction of the arrow shown in Fig. 1 (b), an 8-digit binary number is obtained, which represents the center pixel.
In Ri-LBP, the starting point is changed in turn to acquire eight 8-digit binary numbers, and the minimum value out of the 8 numbers represents the center pixel.
For example, in Fig. 1, the normal LBP value is $0100\ 0011_b = 67_d$, but the Ri-LBP value is $0000\ 1101_b = 13_d$.

1	5	4
8	5	2
7	2	3

(a)

0	1	0
1		0
1	0	0

(b)

Fig. 1. LBP

2.2.2. *Histogram creation*

The cell of the Ri-LBP feature is the same as the cell of Ri-HOG feature.

Since LBP is given as an 8-digit binary number, there are 256 values from 0 to 255, but in Ri-LBP, there is an 8-digit binary number that becomes the same value by changing the start point. For example, $1000\,0000_b = 128_d$ in LBP is expressed as $0000\,0001_b = 1_d$ when the start point is changed. Hence $1000\,0000_b$ and $0000\,0001_b$ are equivalent in Ri-LBP. After all, the total number of Ri-LBP values is 36.

Since there are cases where the number of pixels in one cell is less than 36, the proposed method divides 36 patterns into 9 patterns and creates a histogram with each cell. This histogram is normalized. The cells are rearranged based on the reference angle that gives the peak in the brightness gradient direction obtained in Section 2.1.3, and the histograms of all the cells are concatenated. This provides the Ri-LBP feature.

2.3. *Detection of a fallen person*

The proposed method uses Random Forest as a discriminator. The features of Random Forest include that it is possible to learn efficiently even with high-dimensional features by random learning, and that the influence of noise in the teacher signal can be suppressed by random selection of learning data.

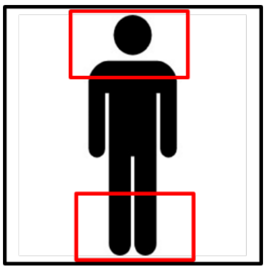


Fig. 2. Example of presumed head positions in the fallen person detection window.

3. Head position estimation

The head position of a fallen person is estimated using the angle at which the gradient histogram in the circular region obtained in Section 2.1.3 has a peak. When a fallen person exists in the detection window in the orientation shown in **Fig. 2**, the horizontal gradient becomes large and the peak angle of the gradient histogram is 0 ° or 180 °. Therefore, the areas shown by the red frame in Fig. 2 are the estimated head position.

4. Experiment

In the experiment, an image of a fallen person is used as an image of the Positive class of the image database. Images other than people are used for Negative class images. The INRIA Person Dataset [11] was also used as an image of the Negative class. A classifier is constructed using these training images. The size of the training image is all 61 pixels vertically and 61 pixels horizontally.

4.1. *Experiment 1*

The first experiment is the detection of a fallen person from an image. Four types of moving images taken with a multicopter are used in the experiment. In all of these videos, there is only one fallen person to be detected. In Experimental Video 1, there are some obstacles other than a fallen person. In the experimental video 2 and the experimental video 3, the amount of movement of the camera is large, and there are some frames in which there is no fallen person in the image. In addition, there is an upright person who is not fallen in the experimental video 4.

The detection is evaluated with each detection window using Intersection over Union (IoU). The threshold of the IoU is experimentally set to 0.5. As a result of the experiment, the recall was 0.990 for movie 1, 0.855 for movie 2, 0.565 for movie 3, and 0.875 for movie 4.

An example of the experimental results is shown in **Fig. 3**. The green rectangle is the manually set Ground Truth, whereas the red rectangle shows a detected fallen person

4.2. *Experiment 2*

In the second experiment, head of a fallen person is searched on the image of a detected fallen person. For the experiment, 137 images of the fallen person area cut out from the video taken by the multicopter are used.

Fig. 3. Example of detection result of Experiment 1 .

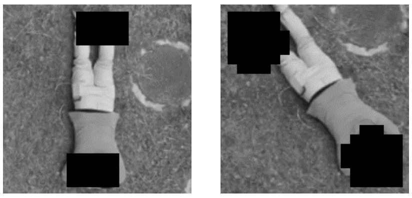


Fig. 4. Example of the estimated head position in Experiment 2

Fig. 5. The result of head detection with a fallen person

In order to visualize the result, the estimated head position is painted black. The case where the estimated head position overlaps with the head of the actual lying person is regarded as true. As the result of the experiment, the correct answer rate was 73.7%. An example of the experimental results is shown in **Fig. 4**.

Fig. 5 shows the results of head detection. It is indicated by a small light-blue square.

5. Conclusion

In this paper, we proposed a method of acquiring aerial images with a camera mounted on a multicopter and detecting a fallen person from the images.

In the experiment, the recall rate of the fallen person detection was about 82% when there was one fallen person in the image frame. The correct answer rate for head position estimation using the peak angle of the gradient histogram was 73.7%. In the future, we aim at further improvement of the detection rate by increasing the number of learning samples and detecting a fallen person with occlusion.

References

1. 平成 25 年版防災白書 付属資料, 内閣府, 2013, p.1.
2. P. Viola, M. Jones: “Rapid object detection using a boosted cascade of simple features”, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2001, pp.511-518.
3. R. Sato, S. Koshimura: “Development of a method for searching survivors in disaster-affected areas with UAV and image analysis”, *J. JSCE*, Ser.B2, Coastal engineering Vol. 69, No. 2,2013, pp.I_1461 - I_1465.
4. A. Gaszczak, T. P. Breckona, J. Hana : “Real-time people and vehicle detection from UAV imagery”, *Proc. SPIE Conference on International Robots and Computer Vision XXVIII:Algorithms and Techniques*, Vol.7878, 2011, pp.78780B-78780B-13.
5. H. Egawa, S. Isihkawa, J. K. Tan: “Detection of a fallen person from UAV images using rotation invariant features”, *Proc. of Biomedical Fuzzy Systems Association*, 2020, pp.1-4(in Japanese).
6. T. Fujimoto, Y. Kameda, I. Matsuda, S. Itoh: “Human detection from top view images using rotation invariant HOG features”, *Proc. of The Institute of Image Information and Television Engineers*, 2014, pp.17_10_1-17_10_2.
7. M. Pietikaine,T. Ojala,Z. Xu: “Rotation-invariant texture classification using feature distribution”, *Pattern Recognition*, Vol.33, 2000, pp. 43-52.
8. L. Breiman: “Random forests”, Machine Learning, Vol.45, No.1, 2001, pp.5-32.
9. N. Dalal, B. Triggs: “Histograms of oriented gradients for human detection”, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp.886-893.
10. T. Ojala, M. Pietikainen, D. Harwood: “A comparative study of texture measures with classification based on featured distributions”, Pattern Recognition, Vol. 29, No. 1, 1996, pp. 51–59.
11. INRIA Person Dataset: （ http://pascal.inrialpes.fr/data/human/(Accessed:10 Jan,2019）

Development of a Pedestrian Crossing Navigation System for a Visually Impaired Person Using MY VISION

Kohei Kitagawa
Department of Mechanical and Control Engineering, Kyushu Institute of Technology, Sensui-cho Tobata-ku Kitakyushu, 804-8550, Japan

Seiji Ishikawa, Joo Kooi Tan
Department of Mechanical and Control Engineering, Kyushu Institute of Technology, Sensui-cho Tobata-ku Kitakyushu, 804-8550, Japan
E-mail: {kitagawa-kohei ,etheltan}@ss10.cntl.kyutech.ac.jp
http://lab.cntl.kyutech.ac.jp/~etheltan/

Abstract

In this paper, we propose a system for a visually impaired person to cross a pedestrian crossing safely by the employment of the state of a traffic light and the remaining distance to the other side obtained from the analysis of MY VISION images. The traffic light at a pedestrian crossing is detected by a discriminator using HOG (Histograms of Oriented Gradients) features and Random Forest. The effectiveness of the proposed method was verified by experiments.

Keywords: MY VISION, Visually Impaired, HOG, Random Forest, Traffic light, Pedestrian crossing behavior.

1. Introduction

The 2011 Ministry of Health, Labor and Welfare report writes that there are approximately 310,000 people in Japan who have visual impairment. Looking at the world, it is predicted that the number of visually impaired people worldwide will triple within the next 40 years [1].

Accidents of visually impaired people are not uncommon, and there is a survey result that about 42% of the subjects experienced accidents with injuries while walking outdoors within the past 5 years, especially against automobiles [2]. It is easy to imagine that accidents with cars may often happen at an intersection, which may be one of the difficult points for a visually impaired person to cross.

It is based on Braille blocks and acoustic signs that a visually impaired person finds an intersection and judges the timing of crossing it. However, neither all intersections are installed Braille blocks, nor they have acoustic instruments. Therefore, it is necessary to have a navigation system that recognizes and analyzes (pedestrian) traffic lights, so that a visually impaired person may cross the intersection safely.

As a traffic light detection method, a method using template matching and a method utilizing position information by GPS or a map have been proposed[3-5]. However, all the studies are limited to the detection of traffic lights and they do not analyze possible timing to cross. However, in practice, it is vital for a visually impaired person to receive the information on the timing to cross. It should be realized by the analysis of traffic light images, once it is detected.

In this paper, we propose a method of detecting a pedestrian traffic light at a pedestrian crossing by MY VISION and analyzing it if one can walk into the crossing now. MY VISION is a self-virtual viewpoint employing a body (or a head) mounted camera. It can be a virtual eye of a visually impaired person or the third eye of a person who is careless ahead. In the proposed method, the traffic light is detected employing the images captured by a camera attached to the head of a person. The HOG feature is used for describing a traffic light. The traffic light is recognized by a Random Forest classifier, which is a multi-class classifier. Then pattern analysis is performed with the traffic light in the image sequence, and the possibility is judged if one may cross the frontal pedestrian crossing.

2. Traffic Light Detection

In this section, we describe detection of traffic lights, extraction of target object candidate regions, feature extraction using HOG features, and learning and identification by Random Forest.

2.1. *Extraction of a signal candidate area*

In the proposed method, instead of searching for a traffic light in a whole input image, it is searched in the upper half of the input image after converting it to a HSV image, which is close to human color recognition[6]. A candidate area of a traffic signal is extracted by searching for the signal color (red or green) in the upper half image. The reason the upper half of an image is considered is that the traffic light normally exists only in the upper half area of a self-virtual viewpoint image.

Once the candidate areas of a traffic signal are extracted, the areas receive expansion processing, since the color area of the traffic light may be extracted smaller than it actually is. After the expansion, labeling is performed to the expanded areas and those noisy areas are removed using a certain threshold value. Thus a candidate area of a traffic light is obtained.

2.2. *Histogram of Oriented Gradients (HOG)*

The HOG feature[7] is used in the proposed method as a feature for detecting a traffic light, as it is robust to lighting changes, misalignment, and rotation. The gradient direction is quantized in four directions to accommodate angular changes.

2.3. *Learning and identification by Random Forest*

In the method, Random forest[8], which is capable of multi-class discrimination, is employed for a classifier. Random forest is possible to learn and identify multiple target objects at the same time. By using it, we can improve generalization ability for unknown data in preparation for future multiclass classification..

3. Crossing Analysis

In this section, the method is explained for analyzing the situation of a traffic light and determining the timing of crossing. The analysis is performed based on the pedestrian crossing behavior principle hypothesis[9].

3.1. *The hypothesis on pedestrian crossing behavior principle*

The hypothesis on pedestrian crossing behavior principle has been proposed, which gives insight into the relationship between pedestrian signals and pedestrian behavior.

The following four empirical patterns are considered in the hypothesis based on the situation of the traffic light.

1. Experience on the change of the light from red to blue: The case where the length of the blue light is perceptible and it is considered to be the safest to cross.
2. Experience on only blue light display: The case where the length of the blue light display cannot be perceived and it is unknown when it will turn to a blinking signal. Depending on the walking speed of a person, it may be dangerous.
3. Experience on the change from blue to blinking blue: This is the case where the length of the blue blinking time is perceptible, and a person on cross can finish crossing.
4. Experience on only blinking blue light. The length of the blue blinking time cannot be perceived, and it is not possible to cross.)

Based on this crossing behavior principle hypothesis, pattern identification and analysis of a traffic light are performed.

3.2. *Input of walking status*

In the proposed method, crossing analysis takes the walking condition of a user into account. The information on what kind of walking situation a user is in is collected in advance.

The followings are two patterns of walking situation with a user.

A) Walks normally.
B) Cannot walk normally.

Once either of the above walking conditions is entered before using the system, it is taken into consideration when the pattern 2 in 3.2 is identified.

3.3. *Discrimination of traffic lights*

This section describes how to discriminate the color and blinking of a traffic light.

3.3.1. *Color discrimination*

The color of the traffic light is determined based on the HSV model. The color of the signal (red/blue) is identified by setting certain thresholds on H, S and V values and counting the number of pixels of each color in the search window.

3.3.2. *Identification of blinking signal*

The blinking cycle of the traffic light is used to determine whether the signal is blinking or it is the frame a traffic light cannot be detected. The blinking cycle of the traffic light is 0.5 seconds in Japan. If this cycle is satisfied, it is determined as a blinking signal, and if it is any other cycle, it is not detected.

3.4. *Pattern discrimination on traffic signals*

Based on the hypothesis on pedestrian crossing behavior principle explained in Section 3.1, pattern discrimination is performed in the sequence of video frames. There are four signal conditions: (a) red to blue, (b) blue, (c) blue to blue blinking, and (d) blue blinking.

3.5. *Cross judgment*

The proposed method determines whether a person is crossing a pedestrian crossing or before crossing it, and calculates approximate crossing distance. Considering that a pedestrian crossing exists in the lower half area of an image, the input image is processed only the lower half.

The pedestrian crossing image is processed in the following way. Image noise is removed by preprocessing using a Gaussian filter (**Fig. 1(1)**). Then it is binarized by the discriminant analysis[10] to make it less insensitive to outdoor weather change.

After performing edge extraction with a Sobel filter, a probabilistic Hough transform[11] is applied to the edge image to detect straight lines composing a pedestrian crossing. This process is faster than normal Hough transform.

Finally, based on the obtained straight lines (Fig. 2 (5)), the number of lines is counted and the length of the crossing is calculated (Fig.2(6)).

3.6. *Judgment on crossing possibility*

Whether to cross or not is judged based on the hypothesis on crossing behavior principle. As shown in **Fig.2**, the signal pattern is classified into 6 patterns and pattern identification is performed.

The output is done by voice. (For example, “The traffic signal will change soon.”, “About 3 meters remaining.”)

4. Experimental Results

An experiment was conducted to examine the performance of the proposed method. The used camera is a Sony FDR-X3000. It is mounted on the head of a subject. Five videos were evaluated for the situation analysis including traffic light detection. The traffic signal situation was analyzed including stationary signal and blinking signal discrimination, the judgment on crossing possibility. The pedestrian crossing distance was also estimated.

The results are shown by *F*-values. The average *F*-value of traffic light detection was 0.89. The average *F*-values of red, blue and blinking signals detection were 0.87, 0.89 and 0.94, respectively. Finally, the correct answer rate for crossing possibility judgment was 0.91, and the relative error of the distance estimation on a

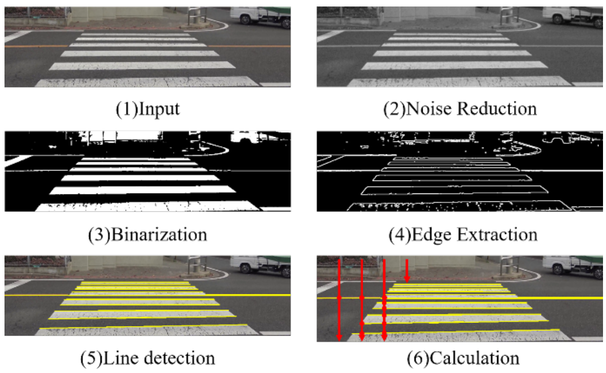


Fig. 1. Flow of the pedestrian crossing distance calculation: (1) Input image, (2) noise reduction, (3) binarization, (4) edge detection, (5) line detection, and (6) calculation of the distance.

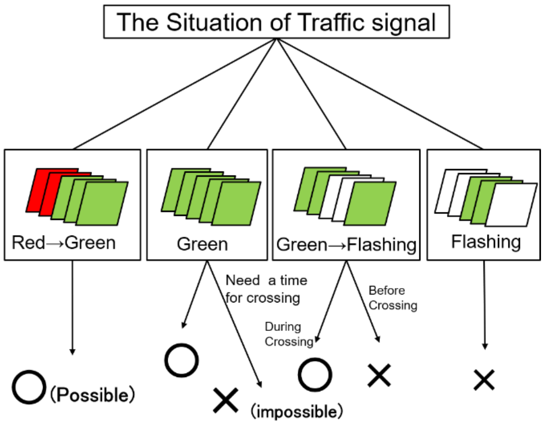


Fig. 2. Pattern identification: Patterns are defined using Pedestrian Crossing Behavior Principle Hypothesis, user status and judgment during crossing.

pedestrian crossing was 25.4%. **Fig. 3** shows a part of the experimental video.

5. Discussion

In the traffic light detection experiment using HOG features and Random Forest, the average *F*-value was 0.89, which was an acceptable detection result. The reason of this good result may be the selection of 4 bins, which could absorb varying inclination of the line segments composing a traffic light.

Regarding the analysis of the traffic light, the signal detection including the blinking signal showed satisfactory results.

6. Conclusion

We proposed a method of detecting and analyzing pedestrian traffic lights from MY VIOSIN images. The effectiveness of the proposed method was verified by experiments. We plan to analyze not only traffic lights but also other kinds of information for a visually impaired person to walk safe in an outdoor environment, and consider developing a navigation system that is more appropriate to the actual environment.

References

1. Bourne, R. R. A., et al., Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis, *The Lancet Glob Health* **5**(9), 2017, e888–e897.

Fig. 3. Part of the experimental video.

2. S. Abe, N. Hashimoto, Survey on walking for the visually impaired, *Bulletin of Hachinohe Institute of Technology* **24**, 2015, pp. 81-92. (In Japanese)
3. M. Imano, T. Hosokawa, et al., Research on autonomous movement of guide robots that guide visually impaired people, *2018 Tohoku chiku wakatekenkyusha happyoukai*, 2018. (In Japanese)
4. Y. Baba , Y. Sakai et al., Real-time recognition of pedestrian signals from moving images, *Proc.77th National Convention of IPSJ* **1**, 2015, pp.543-544. (In Japanese)
5. N. Fairfield, C. Urmson, Traffic light mapping and detection, *2011 IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp.5421-5426.
6. T. Kumano, J. K. Tan, H. Kim, S. Ishikawa, Traffic signs and signals detection employing the My Vision system for a visually impaired person, *International Journal of Research and Surveys, ICIC Express Letters, partB: Applications* **7**(2), 2016, pp.385-391.
7. N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition* **1**, 2005, pp.886-893.
8. L. Breman, Random forests, *Machine Learning* **45**(1), 2001, pp.5-32.
9. K. Hatoyama, S. Sugimori, Analysis on Pedestrian Behavior Considering Spaces before the Crosswalk, *Proceedings of infrastructure planning* **30**, 2004, p.171. (In Japanese)
10. N. Otsu, A threshold selection method from gray-level histograms, *IEEE Transactions on systems, man, and cybernetics* **9**(1), 1979, pp.62-66.
11. N. Kiryati, Y. Eldar, A. M. Bruckstein, A probabilistic Hough transform, *Pattern Recognition* **24**(4), 1991, pp.303-316.

Abnormal Human Action Detection Based on GAN

Tomoya Sano, Seiji Ishikawa, Joo Kooi Tan

Department of Mechanical and Control Engineering, Kyushu Institute of Technology, Sensui-cho, Tobata-ku, Kitakyushu 804-8550, Japan *

E-mail:{sano-tomoya, etheltan}@ss10.cntl.kyutech.ac.jp

http://lab.cntl.kyutech.ac.jp/~etheltan/

Abstract

One of the important roles of a camera surveillance system is to detect abnormal human actions or events. In this study, we propose a method of abnormal human actions/events detection method using Generative Adversarial Nets (GAN). In anomaly action detection, the main problem is that the image data of abnormal human actions is more difficult to obtain than normal human actions. To solve this difficulty, we use only normal human action data in the employed training network and those actions not recognized as normal are judged as abnormal. Experimental results show effectiveness of the proposed method.

Keywords: anomaly detection, camera surveillance system, CNN, GAN, optical flow.

1. Introduction

Recently, surveillance systems using cameras have been widely used according to frequent outbreak of crimes. The number of surveillance cameras installed is on the rise. It is used in various places such as convenience stores, supermarkets, banks, stations, roads, and homes, and not only law enforcement agencies but also private companies and individuals exist as users. However, due to the shortage of manpower, it has not been possible to secure human resources to monitor surveillance camera images. Generally, the frequency of abnormal events is much lower than the frequency of normal events. As a result, most of the time and money is wasted on the work done by the observer. Therefore, there is an urgent need to develop a technology that automatically detects abnormal events from the images of surveillance cameras.

In this paper, we propose a detection method of an anomaly action on a video image using a convolutional neural network. As mentioned above, abnormal events rarely occur compared to normal events. Therefore, it is very difficult to collect data on abnormal events. We solve this problem by using only the data on normal events when training a model. Then an action not recognized as a normal action is judged as an abnormal action.

2. Acquisition of motion information

We estimate optical flows to capture motion features on a video. We use FlowNet [1] for this purpose. FlowNet is a deep convolutional neural network that provides a high-density optical flow for each estimated pixel, given two consecutive video frames. The estimated optical flow is provided in the form of a 2-channel (horizontal and vertical) map. Optical flow expresses the magnitude and direction of movement by a vector.

In the proposed method, we use the 2-channel map that takes the absolute value of the optical flow as motion information. This prevents the loss calculated by the loss function during training (discussed later in Section 3.3) from becoming large due to the inversion of the positive and negative signs and promotes early convergence of the parameters. We call this feature expression Absolute Optical Flow (AOF) for convenience.

3. Framework

3.1. *GAN*

Generative Adversarial Nets (GAN) [2] is a framework for training generative models. It aims at finding a generative data distribution $p_g(\boldsymbol{x})$ that matches the true data distribution $p_{data}(\boldsymbol{x})$ obtained from the training data $\boldsymbol{x}$. Both the generator G, that captures the data distribution, and the discriminator D, that estimates the probability that the input data is obtained from $p_{data}(\boldsymbol{x})$, are trained while competing with each other.

The generator G generates the generated data $G(\boldsymbol{z})$ when random noise $\boldsymbol{z}$ according to the distribution $p_z(\boldsymbol{z})$ is input to it. The input data $\boldsymbol{u}$ to the discriminator D is the training data $\boldsymbol{x}$ or the generated data $G(\boldsymbol{z})$. The output $D(\boldsymbol{u})$ of the discriminator D is the probability that the input data $\boldsymbol{u}$ belongs to the distribution of training data $p_{data}(\boldsymbol{x})$. In GAN, G and D are optimized by the value function $V(G, D)$ of the following equation;

$$\min_G \max_D V(G,D) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(\boldsymbol{x})] + \mathbb{E}_{z\sim p_z(z)}\left[\log\left(1 - D\left(G(\boldsymbol{z})\right)\right)\right] \quad (1)$$

In the above equation, G minimizes the objective function to trick D into generating data that is indistinguishable from true data. On the other hand, D finds the distinguishing boundary between the true data and the generated data by maximizing the objective function and avoids being fooled by G. In this way, G and D have a hostile relationship and are optimized while competing in GAN.

3.2. *Proposed framework*

Our framework learns only normal event data in surveillance camera images. We train two deep convolutional neural networks based on GAN. The first is generator G, which captures the data distribution of normal events. G gives one raw frame obtained from the camera and predicts the estimated motion information (AOF). The second is discriminator D, which estimates the probability that a given sample is AOF obtained by FlowNet. D discriminates whether the given input is AOF obtained by FlowNet or AOF estimated from one frame by G. Figure 1 shows the structure of the proposed framework.

Through this training, we aim at enabling the G to predict the appropriate motion information of the training data (normal event). We use the trained G during test. When the data of a normal action is given to G, the motion information prediction is performed appropriately. On the contrary, when the data including an abnormal action is given to G, the motion information is not estimated properly. As a result, the prediction result generated by G differs from the result obtained by FlowNet. In this way, anomaly detection is executed.

3.3. *Loss function*

The two models G and D minimize the following losses L_G and L_D by learning.

$$L_G = \mathrm{softplus}\left(-D\left(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t, G(\boldsymbol{x}_t)\right)\right) \quad (2)$$

$$L_D = \mathrm{softplus}\left(-D(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t, |F(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t)|)\right) + \mathrm{softplus}\left(D\left(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t, G(\boldsymbol{x}_t)\right)\right) \quad (3)$$

$$\mathrm{softplus}(n) = \log(1 - \exp n) \quad (4)$$

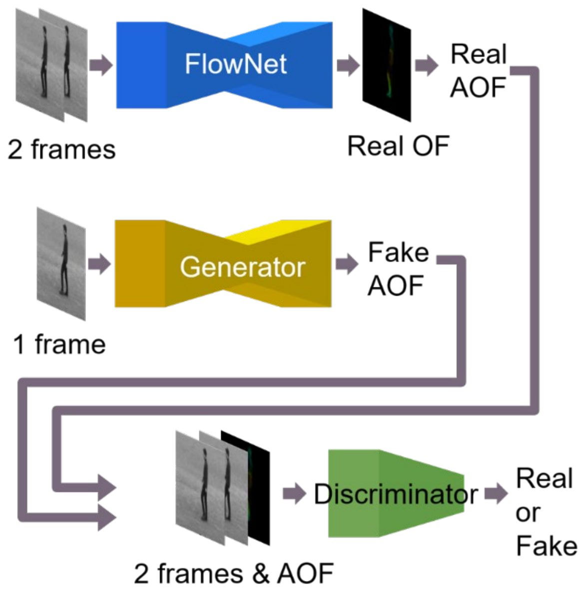


Fig. 1. The structure of the proposed framework.

where $G(\boldsymbol{x}_t)$ is the output of the generator G, $F(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t)$ is the output of FlowNet, $D(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t, G(\boldsymbol{x}_t))$ is the output of discriminator D that estimates the probability that the data generated by G belongs to the distribution of the estimation results by FlowNet.

3.4. *Anomaly detection*

Anomaly detection is executed every frame except the first frame. To detect anomalies, the squared error of the output of generator G and AOF obtained from FlowNet is calculated. The abnormal score at time t is calculated by the following equation.

$$E_t = \sum \left(G(\boldsymbol{x}_t) - F(\boldsymbol{x}_{t-1}, \boldsymbol{x}_t)\right)^2 \qquad (5)$$

When E_t exceeds a threshold value, the video frame at time t is detected as a frame containing an abnormal event.

3.5. *Network architecture*

3.5.1. *Generator G*

Generator G is a Deep Encoder-Decoder that has a skip coupling between the intermediate layer of the contracting part and the expansion part like U-net [3]. The input is one raw video frame (128×128×3). The output is AOF (128×128×2). Details are shown in Fig. 2.

In the 3×3 convolution layer, batch normalization is performed, followed by a rectified linear unit (ReLU). In the contracting part (upper side of Fig. 2), down-sampling is performed by 2×2 max-pooling, which doubles the number of channels in the feature map. After that, in the expansion part, on the contrary, up-sampling is performed by 2×2 deconvolution that halves the number of channels in the feature map.

3.5.2. *Discriminator D*

Discriminator D is a deep convolutional neural network with 10 convolutional layers. The input is a feature map (128×128×8) that connects two consecutive video frames and AOF. The output is the probability (scalar) that the AOF contained in the input is obtained by FlowNet. Details are shown in Fig. 3.

Discriminator D has the same structure as the encoder part of the generator G. Finally, it is output through a fully connected layer. It should be noted that no activation function is applied to the output layer.

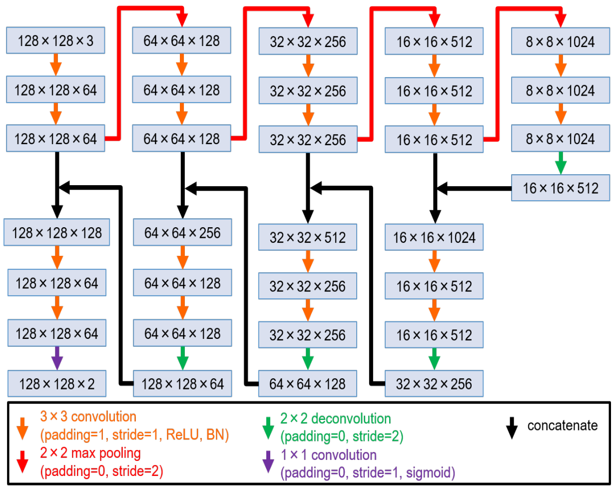


Fig. 2. Details of generator G. The blue block shows the feature map. The orange arrow indicates 3×3 convolution (padding=1, stride=1) with batch normalization, followed by a ReLU. The red arrow indicates 2×2 max-pooling with stride 2. The green arrow indicates 2×2 deconvolution with stride 2. The purple arrow indicates a 1×1 convolution with the sigmoid function applied. The black branched arrow indicates that the two feature maps are connected.

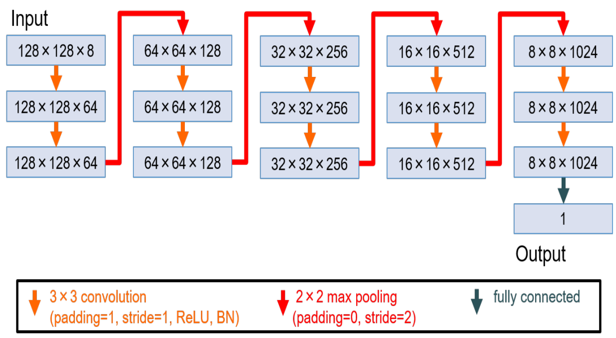


Fig. 3. Details of discriminator D. The dark green arrow indicates a fully connected layer. For other arrows, see the explanation in Fig. 2.

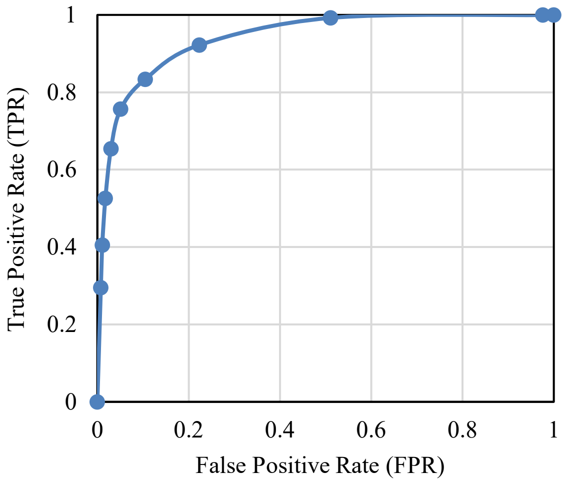


Fig. 4. ROC curve.

4. Experiment

Using the proposed framework, we perform a single-class motion classification experiment. For evaluation, a receiver operating characteristic (ROC) curve and an area under the curve (AUC) are employed. The ROC curve is obtained by calculating the true positive rate (TPR) and the false positive rate (FPR) for each of the normal and abnormal cutoff values for abnormality detection, and plotted on the coordinate plane with TPR on the vertical axis and FPR on the horizontal axis. The area under the curve is calculated to evaluate the effectiveness of the proposed method.

4.1. *Dataset*

The dataset used for the experiment is KTH dataset [4, 5]. KTH dataset is a human motion recognition dataset by 25 actors, including six motion classes: walking, jogging, running, boxing, hand waving, and hand clapping. This dataset contains videos shot under four different conditions (d1, d2, d3, d4), but in this experiment only the data under condition d1 is used. When inputting an image to the network, both ends are cut so that the image becomes square, and it is resized to a 128×128 pixel image. In walking, jogging, and running, there are frames with backgrounds that do not include people. In this experiment, such a frame is excluded in advance.

The walking class in this data set is treated as a normal class, and the other five classes of jogging, running, boxing, hand waving, and hand clapping are tested as abnormal classes. Out of the data of 25 actors, walking videos for 20 actors are used for model training. Six motion videos of walking, jogging, running, boxing, hand waving, and hand clapping by the remaining 5 people are used for the test.

4.2. *Result*

Figure 4 shows the ROC curve of a single-class motion classification experiment using a KTH dataset. The AUC calculated from Fig. 4 was 0.94.

5. Conclusion

We proposed a GAN-based method for detecting abnormal human action using only normal event data for training. Our experiments show that the proposed method is effective in detecting abnormal human actions. Future work includes application to more realistic surveillance video images and improvement of accuracy.

References

1. A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazırbas, V. Golkov, FlowNet: Learning Optical Flow with Convolutional Networks, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 2758–2766.
2. I. J. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D. W. Farley, S. Ozair, A. Courville, Y. Bengio, Generative Adversarial Nets, *Advances in neural information processing systems*, 2014, pp. 2672–2680.
3. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, *International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI*, 2015, pp. 234–241.
4. C. Schuldt, I. Laptev, Barbara Caputo, Recognizing Human Actions: A Local SVM Approach, *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)*, 2004, Vol. 3, pp. 32–36.
5. KTH dataset: https://www.csc.kth.se/cvap/actions/
6. M. Ravanbakhsh, E. Sangineto, M. Nabi, N. Sebe, Training Adversarial Discriminators for Cross-Channel Abnormal Event Detection in Crowds, *Proceedings of 2019 IEEE Winter Conference on Applications of Computer Vision*, 2019, pp. 1896–1904.
7. P. Isola, J. Y. Zhu, T. Zhou, A. A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1125–1134.

Bus Line Number Detection Employing MY VISION

Ye Zhou, Yosiki Hamasaki, Joo Kooi Tan
Department of Mechanical and Control Engineering,
Kyushu Institute of Technology,
Sensui-cho Tobata-ku
Kitakyushu, 804-8550, Japan
E-mail: {zhou-ye, etheltan }@ss10.cntl.kyutech.ac.jp
http://lab.cntl.kyutech.ac.jp/~etheltan/

Abstract

In response to the requests from visually impaired people that they wish better assistance tools in their daily activities, and to the facts that they find difficulty in using public transportation, we propose a bus boarding support system using MY VISION. For the support, we detect a bus approaching a bus stop from MY VISION images employing Haar-like filters, and acquire the moving object area with the pyramidal Lucas-Kanade method. Then, we extract the frontal upper area displaying bus line number and the destination by random forest. Having obtained the area, a method of connecting area marking is used to obtain the bus line number area. The effectiveness of the proposed method is shown by experiments.

Keywords: Bus line number, Moving area detection, Haar-like filters, Random forest, Visually impaired.

1. Introduction

According to the statistics survey from the Ministry of Labor in the past 23 years, there are about 320,000 people with visual impairment in Japan [1]. Meanwhile, about 69% of these people are over 65 years old. It can be seen that with the development of social aging, the number of people with visual impairment is further increasing. For the visually impaired, considering practicality and economic cost, visual impairment usually chooses public transportation. However, public transportation needs a lot of visual information and there are many interference problems, such as a bus. In addition, a bus also contains critical information of the line number and entrance. If the lack of the above information, the visually impaired cannot successfully complete the ride behavior. Therefore, it is necessary to develop a support system that enables visually impaired people to easily use public transportation facilities. In this paper, we propose a visual disability support system that can detect the frontal LED area of a bus and recognize its route number.

Although the computational cost of using learning results to detect specific moving objects in a complex image background is high, the detection results are accurate. This paper proposes a method of detecting a specific moving object. It extracts a moving object area by using the optical flow method, updates the color histogram of the area based on the Camshift [2] tracking method, and corrects the tracking bounding box to complete the entire tracking and recognition process.

The system in this paper contains two structures: learning unit and recognition unit. (1) Learning unit. According to the input image, the Haar-like [3] feature quantity is calculated and extracted, which is based on the brightness difference between the two regions and is robust to noise and brightness changes. After that, the random forest classification method is used to effectively classify the feature quantities. (2) Identification unit. The bus LED area recognition is performed using Harris angle detector to extract the feature points of the input image, and the pyramidal Lucas-Kanade optical flow algorithm[4] to obtain the moving object area [5]. After that, Random Sample Consensus (RANSAC) [6] is used to filter out the interference optical flow information generated by the camera movement, and then use the k-means clustering method to obtain the target area. In the clustering area, the learning-based recognition method is used to complete the task of bus LED area recognition. After a certain target area is obtained, the tracking task is performed.

2. Related work

In the literature, there are many automatic bus number recognition systems. However, most of them use active sensors, such as GPS tracking systems, RFID, etc.

For vision-based methods, Guida[7] proposed the detection and identification of bus line numbers [8]. Their method uses Adaboost-based cascaded classifiers to locate the main geometric entities of the bus facade, and the matching is improved by robust geometric matching. Then the bus façade image is converted to a HSV color space and threshold operations are applied to segment each number.

Pan [9] proposed another vision-based method that detects information from the front of the bus. Their system can be divided into 2 subsystems: bus detection and bus line number detection. In the bus detection, they use HOG and SVM to detect the bus position. In the bus line number detection, they use adjacent character grouping and sophisticated edge detection to find candidate regions. The Haar-like feature is extracted from edge distribution. The acquired features are input into Adaboost to classify each component. Finally, they combine an optical character recognition software with a text-to-speech synthesizer to produce audio.

In short, compared with the aforementioned tracking and recognition method, computational complexity of their methods is high, and the recognition accuracy and the effect of tracking moving objects still need to be improved. Therefore, in the feature area selection, the proposed method uses the optical flow method to effectively filter out the moving object area. At the same time, the Camshift tracking method used in this article has low computational complexity and has certain practicability.

3. Learning unit

3.1. *Haar-like feature*

For the detection of a bus LED area, this article is divided into two types of images for feature learning: the bus LED area and the background area. In order to unify the input image size, it is standardized to 100×30 pixels. In the previous methods based on feature learning, the most commonly used method for vehicle feature learning is HOG feature representation. However, when the feature shapes of two regions are similar, the HOG feature is prone to representation errors, which will affect the subsequent recognition tasks. Therefore, in this article, it is proposed to use Haar-like features for regional characterization. Its main advantage is that it characterizes not only the shape of the object, but also the brightness distribution of the image area effectively. **Figure 1** shows an example diagram of the Haar-like filter and calculation is done by the following formula;

$$H(A,B) = F(A) - F(B) \tag{1}$$

Here, $F(A)$ and $F(B)$ are the average brightness of area A and B.

A B

As shown in **Fig. 2**, the proposed method uses 6 different Haar-like filters. The 6 Haar-like features include two edge features, three linear features and the center-around feature. After obtaining the Haar-like features of the LED area, it is necessary to classify and learn the obtained features. Random forest is used as a classification method in the proposed method.

Fig. 1 Haar feature

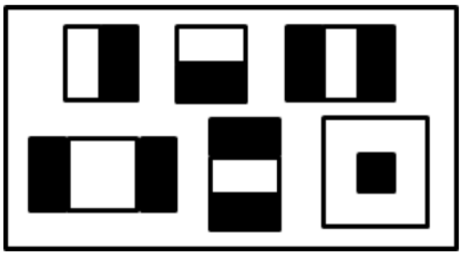


Fig. 2 Six types of filters

3.2. *Random forest*

The system proposed in this paper, it needs to complete the task of identifying multiple types of targets. Therefore, random forest learning [10] meets the needs of this method. It creates a feature tree with low correlation by selecting feature quantities at the non-terminal nodes of each decision tree. It calculates the Haar-like feature value of each input image to create a sample. A subset is randomly generated from the sample, and a random decision tree is created using the branch function of each subset to perform learning. The random forest is determined by inputting the feature vector of the unknown sample into the decision tree. It calculates the category distribution of the terminal node that finally reaches the unknown sample, and uses this result to estimate the category of the unknown sample through the following equation;

$$P(c/S) = \frac{1}{T}\sum_{t=1}^{T} P_t(c/S) \qquad (2)$$

In the above formula, T is the number of trees in the random forest, and $P(c/S)$ is the class distribution.

4. Recognition

The target recognition task is divided into two subtasks: bus LED area recognition and bus line number recognition.

4.1. *Bus LED area recognition*

In order to reduce the amount of data processing, the optical flow method is used to obtain a moving object area [10] to achieve a faster target recognition task. The calculation process is as follows:

1) Use Harris angle detector to acquire the feature points of the target and establish a moving target area. After the feature points are detected, take r frames as the interval. The moving object area is obtained by using the LK pyramid optical flow and k-means clustering method.
2) Since the moving object area obtained by the optical flow method usually includes the noise caused by the camera movement, it is necessary to use the projection transformation model to estimate the camera movement and remove the noise area generated by it. The calculation formula of this model is as follows;

$$\begin{cases} x_2 = \dfrac{h_0x_1+h_1y_1+h_2}{h_6x_1+h_7y_1+1} \\ y_2 = \dfrac{h_3x_1+h_4y_1+h_5}{h_6x_1+h_7y_1+1} \end{cases} \qquad (3)$$

Use RANSAC to determine the location parameters $h_i(i = 0,1 \ldots 7)$of Eq. (3). The RANSAC can effectively divide the internal optical flow area of moving objects and the external optical flow of noise generated by camera movement. Then use the random forest to complete the recognition of the bus LED area.

4.2. *Bus line number recognition*

After obtaining the LED area, it needs to be further processed to obtain the digital area.

According to the relatively fixed position distribution of the LED display area of Japanese buses, a more accurate line number area can be obtained. Then use HSV color processing to obtain the binary image of the digital part, and use the connected area labeling method to segment the digital area.

5. Experimental results

5.1. *Setup*

In this section, we show the results of the method proposed in this paper to learn the characteristics of the bus LED area, as well as the recognition results of the specific moving target area and the line number recognition results obtained from object tracking.

The learning results of the random forests are evaluated by test data. At the same time, the accuracy of the bus LED area detection is also evaluated.

Table 1 shows detailed parameter values for the learning of the random forest. Among them, the dimension of Haar-like features is 5102.

Table 1 Learning parameters of the random forest.

Learning images	Positive image:	1200
	Negative image:	800
Test images	Each class:	500
Size	100 x 30	
Number of subsets	30	
Tree depth	5	
Feature selection times	30	
Threshold selection times	30	

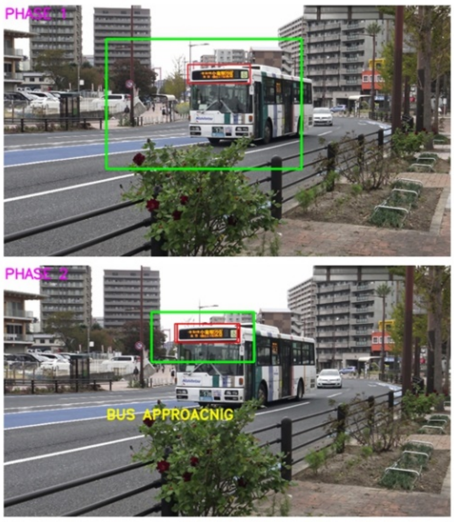
5.2. *Results*

For the learning results of random forest, we use the classification accuracy to evaluate the performance for the plate of bus and the background.

Table 2 shows the result of the classification. It proves that the proposed method can achieve a good performance for the task of classification. **Figure 3** shows the detection performance of the method.

Table 2 Random forest learning results

	Plate	Background
Plate	90.4%	9.6%
Background	4.0%	96.0%



(a)

(b)

Figure 3 Examples of detection result for the bus LED and the line number

The average processing time of each picture is 1.56 [s], and the success rate of detecting the LED panel area in the window of the moving object is 91.3%.

6. Discussion and conclusion

This paper proposes a bus travel assistance system for the visually impaired. The system includes bus LED area recognition and line number recognition tasks. Among them, the overall system includes learning method design, area tracking method design and recognition method design. The final experimental results prove the effectiveness of the proposed method and the practicability of the system.

In the course of the experiment, it was found that when identifying the LED area, there was misrecognition due to lighting conditions, and interference from moving cars. Therefore, in future work, it is necessary to expand the existing data set and adjust the learning parameters to obtain higher recognition accuracy.

7. References

1. Ministry of Health, Labor and Welfare, Social Assistance Bureau, Disability Health and Welfare Department: 2011 Survey on Difficulty in Living (Fact-finding Survey on Children at Home, etc.) Results, Ministry of Health, Labor and Welfare, Social Assistance Bureau, Disability Health and Welfare Department, 2013, pp.13.
2. O.-D. Nouar, G. Ali, C. Raphael Canals. Improved object tracking with camshift algorithm, IEEE International Conference on Acoustics Speech and Signal Processing, 2006, pp.36-42.
3. Gong Haifan, Chen Limin, Li Changhao, Wang Yue. Online tracking and relocation based on a new rotation-invariant Haar-Like statistical descriptor in endoscopic examination, IEEE Access, 99, 2020, pp.69-83.
4. Dowson Nicholas, Bowden Richard. Mutual information for Lucas-Kanade tracking, An inverse compositional formulation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(1), 2008, pp.180-185.
5. Ishitobi K. Tan J K. Kim H. Detection of a specific moving object from head-mounted camera images. IEEE International Symposium on System Integration (SII), 2017, pp. 6pages.
6. M.A. Fischler, R.C. Bolles. Random Sample Consensus: a paradigm for model fitting with applications to image analysis and automated cartography. The ACM Communication, 1981, pp.381-395.
7. C. Guida, D. Comanducci, C. Colombo. Automatic bus line number localization and recognition on mobile phones. A computer vision aid for the visually impaired, Image Analysis and Processing (ICIAP), 2011, pp. 323–332.
8. Pitchakorn Wongta. An automatic bus route number recognition. International Joint Conference on Computer Science and Software Engineering (JCSSE), 2016, pp.19-24.
9. H. Pan, C. Yi, Y. Tian. A primary travelling assistant system of bus detection and recognition for visually impaired people, IEEE International Conference on Multimedia and Expo Workshops (ICMEW), 2013, pp. 1–6.
10. F.Breiman. Random forest. Machine Learning, Vol, 45, No. 1, 2001, pp. 5-32.

A Method for Patterns of Cell-Like Images Based on Distance Transformation

Toru Hiraoka
Department of Information Systems
University of Nagasaki
1-1-1, Manabino, Nagayo-chou, Nishisonogi-gun, Nagasaki-ken, 851-2195, Japan
E-mail: hiraoka@sun.ac.jp

Kohei Maeda
Department of Information Systems
University of Nagasaki
1-1-1, Manabino, Nagayo-chou, Nishisonogi-gun, Nagasaki-ken, 851-2195, Japan
E-mail: bs118031@sun.ac.jp

Abstract

A non-photorealistic rendering method has been proposed to generate cell-like images in which cell-like patterns are represented in photographic images. Cell-like patterns are automatically generated by the change of density of photographic images. However, cell-like patterns are irregularly arranged. In this paper, we propose a method to arrange cell-like patterns along the edges of photographic images. We improves the conventional method by using Euclidean distance from the edges. We show that appealing cell-like images can be generated by our method through experiments using various photographic images.

Keywords: Non-photorealistic rendering, Cell, Euclidean distance, Edge, Automatic generation

1. Introduction

Non-photorealistic rendering is a field of computer graphics that can generate effective illustrations and attractive artistic images. Some researchers have proposed non-photorealistic rendering methods to simulate art expression techniques and to develop unprecedented artistic approaches [1,2,3,4,5,6,7]. One of the unprecedented artistic approaches is a method to generating cell-like images by an iterative calculation using inverse iris filter from photographic images [8]. Cell-like patterns are composed of cell membrane and cell nucleus. Cell-like images are overlaid with cell-like patterns in photographic images. Cell patterns are automatically generated by the change of density of photographic images. However, cell-like patterns are irregularly arranged. Therefore, a method for aligning cell-like patterns in cell-like images has been proposed [9]. The conventional method[9] is implemented by synthesizing sine and cosine waves into photographic images. However, in the conventional method [9], cell-like patterns are mainly arranged in a grid pattern, and therefore the edge preservation of cell-like images is not high. Also, the actual cells are not arranged in the grid pattern.

In this paper, we propose a method to arrange cell-like patterns along the edges of photographic images. We improve the conventional method [8] by using Euclidean distance from the edges. Cell-like images of our method can give an impression different from cell-like images of the conventional method [8,9]. By conducting experiments using various photographic images, it is visually confirmed that cell-like patterns can be generated along the edges. In addition, it is visually confirmed that the size of cell-like patterns can be changed by changing the values of the parameters in our method.

This paper is organized as follows: the second section describes our method to arrange cell-like patterns along the edges, the third section shows experimental results and reveals the effectiveness of our method, and the conclusion of this paper is given in the fourth section.

2. Our method

Our method is implemented in two steps. In the first step, distance-transformed images are created by calculating Euclidean distance from the edges of photographic images. In the second step, cell-like images are generated using distance-transformed images and inverse iris filter. Inverse iris filter is calculated by the procedure that restores the transformed image using iris filter [10] to the original image using inverse filter [11]. The flow chart of our method is shown in Fig. 1.

We explain details of the procedure in Fig. 1 below.

Step 0: Let input pixel values of RGB on coordinates (i, j) in a color photographic image be $f_{R,i,j}$, $f_{G,i,j}$ and $f_{B,i,j}$ ($i = 1, 2, 3, \cdots, I$; $j = 1, 2, 3, \cdots, J$). The pixel values $f_{R,i,j}$, $f_{G,i,j}$ and $f_{B,i,j}$ have value of U gradation from 0 to $U-1$.

Step 1: Gray-scale pixel values $f_{i,j}$ are calculated as follows.

$$f_{i,j} = \frac{f_{R,i,j}+f_{G,i,j}+f_{B,i,j}}{3} \tag{1}$$

Edges are extracted from the gray-scale image using EDISON [12] that is a feature extraction tool that integrates edge detection and image segmentation. Shortest Euclidean distances $d_{1,i,j}$ from the edge pixels are calculated at each pixel.

If the Euclidean distances $d_{1,i,j}$ are within $mD + D/2 \pm 0.5$ ($m = 0, 1, 2, \cdots$), the Euclidean distances $d_{2,i,j}$ are set to 0, where D is a positive constant. Otherwise, the Euclidean distances $d_{2,i,j}$ are set to ∞. When the positions where the Euclidean distances $d_{2,i,j}$ are 0 are expressed on the image, it becomes lines of the interval D along the edges.

Scan from the upper left to the lower right of the image, and in the case that the Euclidean distance $d_{2,i,j}$ are 0 at the target pixel (i, j), the Euclidean distances $d_{2,k,l}$ in the range where the Euclidean distance from the target pixel (i, j) is smaller than D are updated to ∞. Where the pixels (k, l) have the Euclidean distances from the target pixel (i, j) smaller than D, and the Euclidean distance $d_{2,i,j}$ of the target pixel (i, j) is not updated. When the positions where the Euclidean distances $d_{2,i,j}$ are 0 are expressed on the image, it becomes dotted lines with the spacing D.

Scan from the upper left to the lower right of the image, and in the case that the Euclidean distance $d_{2,i,j}$ are ∞ at the target pixel (i, j), the Euclidean distance $d_{2,i,j}$ of the target pixel (i, j) is updated to 0 if there is no pixel where the Euclidean distances $d_{2,k,l}$ are 0 in the range where the Euclidean distance from the target pixel (i, j) is smaller than D. In all pixels (i, j), there are pixels that the Euclidean distances $d_{2,k,l}$ are 0 within the radius D. Shortest Euclidean distances $d_{i,j}$ from the pixels that the Euclidean distances $d_{2,k,l}$ are 0 are calculated at each pixel. An image composed of the shortest Euclidean distances $d_{i,j}$ is called a distance-transformed image.

Step 2: Let output pixel values after processing with iris filter on $d_{i,j}$ be $IF(d_{i,j})$. Iris filter is executed with the $2r+1$ peripheral pixels (k, l) in the window of size r. Angles $\theta_{i,j,k,l}$ between a vector $(i-k, j-l)$ from the peripheral pixels (k, l) to the target pixel (i, j) and a vector $\Big(\big(d_{k+2,l+2}+d_{k+2,l+1}+d_{k+2,l}+d_{k+2,l-1}+d_{k+2,l-2}\big)-\big(d_{k-2,l+2}+d_{k-2,l+1}+d_{k-2,l}+d_{k-2,l-1}+d_{k-2,l-2}\big), \big(d_{k+2,l+2}+d_{k+1,l+2}+d_{k,l+2}+d_{k-1,l+2}+d_{k-2,l+2}\big)-\big(d_{k+2,l-2}+d_{k+1,l-2}+d_{k,l-2}+d_{k-1,l-2}+d_{k-2,l-2}\big)\Big)$ are computed. Let convergence indices of the target pixel (i, j) be

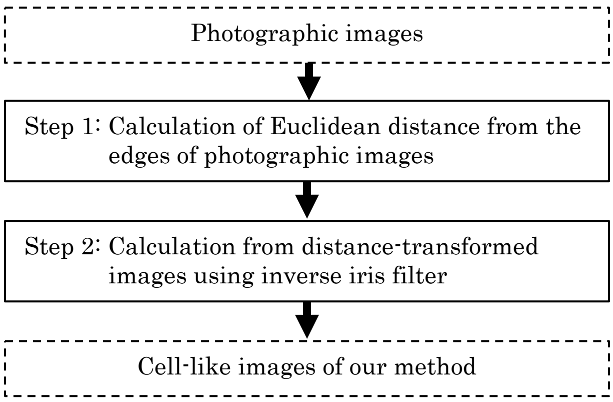


Fig. 1. Flow chart of our method

$c_{i,j}$. The convergence indices $c_{i,j}$ are calculated as follows.

$$c_{i,j} = \frac{1}{(2r+1)^2} \left| \sum_{k=i-r}^{i+r} \sum_{l=j-r}^{j+r} \cos\theta_{i,j,k,l} \right| \quad (2)$$

Let minimum and maximum values of $c_{i,j}$ in all pixels be c_{min} and c_{max}, respectively. The convergence indices $c_{i,j}$ are converted to $C_{i,j}$ as follows.

$$C_{i,j} = 255\left(\frac{c_{i,j} - c_{min}}{c_{max} - c_{min}}\right) \quad (3)$$

The values $IF(d_{i,j})$ and $C_{i,j}$ are the same value. Pixel values $g_{R,i,j}$, $g_{G,i,j}$ and $g_{B,i,j}$ are computed by using inverse iris filter as follows.

$$\begin{aligned} g_{R,i,j} &= a\left(f_{,i,j} - C_{i,j}\right) + f_{R,i,j} \\ g_{G,i,j} &= a\left(f_{,i,j} - C_{i,j}\right) + f_{G,i,j} \\ g_{B,i,j} &= a\left(f_{,i,j} - C_{i,j}\right) + f_{B,i,j} \end{aligned} \quad (4)$$

where a is positive constant. If $g_{R,i,j}$, $g_{G,i,j}$ and $g_{B,i,j}$ are less than 0, then these values must be set to 0, respectively. If $g_{R,i,j}$, $g_{G,i,j}$ and $g_{B,i,j}$ are greater than $U-1$, then these values must be set to $U-1$, respectively. The image composed of the pixel values $g_{R,i,j}$, $g_{G,i,j}$ and $g_{B,i,j}$ is a cell-like image of our method.

3. Experiments

We conducted two experiments. In the first experiment, we visually confirmed the appearance of cell-like images generated from Lenna image shown in Fig. 2 by changing the values of the parameters. In the second experiment, used in the experiments were 512 * 512 pixels and 256 gradation. For reference, the edge extracted image and the distance-transformed image ($D = 15$) of Lenna image are show in Fig. 4 (a) and (b), respectively.

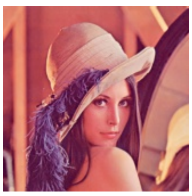


Figure 2. Lenna image.

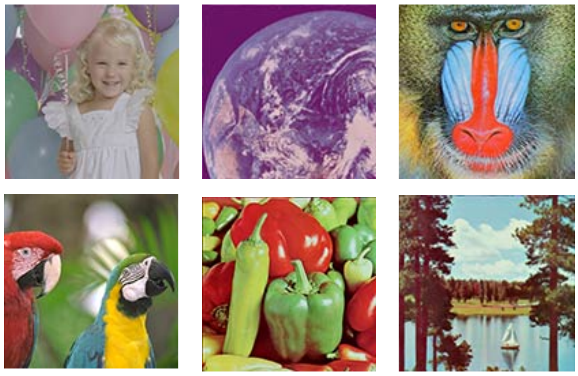
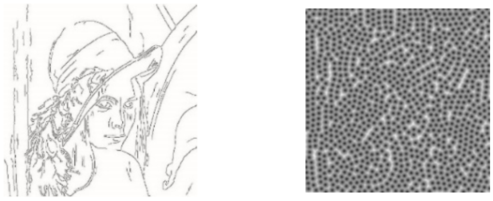


Figure 3. Various photographic images.



(a) Edge-extracted image (b) Distance-transformed image

Figure 4. Edge-extracted image and distance-transformed image.

3.1. Experiments with different parameter values

We visually assessed the change in appearance of cell-like images as the value of the parameter D was changed using Lenna image. The value of the parameter D was set to 10, 15 and 20, and the value of the parameters r and a were set to 3 and 0.4, respectively. Cell-like images in the case are shown in Fig. 5. The larger the value of the parameter D, the larger cell-like patterns were expressed.

We visually assessed the change in appearance of cell-like images as the value of the parameter r was changed using Lenna image. The value of the parameter r was set to 2, 3 and 4, and the value of the parameters D and a were set to 15 and 0.4, respectively. Cell-like images in the case are shown in Fig. 6. The larger the value of the parameter r, the larger the cell nuclei of cell-like patterns were expressed.

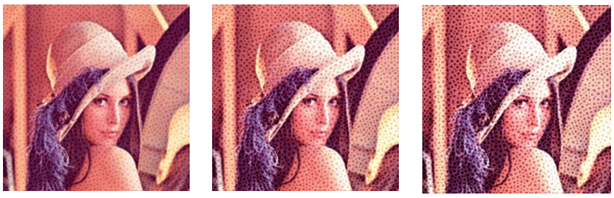
We visually assessed the change in appearance of cell-like images as the value of the parameter a was changed using Lenna image. The value of the parameter

(a) $D = 10$ (b) $D = 15$ (c) $D = 20$

Figure 5. Cell-like images for $D = 10$, 15 and 20.

(a) $r = 2$ (b) $r = 3$ (c) $r = 4$

Figure 6. Cell-like images for $r = 2$, 3 and 4.



(a) $a = 0.2$ (b) $a = 0.4$ (c) $a = 0.6$

Figure 7. Cell-like images for $a = 0.2$, 0.4 and 0.6.

a was set to 0.2, 0.4 and 0.6, and the value of the parameters D and r were set to 15 and 3, respectively. Cell-like images in the case are shown in Fig. 7. The larger the value of the parameter a, the more clearly cell-like patterns were emphasized.

3.2. Experiments with various photographic images

We applied our method to six photographic images shown in Fig. 3. Referring to the results of the experiments in the previous section, the values of the parameters D, r and a were set to 15, 3 and 0.4, respectively. Cell-like images generated from the six photographic images are show in Fig. 8. In all cell-like images, our method could express cell-like patterns along the edges of photographic images. In addition, the cell-like patterns generated by our method were expressed the cell membrane and cell nucleus more clearly than those of the conventional methods [8,9].

Figure 8. Cell-like images.

4. Conclusions

We proposed a method to arrange cell-like patterns along the edges of photographic images. We improved the conventional method [8] by using Euclidean distance from the edges of photographic images. We demonstrated the effectiveness of our method through experiments using various photographic images. The experimental results showed that our method can express cell-like patterns along the edges. In addition, the experimental results showed that the size of cell-like patterns can be changed by changing the values of the parameters in our method. However, cell-like patterns were hard to occur in the fine area and in the black and white areas.

In future work, we will try to make it possible to generate cell-like patterns in the areas where cell-like patterns do not easily occur. And, we will try to apply our method to videos and three-dimensional data.

Acknowledgment. This work was supported by JSPS KAKENHI Grant Number JP19-K12664.

References

1. P. Haeberli, Paint by numbers: abstract image representations, *ACM SIGGRAPH Computer Graphics*, 24(4), 1990, pp.207-214.
2. D. D. Seligmann and S. Feiner, Automated generation of intent-based 3D illustrations, *ACM SIGGRAPH Computer Graphics*, 25(4), 1991, pp.123-132.
3. J. Lansdown and S. Schofield, Expressive rendering: a review of nonphotorealistic techniques, *IEEE Computer Graphics and Applications*, 15(3), 1995, pp.29-37.
4. W. Qian, D. Xu, K. Yue, Z. Guan, Y. Pu and Y. Shi, Gourd pyrography art simulating based on non-photorealistic rendering, *Multimedia Tools and Applications*, 76(13), 2017, pp.14559-14579.
5. D. Martin, G. Arroyo, A. Rodriguez and T. Isenberg, A survey of digital stippling, *Computers & Graphics*, 67, 2017, pp.24-44.

6. W. Qian, D. Xu, J. Cao, Z. Guan and Y. Pu, Aesthetic art simulation for embroidery style, *Multimedia Tools and Applications*, 78(1), 2019, pp.995-1016.
7. T. Hiraoka, Generation of edge-enhancing labyrinth images using inverse filter and two improved Laplacian filters, *ICIC Express Letters*, 14(2), 2020, pp.121-127.
8. T. Hiraoka, M. Hirota, K. Inoue and K. Urahama, Generating cell-like color images by inverse iris filter, *ICIC Express Letters*, 11(2), 2017, pp.399-404.
9. T. Hiraoka, A method for emphasizing and aligning patterns in cell-like images, *ICIC Express Letters*, 13(11), 2019, pp.1031-1037.
10. H. Kobatake and S. Hashimoto, Convergence index filter for vector fields, *IEEE Transactions on Image Processing*, 8(8), 1999, pp.1029-1038.
11. Z. Yu and K. Urahama, Iterative method for inverse nonlinear image processing, *IEICE Transactions on Fundamentals*, E97-A(2), 2014, pp.719-721.
12. The Robust Image Understanding Laboratory at Rutgers University, *Edge detection and image segmentation (EDISON) system*, http://rci.rutgers.edu/˜meer/RIUL/research/code/EDISON/doc/help.html, 2002.

Generating Striped Animations by Inverse Line Convergence Index Filter

Toru Hiraoka
Department of Information Systems, University of Nagasaki
1-1-1, Manabino Nagayo-chou, Nishisonogi-gun, Nagasaki-ken, 815—2195, Japan
E-mail: hiraoka@sun.ac.jp

Ryosuke Takaki
Division of Computer Science, Graduate School of Regional Design and Creation, University of Nagasaki
1-1-1, Manabino, Nagayo-chou, Nishisonogi-gun, Nagasaki-ken, 815-2195, Japan
E-mail: mc220002@sun.ac.jp

Abstract

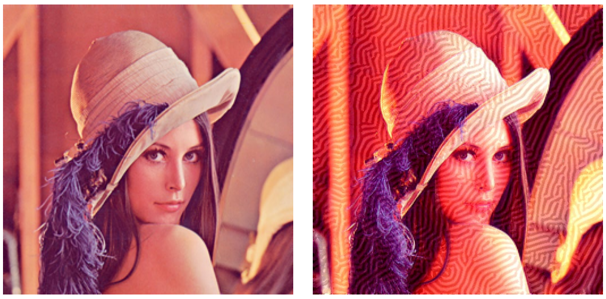
A non-photorealistic rendering method has been proposed for generating a striped image which is overlaid striped patterns in a photograph. The conventional method generates the striped image by an iterative process using an inverse line convergence index filter. When a striped animation is generated by converting each frame of a video by the conventional method, flickering occurs in the generated striped animation. In this paper, we propose a method for generating a striped animation that has characteristic with less flicker from the video. The effectiveness of the proposed method is investigated experimentally. As a result of the experiments, the proposed striped animation had less flicker than the conventional striped animation.

Keywords: Non-photorealistic rendering, animation, striped pattern, inverse line convergence index filter

1. Introduction

Various non-photorealistic rendering methods for a photograph have been proposed[1,2,3,4,5,6]. These methods are used in a wide range of applications, for examples, applications embedded in a personal computer and a portable terminal. Applying these methods to a video is to improve the visual appearance. A Study on non-photorealistic rendering of the video has also been conducted[7], but the conventional method was been known that flickering occurs in the non-photorealistic rendering animation.

In this paper, we focus on the non-photorealistic rendering method for generating a striped image[6] from the photograph. Also we apply the conventional method to the video. The striped image is overlaid the striped patterns in the photograph as shown in Fig. 1. The conventional method generates the striped image by an iterative process using an inverse line convergence index filter, and is characterized by the ability to automatically generate the striped patterns in accordance with the shading and the edge of the photograph. When a striped animation is generated by converting each frames of the video by the conventional method, flickering occurs in the generated striped animation. Therefore, we suppress flicker by using the forward and backward frames of the



(a) Lena image (b) Striped image
Fig. 1. Lena image and striped image

video. The effectiveness of the proposed method is investigated experimentally. By visually and quantitatively comparing the proposed striped animation to the conventional striped animation, we show that the proposed method can suppress flicker better than the conventional method.

2. Method

Let input pixel values of RGB on coordinates (i,j) in o-th frame of the video be $f_{R,i,j,o}$, $f_{G,i,j,o}$ and $f_{B,i,j,o}$ $(i=1,2,\dots,I; j=1,2,\dots,J; o=1,2,\dots,O)$, and let output pixel values after processing with a line convergence index filter[8] be $LF(f_{R,i,j,o})$, $LF(f_{G,i,j,o})$ and $LF(f_{B,i,j,o})$.

The line convergence index filter is executed in the following manner. Consider the straight line $l_{m,o}(m=1,2,\dots,M)$ inclined $\theta_{m,o}(=0,\pi/M,2\pi/M,\dots,(M-1)\pi/M)$ degrees from the x-direction of the target pixel (i,j). Let the length of the straight line $l_{m,o}$ in each direction around the target pixel (i,j) be W_1 pixels. Consider the inclined rectangle which has the center at (i,j) and the length of sides are W_1, W_2. Let N be the number of pixels (k,l) inside the rectangle (Conceptual diagram for W_1, W_2 and N are shown in Fig. 2). Compute a cosine of the angle between a vector perpendicular to a straight line $l_{m,o}$ from a neighboring pixel (k,l) and a vector $((s_{k+2,l+2,o}+s_{k+2,l+1,o}+s_{k+2,l,o}+s_{k+2,l-1,o}+s_{k+2,l-2,o})-(s_{k-2,l+2,o}+s_{k-2,l+1,o}+s_{k-2,l,o}+s_{k-2,l-1,o}+s_{k-2,l-2,o}),(s_{k+2,l+2,o}+s_{k+1,l+2,o}+s_{k,l+2,o}+s_{k-1,l+2,o}+s_{k-2,l+2,o})-(s_{k+2,l-2,o}+s_{k+1,l-2,o}+s_{k,l-2,o}+s_{k-1,l-2,o}+s_{k-2,l-2,o}))$ calculated from the density variation. Let the cosine of the angle be $c_{m,n,o}(n=1,2,\dots,N)$. The term $s_{i,j,o}$ is calculated as the following equation.

$$d_{i,j,o}=\frac{f_{R,i,j,o}+f_{G,i,j,o}+f_{B,i,j,o}}{3}. \qquad (1)$$

$$s_{i,j,o}=\frac{\sum_{p=-P}^{P}\frac{1}{1+|p|}d_{i,j,o+p}}{\sum_{p=-P}^{P}\frac{1}{1+|p|}}. \qquad (2)$$

Where P is a positive constant. Next, compute $C_{m,o}$ that is the average of the absolute value of $c_{m,n,o}$ of N neighboring pixels (k,l) in each straight line $l_{m,o}$ as the following equation.

$$C_{m,o}=\frac{1}{N}\sum_{n=1}^{N}\left|c_{m,n,o}\right|. \qquad (3)$$

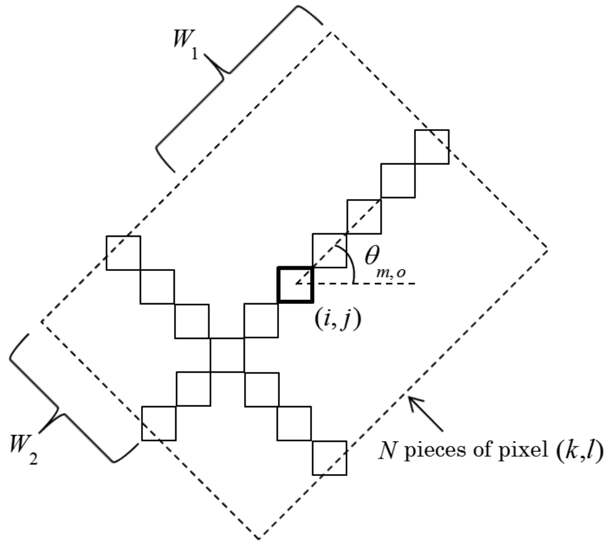


Fig. 2. Conceptual diagram for W_1, W_2 and N

With the maximum value of $C_{m,o}$ in all straight lines $l_{m,o}$ in each pixel (i,j), denote as $g_{i,j,o}$. With the minimum and maximum values of $g_{i,j}$ in all pixels, denoted as $g_{min,o}$ and $g_{max,o}$, respectively. Transform $g_{i,j,o}$ to $h_{i,j,o}$ and thus to the pixel values had from 0 to 255 by thefollowing equation.

$$h_{i,j,o}=255\frac{g_{i,j,o}-g_{min,o}}{g_{max,o}-g_{min,o}}. \qquad (4)$$

$LF(f_{R,i,j,o})$, $LF(f_{R,i,j,o})$ and $LF(f_{R,i,j,o})$ are the same value as $h_{i,j,o}$.

Compute the pixel value $f_{R,i,j,o}^{(t)}$, $f_{G,i,j,o}^{(t)}$ and $f_{B,i,j,o}^{(t)}$ by using the inverse line convergence index filter as

$$f_{R,i,j,o}^{(t)}=a(f_{R,i,j,o}^{(t-1)}-LF(f_{R,i,j,o}^{(t-1)}))+f_{R,i,j,o}. \qquad (5)$$

$$f_{G,i,j,o}^{(t)}=a(f_{G,i,j,o}^{(t-1)}-LF(f_{G,i,j,o}^{(t-1)}))+f_{G,i,j,o}. \qquad (6)$$

$$f_{B,i,j,o}^{(t)}=a(f_{B,i,j,o}^{(t-1)}-LF(f_{B,i,j,o}^{(t-1)}))+f_{B,i,j,o}. \qquad (7)$$

Where a is a positive constant and t is the number of iterations. Let the initial value $f_{R,i,j,o}^{(0)}$, $f_{G,i,j,o}^{(0)}$ and $f_{B,i,j,o}^{(0)}$ be $f_{R,i,j,o}$, $f_{G,i,j,o}$ and $f_{B,i,j,o}$, respectively. $f_{R,i,j,o}^{(t)}$, $f_{G,i,j,o}^{(t)}$ and $f_{B,i,j,o}^{(t)}$ set to 0 if their values are less than 0, and set to 255 if their values are greater than 255.

Finally, the N striped images are obtained after processing of the inverse line convergence index filter of T generated from these striped images.

3. Experiments

We applied the proposed method to the Yuzenzome video[9] which consists of 703 frames, 30 frames / second, $352 * 240$ pixels and 256 tone. The 100, 101, 387 and 388th frames of the Yuzenzome video are shown in Fig. 3. The 387 and 388 th ftames are the scene change frames. In the following experiments, referring to the literature 6), we set the values of the parameters M, W_1, W_2, a and T to 8, 4, 4, 0.4 and 30, respectively.

First, we visually compared the proposed striped animation to the conventional striped animation. The 100, 101, 387 and 388th frames of the conventional striped animation and the proposed striped animation are

(a) 100th

(b) 101th

(c) 387th

(d) 388th

Fig. 3. The frames of the Yuzenzome video

(a) 100th

(b) 101th

(c) 387th

(d) 388th

Fig. 4. The frames of the conventional striped animation

(a) 100th

(b) 101th

(c) 387th

(d) 388th

Fig. 5. The frames of the proposed animation

shown in Fig. 4 and Fig. 5, respectively. We set the value of the parameter P to 3. Observing these animations, the proposed striped animation were suppressed flicker more than the conventional striped animation. Observing Fig. 4 and Fig. 5, the proposed striped animation had less changes in the striped patterns than the conventional striped animation.

Next, we quantitatively compared the proposed striped animation to the conventional striped animation. We calculated the average of the absolute difference value of pixel values between front and rear frames of these animations. As the average become smaller, flickering is less. The average of the conventional striped animation is shown in Table 1. On the other hand, the average of the proposed striped animation is shown in Table 2, when we changed the values of parameter P from 1 to 10. Observing Table 2, as the values of parameter P became bigger, the average became small. Observing Table 1 and Table 2, the average of the proposed striped animation were smaller than the average of the conventional striped animation. Thus, the proposed method can suppress flicker better than the conventional method.

Lastly, we compared the calculation time of the proposed method to that of the conventional method. The calculation time of the conventional method is shown in Table 3. On the other hand, the calculation time of the proposed method is shown in Table 4, when we changed the values of parameter P from 1 to 10. The calculation time in Table 3 and Table 4 is one iteration calculation

Table 1. The average of the conventional striped animation

Average
23.856

Table 2. The average of the proposed striped animation when the value of the parameter P is changed

P	Average
1	20.403
2	20.264
3	20.190
4	20.151
5	20.118
6	20.100
7	20.087
8	20.075
9	20.067
10	20.062

Table 3. The calculation time of the conventional method [second]

Average
4.377

Table 4. The calculation time of the proposed method when the value of the parameter P is changed [second]

P	Average
1	4.381
2	4.390
3	4.394
4	4.399
5	4.401
6	4.405
7	4.410
8	4.413
9	4.420
10	4.426

[second]. Observing Table 4, the calculation time became big as the values of parameter P became bigger, but there are no big differences in these calculation times. Observing Table 3 and Table 4, there was no big difference in the calculation time between the proposed method and the conventional method. Thus, the proposed method is considered to be an effective method compared with the conventional method on the calculation time. However, considering real-time use on portable terminals, it is necessary to speed up the proposed method. Thecomputing environment of the experiment is Windows 7 Professional for OS, 3.40 GHz for CPU, 8.00 GB for RAM and C language.

4. Conclusion

We proposed a method for generating a striped animation that has characteristic with less flicker from the video. The effectiveness of the proposed method was investigated experimentally. As a result of the experiments, the proposed striped animation had less flicker than the conventional striped animation. The future tasks are to speed up the proposed method and to apply the proposed method to other videos.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP19K12664.

References

1. D. Nakagawa, K. Yamaguchi, K. Muraoka, and N. Chiba, Generating Pen and Ink Illustrations from Gray-scale Images : Stippling and Line Drawing Style, *Journal of IIEEJ*. **30**(4), 2001, pp. 350-361.
2. J. Sugita and T. Takahashi, A Generation Method for Colored-Paper-Mosaic Images Based on Hierarchial Poisson Disk Sampling, *Journal of IIEEJ*. **36**(4), 2007, pp. 407-416.
3. Y. Takada and K. Urahama, Dendritic Line Images Generated from Distance Transform and its Application to NPR, *Journal of IIEEJ*. **38**(5), 2009, pp. 791-794.
4. T. Saito, Non-photorealistic Rendering, *Journal of IIEEJ*. **39**(6), 2010, pp. 837-839.
5. R. Ando and R. Tsuruno, An Interactive Brush Stroke Synthesis Using Exemplar Images, *Journal of IIEEJ*. **40**(4), 2011, pp. 578-586.
6. T. Hiraoka and K. Urahama, Generating Striped Color Images by Inverse Line Convergence Index Filter, *IIEEJ Transactions on Image Electronics and Visual Computing*. **2**(2), 2014, pp. 190-194.
7. B. J. Meier, Painterly Rendering for Animation, *Proceedings of ACM SIGGRAPH 1996*, 1996, pp. 477-484.
8. H. Kobatake and S. Hashimoto, Convergence Index Filter for Vector Fields, *IEEE Transactions Image Processing*. **8**, 1999.
9. The Video-Database-Working Group of PRMU, Video Database for Evaluating Video Processing, http://www.nii.ac.jp/dsc/idr/video/VDB/documentary.html, Yuzenzo-me, 2002.

Estimating Home Location of Foreigners in Japan Using Photograph Location

Masaharu Hirota
Department of Information Science, Okayama University of Science
1-1 Ridaicho, Kita-ku, Okayama-shi, 700-0005, Japan

Tetsuya Oda
Department of Information and Computer Engineering, Okayama University of Science
1-1 Ridaicho, Kita-ku, Okayama-shi, 700-0005, Japan
E-mail: hirota@mis.ous.ac.jp, oda@ice.ous.ac.jp

Abstract

The attributes of travelers such as home location and age could be useful for many applications such as information recommendation and targeted advertisement. However, this information is not accessible in most Web services because users do not reveal them. We propose a method to estimate a foreign tourist's home location based on each region's tendency, in which tourists from a region tend to visit certain places when traveling abroad. The feature for the estimation uses the frequency of photograph location in a geohash. In this paper, we use foreigners in Japan as a case study. We evaluate the performance of our proposed method by using photographs obtained from their user accounts on Flickr.

Keywords: home location estimation, user trajectory, machine learning, Flickr.

1. Introduction

With the rapid growth of social media websites such as Flickr or Twitter, it has become common for tourists to upload geo-tagged content from their trips. The contents are recommended to the other users based on user information and its friendship network. Therefore, the content strongly influences tourists' decisions, such as their travel destinations and routes, because tourists may refer to this information when they plan their trips.

The user attribute of users on social media websites (e.g., home location, age, and gender) is important for many applications. The user's home location is especially used for targeting advertisement, information recommendation, and user behavior analysis. However, many users on social media websites do not provide information on the home location. It is also possible that the user provides false information.

To estimate the user's home location, we can use the frequency of geo-tags annotated to the user's content. For example, a simple way is to regard the area with the most geo-tags indicating a particular area as the user's home location. However, this method's limitation is that it cannot be applied when the analyst only has data for a specific region. For example, it is impossible to estimate the country of residence outside of Japan if the analyst only has data posted in Japan. Therefore, a method to estimate the home location from a dataset for a particular region only is necessary.

This paper proposes a method to estimate the user's home location based on geo-tags visiting from abroad using contents obtained from a specific area. Our hypothesis is that travelers in different home locations visit different areas. Some places are widely popular tourist spots, like the Eiffel Tower in France and the

Leaning Tower of Pisa in Italy. However, when travelers go abroad, tourists from a particular country may be interested in specific places, not necessarily conventional tourist spots. These are places associated with a specific region, such as a famous movie in their origin region, a restaurant whose owners are from their region, or a place featured on a local TV show. If we can extract the tendency of the user's visitation by home location, we assume that it is possible to estimate the user's home location from geo-tags.

In this paper, we estimate the region of residence of users who come to Japan from abroad. We used geo-tagged photographs obtained from Flickr. We evaluate the performance of our proposed method by estimating the home location of users in Flickr. Also, we discussed the effect of area size of geohash on the estimation performance.

This paper's remainder is organized as follows: Section 2 describes the work related to this topic. Section 3 presents our proposed method for estimating the user's home location. Section 4 presents the experiment conducted to evaluate the performance of our proposed method. Finally, Section 5 concludes our work and discuss results and areas of future work.

2. Related Work

A wide range of home location methods estimating from social media websites content has been proposed in recent years.

Hironaka et al. used data from Twitter to analyze users' home location based on friend relationships[1]. Hu et al. proposed a method to estimate home location from sparse and noisy Twitter data within 100 by 100-meter squares at high accuracy using users' trajectories in their home country[2]. Jurgens et al. evaluated several methods for geo-location prediction using data from Twitter[3].

The above methods focus on estimating users' home location of social media websites by using content posted in their home country. These researches estimate the home location of users within the posted range of the data. Therefore, it is difficult to estimate the home location of users visiting from other regions. Our previous research[4] proposed a new method for home location inference by analyzing foreign tourists' tendencies in a different region from their home country. In this research, we proposed a method for estimating the user's home location and evaluating the method's performance.

3. Proposed Method

Our proposed method consists of two steps as follows.

(i). Extracting features based on occurrence of users' visits from user's contents.

(ii). Estimating the home location of a tourist from their trajectory using a machine learning algorithm.

3.1. *Feature extraction*

We describe the generation of features for machine learning.

In this study, we use geohash generated from a photographic location as a user's feature. A geohash is a string representation of a geographical area, created by encoding latitude and longitude. Geohash converts a point expressed in latitude and longitude into a geographical area, which reduces the accuracy of representing a geographical point. In addition, depending on the string's length of geohash, it is possible to control the degree of accuracy loss. For example, the latitude and longitude of Okayama Castle (latitude, longitude) = (34.66568196, 133.93593695) is converted to 'wypjpy9' in 7-characters geohash.

To create features, we count the number of photographs taken in each geohash for each user. Then, we delete geohashs with fewer than ten users who took the photographs. Finally, we apply L2 normalization to the features.

3.2. *Estimating Home Location*

In this research, we use Random Forest[5] (RF) to estimate the user's home location based on feature vectors obtained from Section 3.1. This algorithm is a multi-class classifier algorithm in supervised learning. In this research, we regard the predicted class from the feature vector obtained from a user's photostream as the user's estimated home location.

4. Experiment

In this paper, we evaluate the performance of our proposed method through an evaluation experiment based on classification. We describe the experiment conditions of the dataset and the evaluation criteria.

As described in Section 3.1, the geohash allows controlling the degree to which latitude and longitude are quantized depending on the string's length to be converted. Therefore, our experiment evaluates the effect of geohash string length on classification performance.

4.1. *Dataset*

In this section, we describe the dataset used for evaluating our proposed method. We used Japan as a particular country for this experiment. We estimate the home location of users who visit from other regions and using the photographs taken in Japan.

The evaluation experiments used photographs obtained from Flickr. The photographs include the metadata: latitude, longitude, and home location. The home location is the text data where the user in Flickr describes their home location.

Because Flickr's home location field is an open text field, the data in it varies widely. Therefore, we mapped all the abbreviations and significant states/cities to the same country name (i.e., “usa”, “u.s.a” and “Dallas, Texas” were mapped to “usa”).

We excluded photographs where either the latitude or the longitude does not have a value with a precision less than or equal to the third decimal place. We also exclude users who describe more than one home location in their photographs and users who have less than ten photographs.

In this research, we used three home location: Taiwan, America, and United Kingdom. This reason is that the number of users in which we obtained photographs from Flickr is Taiwan, USA, and UK. Those regions are the top three regions in the number of users in our photographs obtained from Flickr. Therefore, our method tackles three-class classification.

Consequently, the number of photographs used in this experiment was 461,454 and the number of users is 1,683. Also, we randomly split into a ratio of 9:1 for train and test data.

4.2. *Evaluation criteria*

We used the following widely used performance measures for classification: Accuracy, Recall, Precision, and F-score. In this research, our classification is multi-class classification. Therefore, we used the macro average of them.

Table 1. Evaluation result.

Geohash	Precision	Recall	F1-score	Accuracy
3	0.499	0.494	0.463	0.552
4	0.543	0.461	0.426	0.554
5	**0.716**	**0.529**	0.455	0.581
6	0.537	0.527	0.479	0.625
7	0.597	0.522	0.491	**0.680**
8	0.570	0.513	**0.515**	0.649

4.3. *Experimental conditions*

This section describes the procedure used for the classifiers to estimate home locations.

This experiment used entropy and Gini impurity for a split of nodes in RF. In addition, the hyperparameters such as max nodes and the number of trees were searched using Optuna[6] with five cross-validations, which is a software for automatically optimizing hyperparameters. We used the parameters with the highest accuracy measured through this experiment. In addition, we used the Python software scikit-learn[7] for the implementation of RF.

4.4. *Evaluation results*

This section describes and discusses the evaluation results of classifying home locations and the effects on string length of geohash.

Table 1 shows the evaluation results of Accuracy and macro average of Recall, Precision, and F-score. The first column represents the string length of geohash. The best value of the length in Precision and Recall is 5. In this case, the UK's Precision value was less than 0.5 when other parameters were used, but it was 1.0 when this parameter was used. The Recall values for the other two regions were also high. Therefore, the result of this parameter may overfit to the UK. Also, in Table 1, the larger the value of the length of geohash, the larger the value of Accuracy. The proposed method may have the best classification performance when the length of the geohash string is 7.

Table 1 shows that the proposed method works well because some performance is obtained for each evaluation criterion.

5. Conclusion

In this paper, we proposed a method to estimate the home location using the user’s trajectory. Our approach uses the tendency of users from the same location to travel to the same places. Also, our approach uses the geohash obtained from the photograph location. Experimental results showed that our classifier could estimate the candidates of home location.

Our future work will include a more detailed experiment. This paper is not enough in the analysis of the home location predicted by our proposed method. Also, the experiment in this paper uses a few photographs. As a result, the experiment result has the limitation of the validity of our proposed method's effectiveness. Therefore, another further study may make the data which includes more number of a dataset.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP19K20418 and Grant for Promotion of OUS Research Project (OUS-RP-20-3).

References

1. Shiori Hironaka, Mitsuo Yoshida, Kyoji Umemura, User's Centrality Analysis for Home Location Estimation. Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence, 14–17, 2019.
2. Tianran Hu, Jiebo Luo, Henry Kautz and Adam Sadilek, Home Location Inference from Sparse and Noisy Data: Models and Applications. Proceedings of IEEE 15th International Conference on Data Mining Workshops, Atlantic City, 14-17, 2015.
3. David Jurgens, Tyler Finethy, James McCorriston, Yi Tian Xu, Derek Ruths, Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice. Proceedings of the Ninth International AAAI Conference on Web and Social Media, 26-29, 2015.
4. K. Lugasi, M.Hirota, Inferring Home Location of Foreign Tourists Based on Travel Routes Extracted from Social Media Sites, The 2020 International Conference on Artificial ALife and Robotics, 2020.
5. L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5–32, 2001.
6. T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, Optuna: A next-generation hyperparameter optimization framework, 2019.
7. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

A Proposal of Online Map-matching Based Trajectory Compression Algorithm Using Road Networks

Shota Iiyama, Masaharu Hirota
Department of Information Science, Okayama University of Science
1-1 Ridaicho, Kita-ku, Okayama-shi, 700-0005, Japan

Tetsuya Oda
Department of Information and Computer Engineering, Okayama University of Science
1-1 Ridaicho, Kita-ku, Okayama-shi, 700-0005, Japan
E-mail: i17i005is@ous.jp, oda@ice.ous.ac.jp, hirota@mis.ous.ac.jp

Abstract

As the data size of GPS logs increases, the amount of data transferred from mobile devices to the server, and the computational cost of analysis of GPS logs increase. One of the solutions to these problems is to compress the GPS logs. However, it is difficult to compress a sparse GPS log while preserving the feature points in the GPS log, such as the user's movement speed, the shape of the GPS log trajectory, and the direction of movement of the GPS log. This study proposes an online compression algorithm for GPS logs that maintains the compression rate while preserving GPS logs' feature points. Our proposed method compresses GPS logs by using information from a road network to identify roads traveled by the user. We evaluate the performance of this method using the GPS data of the bus.

Keywords: GPS, compression algorithm, road network, characteristics of user movement.

1. Introduction

The number of GPS logs that record users' activities increases. Therefore, the data transferred from mobile devices to the server is a critical problem. Another problem is that the processing speed of analysis of GPS logs decreases when analyzing the GPS logs.

One solution to these issues is to compress the GPS logs. Compressing the GPS logs reduces the amount of them transmitted from the mobile device to the server. Also, adequately compressed GPS logs reduce the amount of data processing for analysis.

However, if the GPS log compression lost feature points, the analysis's performance may be worse than the uncompressed. Therefore, it is important to preserve as much of the features as possible. This paper proposes an algorithm for compressing GPS logs moving on the road network while preserving the trajectory's feature points.

2. Related work

The method for GPS log compression without using the road network are Douglas-Peucker Algorithm (DP algorithm)[1] and top-down time-ratio algorithm (TD-TR algorithm)[2]. These methods focus on the shape of the GPS log and compress the GPS log.

Also, the methods for GPS log compression using the road network include Schmid et al.'s method[3]. This method is a compression method for GPS logs of users moving on roads in urban areas. In this method, map-matching is applied to the GPS logs to identify the roads moved by the user. Next, the method uses compression

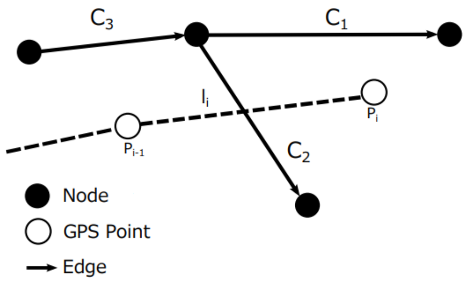


Figure 1 Map-matching.

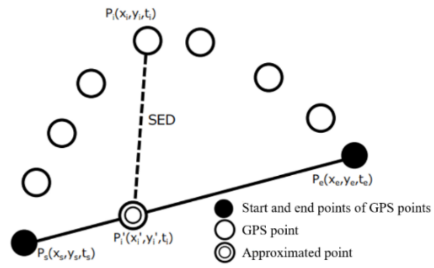


Figure 2 Synchronized Euclidean Distance.

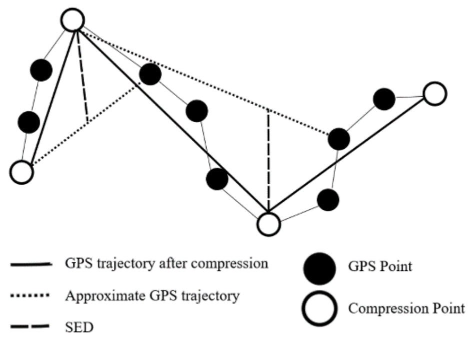


Figure 3 Opening Window Compression Algorithm.

points the GPS logs obtained at intersections and bus stops on the roads moved by the user.

Our method compresses GPS logs not only by focusing on the shape, but the features of movement.

3. Method

Our compression method employs three methods to preserve the feature points of GPS logs: Opening Window Compression Algorithm (OW)[2], CriticalPoint Algorithm (CP)[4], and StayPoint (SP)[5]. OW is a shape-oriented compression algorithm. CP is a compression algorithm that focuses on changes in velocity and direction of motion. SP is the point at which the user was staying. Our method makes a compression result by combining each of the OW, CP, and SP extracted from the GPS logs.

3.1. *Map-matching*

Our method uses the offline map-matching method proposed by Brakatsoulas et al.[6]. In our method, we extend this method to apply to online data.

Figure 1 shows the example of a network. C_1, C_2, and C_3 in Figure 1 represent the vectors from the start to the end node of the road. A node represents an intersection.

This method finds the candidate roads C_1 and C_2 and selects one of them using equation (1) for the point P_i.

$$S_{p_i} = \mu_d - a \cdot D(P_i, C_i)^{n_d} + \mu_\alpha - \cos(\alpha_{i,j})^{n_\alpha} \quad (1)$$

μ_d, μ_α, n_d, and n_α are constants. C_j is a candidate road in P_i. $D(P_i,\ C_j)$ is the geographical distance between P_i and C_j. $\alpha_{i,j}$ is the difference between the angles of l_i and C_j. l_i is the vector from P_{i-1} to P_i. $\cos(\alpha_{i,j})$ is the cosine value of $\alpha_{i,j}$. Among the candidate roads, the road with the highest value of S_{p_i} is the destination of P_i.

3.2. *Opening Window Compression Algorithms*

First, we describe the SED that we uses in OW. The SED is the value of the distance between point P'_i of the GPS log after compression and point P of the GPS log before compression, considering the time information.

We use the method of Meratnia et al.[2] to obtain the SED. Figure 2 shows an example of SED. The P_i shows one of the GPS points in the GPS log. P'_i is the point at which P_i is approximated on the line P_s to P_e. x'_i and y'_i are found by the following equations.

$$\Delta e = t_e - t_s, \qquad \Delta i = t_i - t_s \quad (2)$$

$$x'_i = x_e + \frac{\Delta i}{\Delta e}(x_e - x_s) \quad (3)$$

$$y'_i = y_e + \frac{\Delta i}{\Delta e}(y_e - y_s) \quad (4)$$

OW defines the $anchor$ and $float$. The $anchor$ p_s is the first point in the GPS log. The $float$ p_e is the third point from the beginning of the GPS log. The SED is calculated for the GPS point between the $anchor$ and the $float$, respectively. If there are no points at which the SEDs are above the threshold, the $float$ is changed to the next GPS point. Otherwise, the point with the highest SED among the points whose SED are above the threshold is the compressed point and the next $anchor$.

The process of computing the SED of a GPS point between each $anchor$ and $float$ is repeated. This process is repeated until the OW terminates when GPS positioning is finished. Figure 3 shows an example of a compressed GPS log with OW applied.

The difference between our proposed OW and the original OW is how to choose $anchor$ and $float$. In original OW, where an $anchor$ is the first point in the GPS log, and $float$ is the third point from the GPS log's beginning. On the other hand, our method uses the results of map-matching to select $anchor$ and $float$.

3.3. *StayPoint*

The StayPoint represents a point where the user was in a range for a amount of time. A GPS log that becomes a StayPoint satisfies the following conditions.

$$m < i \leq n, Int(P_m, P_n) \geq T_r \quad (5)$$

$$D(P_m, P_i) \leq D_r, Dist(P_m, P_{n+1}) > D_r \quad (6)$$

P_i is a point of trajectory in the GPS log. $D(P_m, P_i)$ is the distance between the two points P_{m} and P_i based on the Hubeny formula. D_r is the threshold for the distance between P_m and P_i. $Int(P_m, P_n)$ is the difference in time between P_m and P_n. T_r is the threshold for the difference in time between P_m and P_n. From the GPS log that satisfies these equations, we calculate the StayPoint. Let StayPoint be s, then $s = (x, y, t_a, t_l)$. $s.x$ and $s.y$ represent the coordinates of s. $s.t_a$ represents the start time of the user's stay at the StayPoint. $s.t_l$ represents the time when the user ends his stay at the StayPoint. $s.x$, $s.y$, $s.t_a$, and $s.t_l$ are computed by following equations.

$$s.x = \sum_{i=m}^{n} \frac{P_i.x}{|P|}, s.y = \sum_{i=m}^{n} \frac{P_i.y}{|P|} \quad (7)$$

$$s.t_a = P_m.t_m, s.t_a = P_n.t_n \quad (8)$$

Our method extracts the StayPoint from the GPS log and preserves the state of staying at one point in the GPS log before compression. In our method, s is represented by $sa = (x, y, t_a)$ and $sl = (x, y, t_l)$ to describe the same format as the traditional GPS log.

3.4. *CriticalPoint Algorithm*

CP is a compression method that focuses on the speed and direction of the user's movement. At two consecutive points P_i, P_{i-1} in the GPS log, the compression point satisfies either of the following conditions.

$$SD(P_i, P_{i-1}) > Sr, AD(P_i, P_{i-1}) > Ar \quad (9)$$

$SD(P_i, P_{i-1})$ calculates the difference in speed between P_i and P_{i-1}.$AD(P_i, P_{i-1})$ calculates the angle difference between P_i and P_{i-1}. Sr is the threshold for the difference in velocity between P_i and P_{i-1}. Ar is the threshold for the angle difference between P_i and P_{i-1}.

3.5. *Integrated processing of compression points*

If there is SP or CP compression point within a certain distance from the OW compression points, our method deletes the OW compression points. This reason is compression points by OW preserve only the shape features, but the SP or CP preserve velocity or state of staying in addition to the shape features. The proposed method regards the points obtained as a result of this process as compressed points.

4. Experiment

We evaluate the performance of our method for the GPS logs compression preserving the features of the GPS logs while maintaining the Compression Rate. We used the GPS log of buses in Okayama Prefecture, Japan. The number of GPS logs is 297. The average number of GPS points in the GPS log is about 185.9. The average positioning interval of the GPS logs is about 18.3 seconds. Also, map-matching uses the road network of Okayama Prefecture obtained from OpenStreetMap.

This experiment uses three evaluation criteria: SED Error, Speed Error, and Compression Rate. The SED Error is the difference between the shape of the GPS log before and after compression. This value is the median and average of the SED calculated from the GPS logs before and after compression. Speed Error is the difference between the speed of the GPS log before and after compression. Compression Rate is the compression rate of the GPS logs

The proposed method parameter is the threshold of the SED, which is used to find the compression point in the OW. We changed the threshold of the SED by ten from 30 to 200. For the GPS logs compressed by the proposed method, we evaluate them by each evaluation criterion.

4.1. *Experiment Results and Discussion*

Figure 4, Figure 5, and Figure 6 show the evaluation results of SED Error, Speed Error, and Compression Rate of the GPS logs compressed by our method. Their

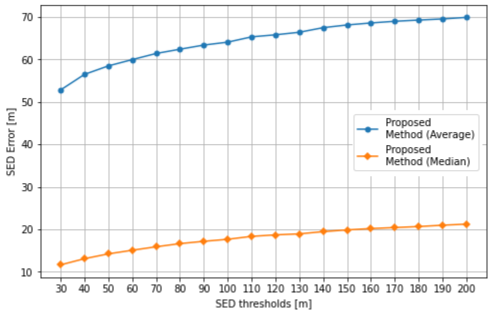


Figure 4 SED Error.

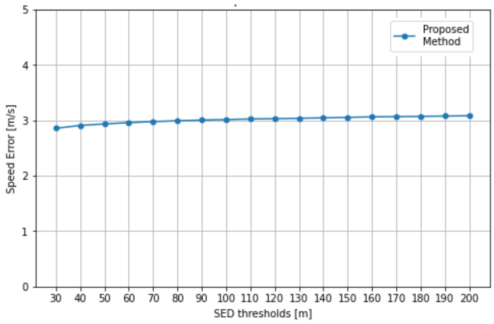


Figure 5 Speed Error.

horizontal axis is the threshold value of the SED. The vertical axis is the value of each evaluation criterion.
In Figure 4, the average value of SED Error is large more than the median. This result shows that our method can compress almost GPS logs preserving the feature, but the method when compressing some GPS logs failed.

The reason why the proposed method has tremendous SED Error values in some GPS logs is due to calculating the pseudo arrival time of the $float$ used in the OW of the proposed method. The pseudo arrival time assumes that the user moves straight from the GPS point to the $float$. However, because roads are not always straight but have curves, there is a difference between the expected and the actual arrival time.

The Speed Error in Figure 5 shows that the Speed Errors are about 3m/s in all SED thresholds. When the threshold of SED was changed, the value of Speed Error did not change much. This result indicates that the proposed method is not much affected by the threshold of SED and can preserve velocity features.

Figure 6 shows the Compression Rate results of the proposed method. As the threshold of SED increases, the Compression Rate also increases. This result shows that the proposed method can change the Compression Rate by varying SED's threshold value. Figure 4 and Figure 5 show that the proposed method can preserve velocity and shape features, even when the Compression Rate is

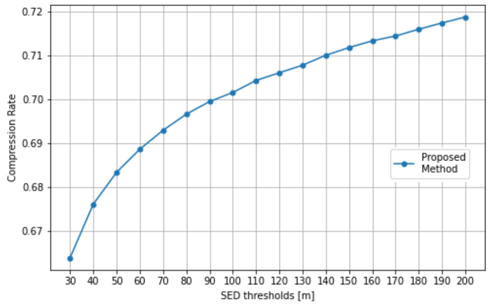


Figure 6 Compression Rate

increased. Therefore, our method can compress GPS logs while considering the trade-off between the shape and velocity information and the Compression Rate.

5. Conclusion and future work

This paper proposes an online compression algorithm for GPS logs that uses three methods to preserve the feature points. Our experiments show our method can compress GPS logs preserving velocity and shape features.

Future work will improve map-matching performance because the compression of our proposed method depends on the performance of the map-matching.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP19K20418 and Grant for Promotion of OUS Research Project (OUS-RP-20-3).

References

1. Douglas, D. H. and Peucker, T. K.: Algorithms for the reduction of the number of points required to represent a line of its caricature, IJGIG, Vol. 10, No 2, pp. 112–122, 1973.
2. Meratnia, N. and de By, R. A.: Spatiotemporal compression techniques for moving point objects, EDBT, pp. 765–782, 2004.
3. Schmid, F., Richter, K. F. and Laube. P.: Semantic trajectory compression, SSTD, Vol.5644, pp. 411–416, 2009.
4. Barbeau, S., Labrador, M. A., Perez, A., Winters, P. Georggi, N., Aguilar, D. and Perez, R.: Dynamic management of real-time location data on GPS-enabled mobile phones, UBICOMM, pp. 343– 348, 2008.
5. Zheng, V. W., Zheng, Yu., Xie, X. and Yang, Q.: Collaborative location and activity recommendations with GPS history data, WWW, pp. 1029–1038, 2010.
6. Brakatsoulas, S., Pfoser, D., Salas, R. and Wenk, C.: On map-matching vehicle tracking data, VLDB, pp. 853–864, 2005

Case Study and Direction of Bicycle-Sharing Systems in Japanese Cities

Minoru Kumano
Faculty of Regional Innovation, University of Miyazaki, Japan
E-mail: kumano@cc.miyazaki-u.ac.jp

Toru Hiraoka
Department of Information Systems, University of Nagasaki, Japan
E-mail: hiraoka@sun.ac.jp

Abstract

Bicycle-sharing systems allow people to share bicycles with others, borrow and return them using multiple cycle ports, and register and pay for them with an IC card or smartphone. Because these systems are still new in Japan, few papers can be found on them, even in the National Institute of Informatics. The purpose of this article is to clarify the business purpose, issues, effects, features, numerical values, and direction of bicycle-sharing systems. Therefore, we selected 16 advanced cases nationwide and conducted a survey in 2019. As a result, it was clarified that the main issues facing bicycle-sharing systems are improving the ease of migration and supplementing public transportation, improving profitability, and so on.

Keywords: Bicycle-Sharing System, Case Study, Direction

1. Introduction (Background and purpose of research, method)

In Japan, the Bicycle Utilization Promotion Law was passed in 2017, and town development and bicycle-sharing businesses have been progressing since. Bicycle-sharing is a system that allows users to share bicycles and freely get on and off at multiple ports.

In addition, bicycle-sharing systems alleviate traffic congestion in cities, save energy, prevent air pollution, reduce travel costs, and improve public health. Therefore, they have recently become widespread in urban municipalities nationwide. However, previous research is not sufficient, and no paper has considered the management trends, issues, and directions of bicycle-sharing systems by conducting a survey of advanced local governments all over Japan.

To fill this gap, 16 advanced cases nationwide were selected and a survey was conducted in 2019. The purpose of this article is to clarify the business purpose, issues, effects, features, numerical values (number of ports, number of bicycles, etc.), and direction of the 16 cases and to secure useful materials for the introduction of bicycle-sharing systems in the future. For the selection of the precedent cases in April 2019, local governments with a proven track record were selected from online searches, literature, and interviews with the Ministry of Land, Infrastructure, Transport and Tourism.

A survey of local governments was conducted from May to October 2019. In addition, we visited with the majority of the selected local governments to conduct interviews and collect materials.

2. Research results

① The most common purpose of the bicycle-sharing businesses is to improve the mobility of the town (10 cases). In particular, the effects of revitalizing the central city area and enhancing the convenience of transportation for tourists have increased.

Next, bicycle-sharing systems aim to help society break away from its dependence on automobiles (4 cases) and reduce the environmental burden (4 cases) by switching to using bicycles. The third most common purpose is to supplement public transportation (3 cases) and focus on bike-sharing as a means of promoting the use of public transportation.

Interregional exchange or the promotion of recreation (1 case) is also the goal of a wide-area bicycle-sharing system that spans multiple prefectures.

② In terms of actual effects or advantages, bicycle-sharing systems contribute most to tourism (7 cases). In particular, they are highly effective as a means of transportation that could increase the number of users for sightseeing purposes and help tourists plan sightseeing excursions at their own pace.

Next, we have achieved about 50% of the purpose of improving the ease of migration (5 cases). In addition, the supplementation of public transportation (4 cases) has also been effective, including the development of regional transportation and the establishment of regional transportation networks. The introduction of electric bicycles (4 cases) has been well received by users, and people are able to travel smoothly. PR (6 cases) concerning these efforts has also been effective.

There were three PR projects that improved the recognition of this project throughout the region. According to survey participants, the development of guided bicycle tours, the suppression of the inflow of automobiles into the city (2 cases), and the conversion from automobiles to bicycle use (1 case) have been achieved.

③ The most common problems/issues are the improvement of expenses such as balance of income and expenditures and costs (8 cases). Regarding this, survey participants expressed the opinion that a review of the fee structure should be considered in consideration of the profitability of the business.

The next most common issue is securing more users and creating PR for that purpose (5 cases). The third issue is the relocation of bicycles (4 cases), which is based at each port. Since it is possible to return a bicycle to a port different from the one from which you rented, the number of bicycles at the ports will be uneven. The high labor costs necessary for the relocation of bicycles is still an issue. In addition, there is a problem of selecting candidate locations for port installation (3 cases). It is necessary to take into consideration the assumption of possible harmful effects such as new construction, location conditions that are more effective in improving the utilization rate, and an increase in the number of illegally parked bicycles. The problem of manners when users use bicycles and the risk of accidents (3 cases) were also issues. Problems with maintenance of management systems and equipment (2 cases), simplification of membership registration (1 case), improvement of the environment for inbound tourists (1 case), and low use by citizens (1 case) were also mentioned.

④ Relationship between population size, number of bicycles, and number of ports

See the correlation between the population size of each municipality, the number of bicycles installed, and the number of ports in the bike-share example.

It is expected that the bicycle-sharing system of a city with a large population will have a large number of bicycles and ports. Figures 1 and 2 are correlation diagrams of population size, the number of bicycles installed, and the number of ports; they were created from the data in Table 1. Regarding population size and the number of ports, the larger the population, the larger the number of ports and the correlation. On the other hand, regarding the relationship between population size and the number of bicycles, there are some data with large outliers, and there is little correlation in the overall value. The data with particularly large outliers is that of "Shimanami Rental Cycle (Imabari City / Onomichi City)." Unlike other local governments, Shimanami Rental Cycle emphasizes external effects such as wide-area sightseeing tours. As a result of arranging bicycles for tourists, the proportion of the local population is higher than in other cases, i.e., the number of bicycles was large. However, Imabari and Onomichi on the Shimanami route; Takamatsu, a city that frequently uses bicycles; and Koto Ward, which is located in Megalopolis in the metropolitan area, are considered to be special, and there is a correlation in cities other than these.

Table 1. Contents of advanced cases of bicycle-sharing systems in each local government in Japan

Municipality	Okayama City	Miyazaki City	Koto-ku, Tokyo	Takamatsu City	Imabari City	Onomichi City	Fujieda City	Fukuoka City
Name	Momochari	Miyako PiPPA	Koto Ward Seaside Community Cycle	Takamatsu City Bicycle Rental	Shimanami Rental Cycle		Fujieda Trip (bike share in the city)	Melchary (currently in the demonstration experiment stage)
Start date	7/27/2013	Jul-17	Nov-12	May-01	Apr-99	May-99	Mar-18	Mar-15
Number of bicycles	332 units (as of H27)	200 units	1,207 units (as of R1.8.30)	1,250 units (as of R1)	1,800 units	1,025 (Onomichi City side as of August 2019)	45 units (with electric assist)	About 1,000
Number of ports	35 locations (1 of which is closed as of R1.8)	42 places	117 locations (as of R1.8.30)	7 places	13 locations (9 locations in Imabari City)	13 locations (5 locations on the Onomichi side)	16 places	More than 180 locations (as of H30)
Registration number	91,714 people (as of H30 degrees)	2,700 (as of H31.6.30)	128,804 people (as of December 31, H30)	---	---	---	---	---
User registration [A: port, B: smartphone app, C: web, D: window]	A	B	C	D	D		B	B
Payment method [A: cash, B: credit card, C: mobile phone payment, D: IC card, E: electronic money, F: other]	A, B, D, F	B, C, D, E, F	A, B, C, D,	A, D	A		B, C	B, F
Operational trends	[Number of uses] Average 1,453 times a day, total number of uses 530,324 times, turnover rate 3.85 times / day / unit (H29 degrees), H30 degrees cumulative number of uses about 1.63 million times	The breakdown of the occupancy rate of the port utilization rate is 62.1% for public land (17 locations), 37.7% for private land (25 locations), and 0.2% unknown.	Bicycles may be parked excessively, and collection and relocation work to level the number of bicycles is increasing.	Operated by bicycles with abandoned bicycles Annual usage: Approximately 310,000	Approximately 150,000 annual users	[Number of loans (fiscal year)] H25: 81,851 H26: 116,303 H27: 135,229 H28: 141,205 H29: 149,740 H30: 132,075	Reiwa first year budget amount: 10,000,000 yen * 2017 budget amount: 2,000,000 yen, 2018 budget amount: 4,500,000 yen	Mercari dismantled its consolidated subsidiary Sozo in June 2019 and started a new company "neuet" in August. Operated in the form of a joint system between "Co., Ltd." and "Clara Online Co., Ltd."
Port with many users	---	High utilization rate of ports on public land under demonstration experiment	Mostly used for commuting on weekday mornings and evenings. In particular, there are many uses of ports near tower apartments and stations.	JR Takamatsu station square underground port and Kotoden Kawaramachi station underground port	---	---	---	---
Pros	・Change of traffic behavior, revitalization of town ・Reduction of environmental load	Environmentally friendly and eco-friendly, contributing to health promotion	Reduction of CO_2 by reducing the movement of automobiles, improvement of attractiveness of the introduction area, improvement of convenience of movement, creation of liveliness by improvement of mobility in the area, etc.	Control the total amount of bicycles through mutual use and effectively utilize illegally parked bicycles	Promotion of stay-type tourism, promotion of health, promotion of purpose of life and friendship, realization of a society with reduced environmental load, promotion of inbound tourism due to international popularity	You can enjoy the beauty of the islands of Setouchi, interact with the locals, and enjoy gourmet food at your own pace.	① Cheaper than a bus and convenient with no waiting time. You can easily climb a slope with an electric bicycle. ② There is a station at Fujieda station. Available 24 hours a day. ③ There is no problem even if you use it frequently. It's cheaper than parking in the parking lot.	・Use unused space as a port (individual application required) ・Can travel short distances faster than using public transportation
Problems/issues	・Simplification of membership registration method ・Dissemination of first-time users ・PR for increasing users for sightseeing purposes ・Elimination of bicycle bias at each port	There is concern that the risk of accidents due to crossing with automobiles and pedestrians is increasing due to the lack of following traffic rules such as driving on the left side of the road.	・Improved income and expenditure by streamlining the collection and relocation of expensive bicycles ・Optimization of the number of bicycles at the cycle port	・Improvement of balance of income and expenditure ・Update of management system due to deterioration over time	Inbound environment maintenance (Wi-Fi environment, communication means, cashless payment)	As the number of cyclists increased, we began to hear complaints about driving manners	・Implementation of effective PR to increase the number of users ・Selection of station installation locations that are expected to be used ・Reduction of running costs	・ There is a possibility that the correct position of the bicycle will not be displayed on the app due to GPS accuracy ・There are users with bad manners such as personalization ・Because there is no reservation system, other people will use the bicycle first (possibility of not being able to ride when you want to)
Direction	---	---	Examining and implementing problems and solutions to problems in demonstration experiments, and shifting to full-scale implementation	With the entry of private businesses, we will consider the future of the rental cycle business while considering a certain division of living according to each role and in cooperation with private businesses.	Expanding the base of cyclists by attracting non-cyclists such as the elderly and women	We will continue to disseminate information on the appeal of Shimanami Kaido Cycling both domestically and internationally, and strive to improve the driving environment and manners.	While taking in the opinions of users through effective PR and surveys, we will improve the utilization rate and ease of migration. Since it is an initiative that utilizes IoT, we will consider how to utilize usage data.	Business succession was passed to "neuet Co., Ltd." and the Mercari business as a new system started in collaboration with "Clara Online". "The purpose of the new system is to focus on Mercari and Mercari. Results of business selection and further business expansion"

Table 1. (Continued)

Municipality	Kagoshima City	Kitakyushu	Oita City	Hiroshima City	Fukui City	Toyama City	Himeji City	Kanazawa
Name	Kagoshima City Community Cycle "Kagorin"	Kitakyushu City Bike	Oita Cycle Share	Hiroshima City Peasukuru	Fukui City	Toyama City Avile	Himeji City Himeji	Kanazawa City Machinori
Start date	Mar-15	Mar-10	Oct-18	Feb-15	2013	Mar-10	Jul-16	Mar-12
Number of bicycles	192 units	246 units	300 units	About 500	44 units	255 units	150 units	155 units
Number of ports	25 locations (as of R1 / 7)	21 places	35 places	62 locations	[Number of ports] 16 locations (as of R1 / 7)	[Number of ports] 23 locations (as of R1 / 9)	[Number of ports] 20 locations (as of R1 / 7)	[Number of ports] 21 locations
Registration number	---	535 general members Shear type (corporate) 45 (as of R1 / 5)	Approximately 1,000 people (as of R1 / 7)	---	---	---	180 units (as of R1 / 7)	296,616 people (H24 / 3 / 21-H31 / 3/31)
User registration [A: port, B: smartphone app, C: web, D: window]	A, D	C, D	B, C	C, D	D	D	A, C	A, D
Payment method [A: cash, B: credit card, C: mobile phone payment, D: IC card, E: electronic money, F: other]	A, B, E	A, B, C	B, F	A, B, C, D	A	A, B, F	A, B	A, B
Operational trends	The number of uses increases year by year (the total number of uses until the end of July of the first year of Reiwa is about 670,000)	Number of users: 147 units / day (2018)	The monthly membership free campaign was held from February 12, 2019 to June 30, 2019, and the usage increased significantly.	Consignment contract amount 2014 ... 20,000,000 yen 2015 ... 18,000,000 yen 2016 ... 14,000,000 yen 2017 ... 0 yen 2018 ... 4,300,000 yen	The number of uses is increasing year by year, and the demand is increasing, especially in. We are able to operate the business on a low budget while coordinating with the private sector.	---	The share cycle has become established as a secondary means of transportation, and the number of users and the number of times of use are increasing year by year.	Number of uses 178,619 times, number of users 63,284 (2018)
Port with many users	---	---	---	---	Port around Fukui Station	Three stations at the south exit of Toyama Station, Nishicho Station (located in front of Toyama City Hall), and in the city center	Ports around Himeji Station, often used for commuting and sightseeing	Most often used at Kanazawa Station, the starting point, and ports around major tourist destinations
Pros	Reducing environmental impact, improving mobility and convenience, responding to a wide range of tourism needs (attractiveness of town tours unique to bicycles)	By shifting to the use of bicycles, it is possible to reduce the environmental load by means such as reduction of carbon dioxide emissions, electronic assistance, and "drop offs" between stations.	・You can drop off ・You don't have to wait like for a bus ・You can use it 24 hours a day ・ Contribute to the establishment of a regional transportation network by combining with the public transportation network	Because it is a city of deltas, you can smoothly tour the city with many undulating bridges by electric bicycle	All of them are equipped with electric power assist, and rentals are accepted face-to-face, so we can provide tourists information as well as rental services.	Cyclocity Co., Ltd., the operator, is operating smoothly and safely by making the best use of the know-how cultivated so far.	It can be used with cash, and it can be used immediately without the need for pre-registration or other operations.	・If you have a credit card, you can easily register on the spot ・There are many narrow streets in the city of Kanazawa, and the main tourist spots are concentrated in a narrow area, so it is suitable as a means of transportation for tourists
Problems/issues	Continue to strive for public relations and maintenance of equipment, etc.	・Ensuring business profitability ・A large ratio of labor costs to correct the bias of bicycles between stations	・It costs a lot of "relocation" to correct the bias of bicycle placement ・It takes a certain period of time to increase the number of users	As a candidate for port installation, it is a highly convenient place, but there are times when it is necessary to abandon the installation due to concerns about problems caused by installing the port, such as the number of illegally parked bicycles increasing	Business management that is conscious of the difficulty of establishing a new lending port and profitability	Stable securing of financial resources such as operation and maintenance costs for properly conducting the bicycle-sharing business	・Efforts to increase the number of users ・Naming rights and other advertising revenues need to be expanded, fee structure reviewed and costs reduced	・Low citizen use (about 90% of the total is for tourism) ・Increase in repair costs over time
Direction	By communicating the significance of riding a bicycle in an easy-to-understand and polite manner, we will further increase interest and understanding of bicycles	Aiming to continue business while selecting stations that better meet needs	We would like to set up a service area so that the relocation work will not be excessive, balance the income and expenditures, and verify that it will be established as a sustainable new transportation system.	High density of cycle ports so that domestic and foreign tourists and other visitors can comfortably visit tourist facilities in the city area and citizens can conveniently use the ports as a means of daily use that complements public transportation. Promote tourism and regional revitalization using bicycles	We increased the operating rate of bicycles by driving up demand, and based on that situation, we considered establishing new rental ports and increasing the number of turntables.	The expansion of stations and services has just been implemented by the operator in consideration of profitability, etc., and there are no plans to expand at this time.	While considering the number of bicycles and the arrangement of stations that meet the needs, we will proceed with the introduction of an "in-vehicle share cycle" that can reduce operating costs compared to the current system as a "next generation Hime-chari."	A new system started from March 2nd, year of Reiwa. Basic direction: (1) Encourage citizens to use (currently, users are mainly tourists); (2) Respond to new tourism demand based on the central city area; (3) Maintain public-private management and other electric assists in order to comply with the policy of town development Equipped with functions and GPS functions

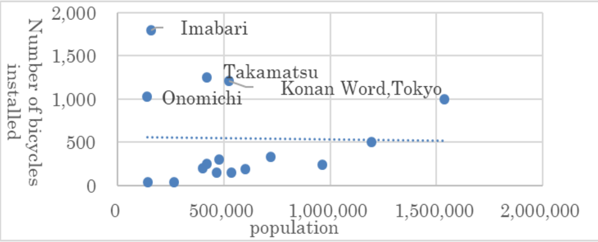


Figure1. Relationship between population size and number of bicycles

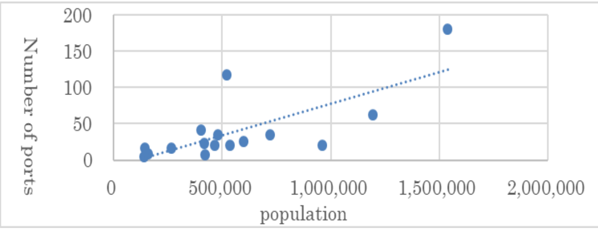


Figure2. Correlation diagram between population size and number of port

3. Conclusion

According to a nationwide survey, the number of cycle ports and the number of users are correlated in the bicycle-sharing businesses located in urban areas, and the cycle ports located at railway stations representing the cities have the highest utilization rates. Cycle ports with heavy traffic at transportation hubs have many users, complementing public transportation; shortening travel times for sightseeing, commuting to school, business, etc.; and increasing mobility. However, it is difficult to make a profit in the bicycle-sharing business, so most of these businesses are subsidized by the local government. Large-scale bicycle-sharing businesses with a relatively large number of cycle ports and bicycles have a higher turnover rate and are more likely to improve profitability than small bicycle-sharing businesses.

Therefore, for the sustainable development of bicycle-sharing businesses in the future, it is important to secure the number of bike-share ports and the number of bicycles in a convenient position commensurate with the size of the population and to devise ways to improve profitability, for example, in Toyama City. As you can see, new ideas and implementations are needed, such as attaching electronic public notices to cycle ports to earn sponsorship income.

References

1. https://shimanami-cycle.or.jp/rental/course
2. http://takamatsu-parking.com/for_use_onetime_rider_2017/
3. https://www.city.toyama.toyama.jp/kankyobu/kankyoseisakuka/ondankataisakukikaku/jitenshakyodoriyo.html
4. https://www.machi-nori.jp/
5. https://docomo-cycle.jp/koto/whatiscs/
6. https://www.keifuku-hotels.com/news/?p=65
7. https://www.momochari.jp/howto/index.html
8. http://www.city.hiroshima.lg.jp/www/contents/1581552607625/index.html
9. http://www.city.kagoshima.lg.jp/faq-kankyomachizukuri/kanseisaku/q9.html
10. https://hachimansama.jp/blog/9305#
11. https://www.itmedia.co.jp/news/articles/1802/13/news115.html
12. https://www.kobe-np.co.jp/news/odekake-plus/news/detail.shtml?news/odekake-plus/news/pickup/201809/11648577
13. https://www.city.miyazaki.miyazaki.jp/city/policy/urban_planning/164580.html
14. https://www.city.fujieda.shizuoka.jp/trip/oshirase/1521184175837.html
15. https://rembrandt-group.com/oita/7842.html

The IoT Solution to Learn English Words

Shogo Aizawa, Motohide Yoshimura
Department of Information Systems, Faculty of Information Systems, University of Nagasaki
Japan

1-1-1 Manabino, Nagayo-cho, Nishi-Sonogi-gun, Nagasaki 851-2195, JAPAN
E-mail: yxsimura@sun.ac.jp
http://sun.ac.jp/e

Abstract

We innovate an IoT solution to play the English word learning system. It uses a Raspberry Pi and two NFC readers. You can learn English words in three steps by using our system. First, prepare NFC tags corresponding to English words and illustrations, such as a cat, dog, chicken etc. Next, select a pair of the tag corresponding to an English word and illustration. Finally, hold them over the each NFC reader, then the answer that is correct or incorrect are displayed on a touch panel. Our solution has two aspects. The one is an educational system for children and the other is an IoT toy playing with children. As to the former, the children can learn English words by intuitive operation such as operating the touch panel and holding the NFC tags. As to the latter, they can experience a part of an IoT technology. In this paper, we report findings through the construction of the English word learning system..

Keywords: English word learning system, IoT, NFC, educational system, IoT toy

1. Introduction

Information and communication technology is now an integral part of our lives. In 2019, 96.1% of households own mobile terminals, and 83.4% of households own smartphones, there is more than 80% of households [1]. The development and spread of ICT has brought about major changes outside the ICT field. The role of ICT in the field of education is also important. In 2020, programming education became compulsory in elementary schools, and the new teaching guidelines for elementary schools announced in March 2017 stated as follows, "*In order to cultivate the ability of information utilization, ICT equipment such as computers and networks will be provided at every school. We will equip the appropriate environment to utilize these tools and enhance learning activities* [2]." The basic idea of ICT utilization by the Ministry of Education is as follows, "*Based on the new teaching guidelines, there are three important aspects. The one is the balanced development of the three pillars of qualities and abilities. The second is the aggressive utilization of ICT as the teaching materials, the teaching and learning tools according to both of the actual conditions of children and schools, and characteristics and learning steps of each subject. The third is the class improvement for the proactive, interactive and deep learning.*[3]" In addition to ICT utilization such as efficient and effective presentation using the large presentation screen, the goal is to prepare a terminal for each student and promote learning independently and interactively. By using ICT equipment, it is possible to study efficiently during the limited learning time. Aggressive utilization of ICT is an important issue for enhancing future educational activities.

The utilization of ICT equipment is underway in all subjects. Especially, the use of ICT in English education is being actively promoted, and 99.1% of schools utilize ICT equipment in English education in elementary schools [4]. However, according to a survey of Japanese children's use of digital devices in schools at OECD PISA 2018, the percentage of students who answered that they did not use ICT devices in their school's "*foreign language*" class was 67.0%, it is the worst among participating countries [4]. It can be said that the environment in which teachers utilize ICT in English education is in place, but the environment in which students themselves use ICT to learn English is not sufficiently developed.

In this research, we will develop an English learning system by using ICT, especially IoT. The IoT is a technology that connects things via the Internet, as translated as the "*Internet of Things*". Simply speaking, it is the immediate reflection of happenings in the real world to the digital world. With this technology, both analog and digital advantages can be leveraged. An IoT toy is the one that incorporates the elements of IoT. By utilizing this element, not only will it attract the interest of children, but it will also be possible for them to experience ICT with simple operations. By using our system, users can learn English while playing and can come into contact with IoT, which is a part of ICT. In this paper, we report the findings obtained in the construction of the system. We will show examples of using English learning tools in Section 2, and describe examples of IoT toys for children in Section 3. In Section 4, we will describes our system using NFC. In Section 5, we will summarize the paper.

2. Examples of using English learning tools

2.1. *Current status of English learning*

As in the case of compulsory English education in elementary schools in 2020, the timing to start learning English is getting earlier year by year. Specifically, the "*foreign language activities*" that were conducted in the 5th and 6th grades in the teaching guidelines before 2020 will be advanced to the 3rd and 4th grades, and English will be started as a subject of "*foreign language*" from the 5th grade. The goal of "*foreign language activities*", which will be held in three or four years in the elementary school, is to familiarize yourself with English through activities such as quizzes, dance, and singing. By starting to learn English from an early stage, children can take advantage of their flexible adaptability, develop a positive attitude toward communication, and become familiar with English expressions. In addition, more than 70% of junior high school students wanted to read English words and sentences in elementary school, and felt that they wanted to learn English words among elementary school [5]. By becoming accustomed to English in elementary school, children will be able to acquire a foundation for English education after junior high school, which will lead to improvement of their English proficiency.

Familiarity with English in active English learning in elementary school is very important in learning English after junior high school. The liking and proficiency of English in junior high school has a great influence on the frequency and amount of learning of English [6]. It is effective to "*like*" English in order to effectively advance learning English. For that purpose, it is necessary for students to "*like*" English from the English education of elementary schools and to maintain their interest in English. A game that utilizes English is used as a way to invite the interest. It has been found that learning English in combination with games is effective in familiarizing yourself with English and practicing and establishing the vocabulary that is the basis of foreign language activities [7].

2.2. *Flashcard case*

Flashcards are used as teaching materials when you start learning English. A flash card is a set of a card with an illustration and a card with words written on it. It is used in school lessons and English conversation classes, and the teacher can show the illustration card and the English word card at the same time to teach pronunciation, or show only the illustration card and let them answer the English word. You can also incorporate game elements such as playing with listening to English and taking illustration cards like playing cards. Since it is a simple structure, it can be used in various ways by the ingenuity of the teaching side. Currently, various types of flash cards are on sale, so you can choose a flash card that suits your purpose, such as age and type of English words.

There are also flashcards that incorporate ICT elements [8]. A flash card can be displayed on an ICT terminal and used in the same way as a conventional flash card. In addition, since the native speaker's voice is attached, you can hear and acquire the correct pronunciation of the native speaker. By using an ICT terminal, students can be made to hear the correct pronunciation of native speakers without the teacher knowing the correct pronunciation of English.

3. Examples of IoT toys for children

3.1. *Overview of IoT toys*

With the development of ICT in recent years, IoT is also being incorporated into toys for children. Products so-called IoT toys are on the market. You can experience the world of the Internet with more intuitive operations than using a smartphone or PC, expanding the range of play from the analog world to the digital world. It is also possible to get interested in the ICT field by playing while touching a part of IoT technology. In addition, IoT toys are more attractive than their toys by connecting smartphones, game consoles, and toys via the Internet. IoT toys also have an aspect as educational toys, and are attracting attention as toys that can be learned while playing. It can be said that IoT toys are also very useful as a means to improve their abilities in the present age when programming education is emphasized. Here, we will introduce specific examples of IoT toys in Japan.

3.2. *CORO-PUTER-Ja.*

"*CORO-PUTER-Ja.*" is a wooden IoT educational toy developed by TDS Co., Ltd. [9]. When a wooden hiragana block with an embedded RFID tag is rolled onto a wooden frame with an embedded reader, the reader reads the tag and the hiragana is displayed on a smartphone connected via Bluetooth. This allows you to learn letters and words and get in touch with technology through the play of rolling blocks. Even infants can easily pick up blocks, carry them to the frame, and roll them, so they can learn while playing. It can be said that the strength of IoT toys is that a lot of information can be obtained with this simple operation.

4. English learning system using NFC

4.1. *Features of our English learning system*

In this research, we innovate an IoT solution to play the English word learning tool as a framework of English learning by IoT toys and prototype an English learning system for children that aims to integrate English learning and ICT. By using this tool, you can learn English while playing with IoT toys, and you can also experience some of the IoT technologies.

4.2. *Development environment*

In this system, the users proceed the English word learning, while the contents of the NFC tag read by the NFC reader are displayed on the touch panel. The system consists of Raspberry Pi 4, and the OS Raspbian (version 10 (buster)) as the Platform, and two NFC readers and English words and illustration cards at which the NFC tags attached. Python (version 3.7.3) is used to construct the system. NFC Py (version 1.0.3) is used to operate the NFC reader.

4.3. *System overview*

We will explain the outline of our system. Figure 1 shows the system configuration diagram.

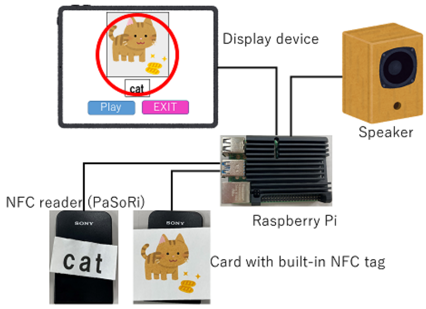


Fig1. System configuration diagram

This tool processes the content read by the NFC reader with Raspberry Pi and reflects it on the touch panel. Sony's PaSoRi RC-S380 / S is used to read NFC tags, Kuman's 7-inch touch panel is used to display the screen, ZoeeTree S1 portable Bluetooth speaker is used to play audio, and Raspberry Pi 4 is used to process and execute codes. The purpose of this system is to learn the English names of animals. There are two kinds of cards, "*Animal Card*" with animal illustrations and "*Animal Name Card*" with animal English names as shown in Fig. 2.

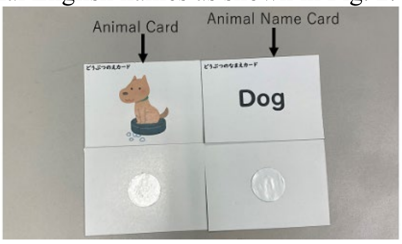


Fig2. Card with NFC tag attached

We attach the NFC tag to the back of each card. In this system, unique IDm of each NFC tag is associated with each of the "*Animal Card*" and "*Animal Name Card*", and the correct answer or incorrect answer is determined by pairing two IDm in the code. The NFC tag used for each card is Sanwa Supply's NFC tag (product part number: MM-NFCT). Two NFC readers are prepared so that the "*Animal Card*" and "*Animal Name Card*" can be read at the same time. They are decorate as shown in Fig. 3 to clarify the reader on which each card should be placed.

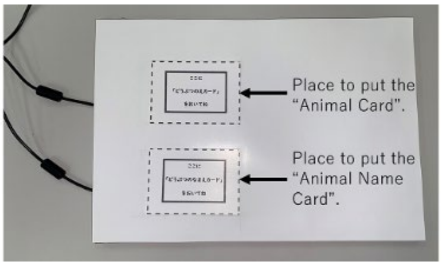


Fig3. Decorated NFC readers

As shown in Fig. 4, place the "*Animal Card*" and "*Animal Name Card*" on each reader, and press the "*Play*" button on the touch panel to execute the code and display the correct or incorrect answers. When reading a card, the IDm of the NFC tag is read, and the illustration and name of the animal on the card corresponding to that IDm are displayed on the touch panel.

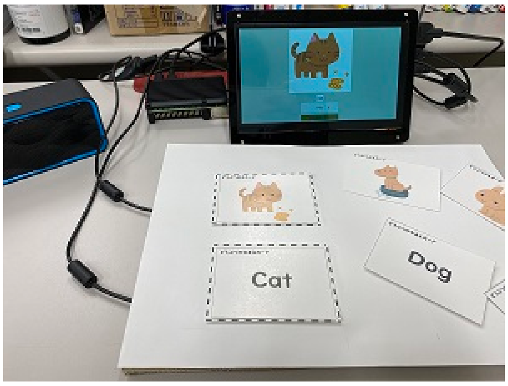


Fig4. Operating environment

4.4. *System behavior*

The user places the "*Animal Card*" and "*Animal Name Card*" with the NFC tag attached on the corresponding NFC reader, and presses the "*Play*" button to confirm the correct or incorrect. In case that the answer is correct, for example, the dog's "*Animal Card*" and the dog's "*Animal Name Card*", a sound indicating the correct answer and an illustration (〇) indicating the correct answer are displayed on the touch panel. In case that the answer is incorrect, for example, a cat's "*Animal Card*" and a dog's "*Animal Name Card*", a sound indicating an incorrect answer and an illustration (×) indicating an incorrect answer are displayed on the touch panel. In addition, only if the answer is incorrect, the correct animal bark that matches the "*Animal Card*" is played through the speaker, and the frame of the "*Animal Name Card*" on the screen is blinked to suggest another "*Animal Name Card*". The reason for playing the correct animal bark that matches the "*Animal Name Card*" is to make it easier to imagine an animal illustration from the bark.

5. Conclusions

In this research, we prototyped an English word learning system that integrates English learning and IoT by using NFC tags. Our system can memorize English words and illustrations as a word meaning in combination by reading NFC tags with two NFC readers. In the future, we have a plan to enhance the mechanism so that children can learn with more interest.

References

1. Ministry of Internal Affairs and Communications, *2020 WHITE PAPER Information and Communications in Japan*, 2020, in Japanese.
2. Ministry of Education, *Elementary School Programming Education Guide (3rd Edition)*, https://www.mext.go.jp/content/20200218-mxt_jogai02-100003171_002.pdf(accessed 2020-12-8), 2020, in Japanese.
3. Ministry of Education, *Effective use of ICT in teaching each subject*, https://www.mext.go.jp/content/20200911-mxt_jogai01-000009772_19.pdf(accessed 2020-12-8), 2020, in Japanese.
4. Ministry of Education, *Effective use of ICT in teaching foreign languages*, https://www.mext.go.jp/content/20200911-mxt_jogai01-000009772_13.pdf (accessed 2020-12-8), 2020, in Japanese.
5. Ministry of Education, *Current status, achievements, and challenges of foreign language activities in elementary schools*, https://www.mext.go.jp/b_menu/shingi/chousa/shotou/102/shiryo/attach/1347444.htm (accessed 2020-12-8), 2020, in Japanese.
6. KANEKO Yoshitaka, *Educational Implications to Help Students to Enjoy Learning English*, Bulletin of Meikai Teacher-Training Support Center, vol.1, pp.1-15, 2018, in Japanese.
7. Chihiro Tachibana, *Game activity functions used in the teaching process of elementary school foreign language activities: Focusing on the functions of game activities and the effects of learning*, Himeji University Academic Institutional Repository, no.12, pp.9-22, 2019, in Japanese.
8. CHIeru Co, *Elementary school flash English words*, https://www.chieru.co.jp/lp_english/(accessed 2020-12-8), in Japanese.
9. TDS Co., LTD., *CORO-PUTER-Ja.*, https://www.replug.jp/coro-puter/(accessed 2020-12-8), in Japanese.

Using Various Evaluation Standards to Determine an Error Recovery Process in an Automation Plant

Akira Nakamura[*1], Natsuki Yamanobe[*1], Ixchel Ramirez Alpizar[*1], Kensuke Harada[*1, *2] and Yukiyasu Domae[*1]

*[*1]Industrial Cyber Phsical System Research Center*
National Institute of Advanced Industrial Science and Technology (AIST)
Second Annex, AIST Tokyo Waterfront, 2-4-7 Aomi, Koto-ku, Tokyo 135-0064 Japan

*[*2]Robotic Manipulation Research Group*
Systems Innovation Department
Graduate School of Engineering Science, Osaka University
1-3 Machikaneyama, Toyonaka 560-8531, Japan

E-mail: a-nakamura@aist.go.jp, n-yamanobe@aist.go.jp, ixchel-ramirezalpizar@aist.go.jp,
harada@sys.es.osaka-u.ac.jp, domae.yukiyasu@aist.go.jp
www.aist.go.jp

Abstract

In an automated plant, an error is more likely to occur in difficult tasks, which are complicated in nature. Such a task is often re-executed after returning to the previous step when a large-scale error occurs. Therefore, deciding both the past step that the task should return to and the recovery planning following this return becomes important. In this study, error recovery planning considering these two factors using various evaluation standards is realized.

Keywords: error recovery, task stratification, error classification, automation plant

1. Introduction

Automated plants have been used for production in various industries due to which the efficiency of production increases. On the contrary, system automation also leads to errors. Therefore, the progress of the error recovery method is crucial, which has demanded the increased study of error recovery in recent years.[1-5] However, the recovery technique used in these studies is ad hoc, and it is difficult to utilize them in various actual plants.

For several years, we performed a study on the systematization of the error recovery theory. We proposed a novel error recovery method based on the concepts of both task stratification and error classifications.[6-8] The main part of this method consists of fundamental elements with sequences of sensing, modeling, planning, and execution (Fig. 1). If an error occurs here, the process advances to the recovery part. In this part, the error cause is estimated, error is classified, system is corrected, and process is re-executed using the corrected system with an improved reliability.

Currently, deciding both the past step that the process should return to and the recovery planning after returning has become problematic. For this, we proposed a planning of the error recovery by deriving these two factors in consideration of cost.[8] In this study, we proposed a planning method for error recovery derived using various evaluation standards.

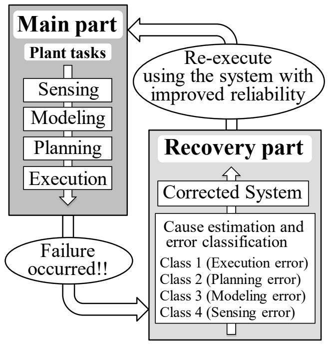
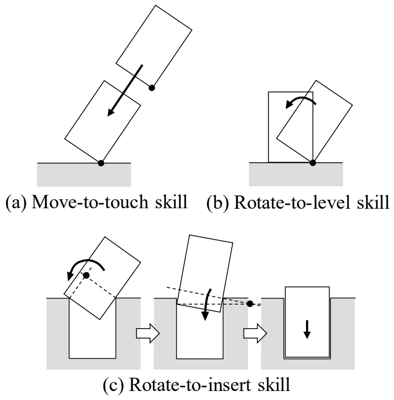


Fig. 1 Robot task system with an error recovery function



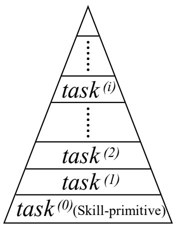


Fig. 2 Three fundamental skills

Fig. 3 Hierarchy of tasks

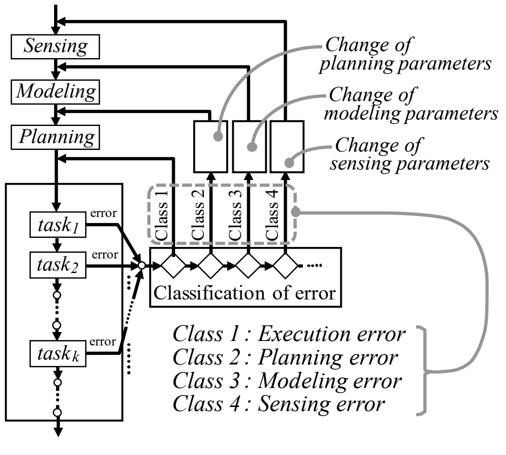


Fig. 4 Fundamental process flow with error recovery

The concept of skill, which is motion primitive, is described in Section 2. The error recovery technique is explained in Section 3. A method of recovery planning using plural evaluation standards is proposed in Section 4, and finally, a simple sample is presented in Section 5.

2. Concept of Skill

In this study, it is assumed that "*skill*" represents the unit of motion. In this section, we describe the indispensable aspects of skills, which are components of human behavior and automated plant operations.[9-11]

2.1. *Skill primitives*

We derived the motion primitives constituting tasks such as assembly and carrying by analyzing a person's behavior and called this motion unit "*skill*".[9-11] Thus, three basic skills of "*move-to-touch*", "*rotate-to-level*" and "*rotate-to-insert*" are important (Fig. 2). The representative person's behavior can be composed of these three basic skills and other resembled skills. The unit of machine motion can be considered to be similar to the motion primitives of human behavior.

2.2. *Stratification of tasks*

In general, it is possible to regard a task performed by an equipment in an automated plant as a hierarchical structure (Fig. 3).[9-11] The layer "*task*$^{(i+1)}$" occurs one tier above the layer "*task*$^{(i)}$", and the layer "*skill primitive*" is represented by the lowest layer "*task*$^{(0)}$."

3. Error Recovery

In an actual environment unlike the ideal, errors in equipment performance often occur due to various causes. In this section, we describe the concept of error classification and our proposed technique of error recovery.[6-8]

3.1. *Error classification*

Errors can be classified into four groups: execution, planning, modeling, and sensing errors on the basis of possible causes (Fig. 4).[6-8]

3.2. *Error recovery based on classification*

When an error occurs in automated plants, its cause is first estimated, and a suitable correction derived according to this tentative cause is performed at the equipment system. Then, the executive process returns to the previous step, and the task is restarted from the step (Fig. 4).[6-8] The same error has a lower probability of occurrence due to the corrected operation of the equipment.

If a small-scale error arises, the process returns to the previous step in the lowest layer of the hierarchy (Fig. 4, Fig. 5). If an error is large-scale, the process goes back to the previous step in the high-ranking hierarchy layer. In both cases, it is restarted from the previous step (Fig. 5).

3.3. *Candidate processes for recovery*

Let us consider candidates for error recovery processes. In the previous subsection, it was shown that the step in which the process returns to after the error occurs varies according to its scale. However, it is possible to back up bigger than a retreat step of necessity minimum.

Figure 6 shows various possible recovery processes when an error occurs at $subtask_m$ in task T with start S and goal G. Then, task T consists of a sequence of $subtask_k$ (k = 1, 2, 3, ⋯, n), which comprises a skill primitive sequence [$skill^{k}_{1}, skill^{k}_{2}, \cdots, skill^{k}_{n\_k(max)}$], where $subtask_m$ indicates the minimum traceable unit explained in the study,[6] that is, the smallest unit, wherein it is necessary to return to the first node of a skill primitive sequence if an error occurs. In the absence of any problem, there is a possibility that the execution of several skills will be continued within $subtask_m$ without returning to the previous step immediately after the error occurrence.

After the process returns to the j-th node before $subtask_j$ (j = 1, 2, 3, ⋯, m), the process [$subtask_j$ $subtask_{j+1}, \cdots, subtask_n$] proceeds along the center bold arrows in the corrected plant (Fig. 6). However, re-execution using the same sequence of subtasks is often impossible for various reasons such as the displacement of the target object and the change in the arrangement of objects. In such a case, an equivalent task of the original sequence, which is derived by large-scale re-planning of subtasks, is carried out as shown in the right-side arrows of Fig. 6.

4. Evaluation Standards for Selection of a Recovery Process

The previous section shows that there is a possibility that many recovery processes are generated. Therefore, it becomes important to restrict routes by deciding both the prior step that should be returned to and the recovery process after return.

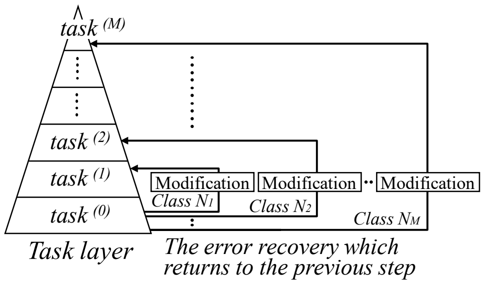


Fig. 5 The expression of task stratification and the process flow of the error recovery

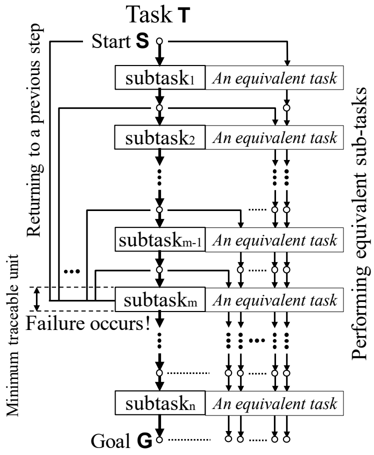


Fig. 6 Various processes of error recovery considered for a failure occurred in $subtask_m$

We proposed a suitable recovery planning method by considering the practical cost as an evaluation standard as in the study.[8] In this study, we proposed a method to derive the most suitable recovery process using various evaluation standards.

4.1. *Evaluation standards*

Let us consider the following eight evaluation standards to select a recovery process.

(i) Cost

We considered only the cost as the evaluation standard,[8] where costs include material costs, part costs, electricity bills, and planning expenses. A process with minimum practical costs is selected.

(ii) Time

Time is considered as the evaluation standard. The process where it is expected that the recovery sequence takes the shortest time is chosen.

(iii) Reliability

Reliability is considered as the evaluation standard. The process that emphasizes most on the operational reliability is chosen.

(iv) Safety

Safety is considered as the evaluation standard. A safety problem to a person and singular point problem are included in this standard. The process that ensures maximum safety of the operation is chosen.

(v) Finishing

The finishing of operation is considered as the evaluation standard. In the example of production and repair, a process whose outward appearance and feel are the cleanest is chosen.

(vi) Recovery data

Data on the recovery process are considered as the evaluation standard. A recovery process with a sufficient amount of data is chosen.

(vii) Tool

The tool used in the recovery process is considered as the evaluation standard. The recovery process with the special tool necessary for the practice of the work is chosen.

(viii) Operator skill

The skill of an operator in the recovery process is considered as the evaluation standard. When an expert is present, the necessary process of the skill is carried out with precedence.

4.2. *Simultaneous use of evaluation standards*

Eight types of evaluation standards were used as mentioned above, which are assumed to be used for the evaluation of the process independently. However, it is necessary to consider using several evaluation standards simultaneously considering the practical use. Of course, in this case, it is possible that the calculation of the evaluation function may become complicated.

5. Precedence of Recovery Processes Based on Various Evaluation Standards

In this section, the change in the priority of the recovery process based on one evaluation standard chosen among various standards discussed in Section 4 is observed via a simple sample exercise.

5.1. *Several types of recovery process*

Let us consider an assembly task in which a hook is to be stuck to the surface of a product using four precision screws, which resembles the exercise illustrated in the study.[8] Figure 7 shows a sequence of this task: the tacking task is performed in the first phase (Fig. 7(a)–(d)) so that these four screws are fastened temporarily to the plate part of the hook placed levelly. The erecting task is performed in the second phase (Fig. 7(e)–(f)) so that the plate part of hook with four screws stood vertically. The touching task is performed in the third phase (Fig. 7(g)–(h)) so that the hook is moved in the mounting location. The installation task is performed in the final phase (Fig. 7(i)–(m)) so that the hook is fixed on the product plane.

In this section, it is supposed that the drop of a screw occurring at the step of Fig. 7(h) represents the error as shown in Fig. 8, similar to the study.[8] Let us consider the following three types of recovery sequences chosen from a lot. Please refer to this study[8] for details.

- [ER-I] Error Recovery I

  The first sequence is a recovery process based on rerun from start S, using a new hook and four new screws, where the tacking task (Fig. 7(a)–(d)) is re-executed using new parts (Fig. 9).

- [ER-II] Error Recovery II

  The second sequence is a recovery process carried out by returning to the step of Fig. 7(d) in the tacking task, as shown in Fig. 10. Two types of methods exist for the use of a substitute of the dropped screw. One is the use of a new screw in the parts box, while the other is the reuse of the dropped screw by searching and picking up. The former is transcribed with [ER-II(N)] and the latter with [ER-II(F)].

- [ER-III] Error Recovery III

  The final sequence is a recovery process carried out by returning to a skill primitive in $subtask_m$ of the failure occurrence as shown in Fig. 11. A disadvantage of this method is that the task to fasten a screw temporarily to the hole in the plate part of the hook stuck perpendicularly is more difficult than when it is placed levelly. The method uses a new screw [ER-III(N)], and the method reuses the dropped screw [ER-III(F)].

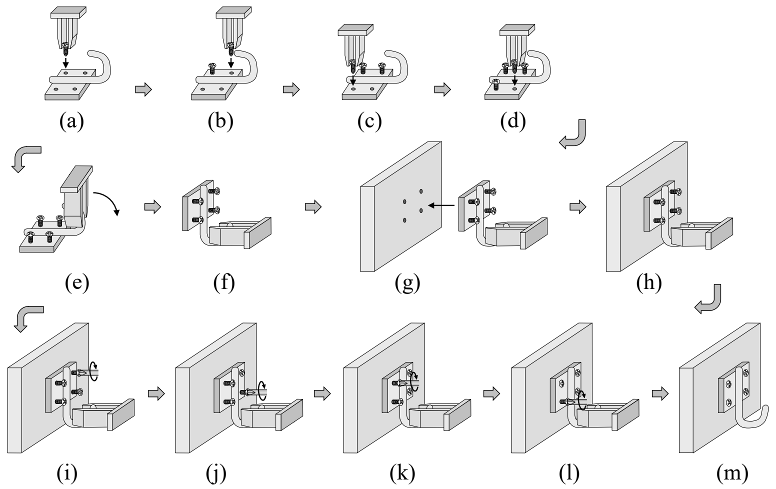


Fig. 7 An assembly task in which a hook is stuck to a plate by four precision screws

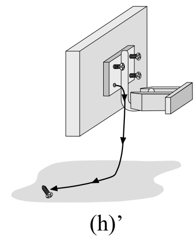


Fig. 8 An error in which a screw is dropped at (h) in Fig. 7

5.2. *Suitable process in each evaluation standard*

In the case of each evaluation standard shown in Section 4, let us consider a priority order of five recovery processes [ER-I], [ER-II(N)], [ER-II(F)], [ER-III(N)], and [ER-III(F)]. The explanation is carried out qualitatively, not quantitatively, for simplicity.

(i) Cost

Priority order: [ER-II(N)], [ER-III(N)], [ER-II(F)], [ER-III(F)], [ER-I].

When considering cost as the only evaluation standard, the selection method of a suitable process has already been explained.[8] The costs include material costs, part costs, electricity bills, and planning expenses. [ER-II(N)] without search operation and difficulty becomes the most suitable process, and [ER-I] where many parts are exchanged become the most unsuitable one.

(ii) Time

Priority order: [ER-III(N)], [ER-II(N)], [ER-I], [ER-III(F)], [ER-II(F)].

Here, [ER-III(N)] with a small return operation and without any search operation becomes the most suitable process, while [ER-II(F)] with return and search operation becomes the most unsuitable.

(iii) Reliability

Priority order: [ER-I], [ER-II(N)], [ER-III(N)], [ER-II(F)], [ER-III(F)].

[ER-I] with a restart operation becomes the most suitable process. [ER-II(N)] and [ER-III(N)] with planned operation and without search operation become the next suitable processes, whereas [ER-II(F)] and [ER-III(F)] with search operation become unsuitable processes. [ER-II(N)] and [ER-II(F)] without difficult operations are suitable for each order.

(iv) Safety

Priority order: [ER-I], [ER-II(N)], [ER-II(F)], [ER-III(N)], [ER-III(F)].

[ER-I] is the most suitable process. [ER-II(N)] and [ER-II(F)] without difficult operation and unstable posture become the next most suitable processes, whereas [ER-III(N)] and [ER-III(F)] with difficult operation becomes a process that is not suitable. [ER-II(N)] and [ER-III(N)] without a search operation are suitable for each order.

(v) Finishing

Priority order: Same for all.

There was no difference in finishing with respect to this task. However, the operation that may adversely affect a target item or the surroundings, such as scratch or scatter, should be avoided.

(vi) Recovery data

Priority order: A process with a large amount of data is given priority.

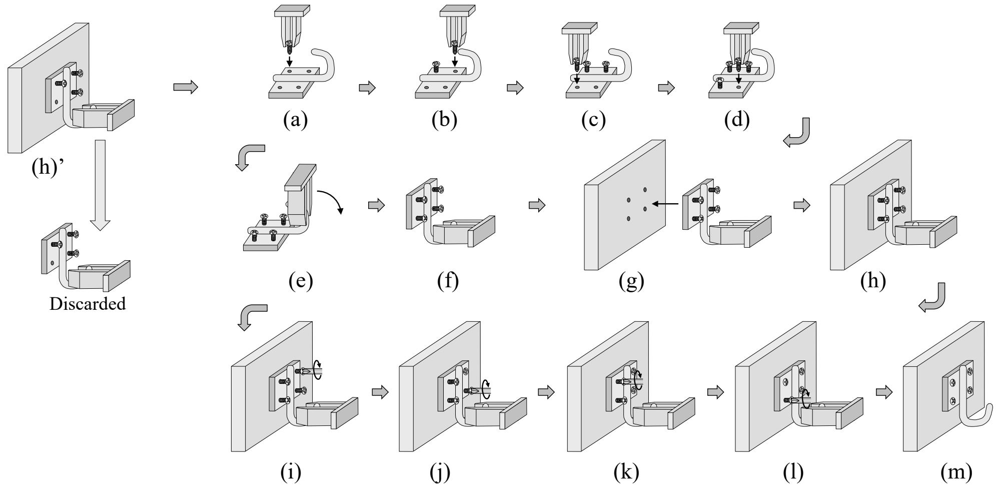


Fig. 9 [ER-I] Error Recovery I

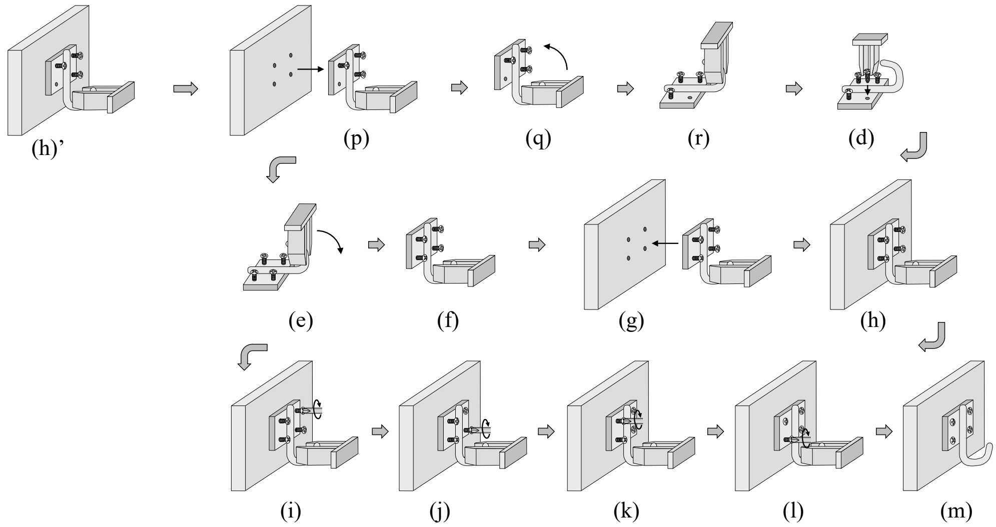


Fig. 10 [ER-II] Error Recovery II

The amount of data set is important when using a data science method. Even if the execution of operation is extremely difficult, a considerable amount of data enables the work to be performed skillfully.

(vii) Tool

Priority order: A process with an exclusive tool is given priority.

If a special tool for recovery is available, there is a high possibility that the process surely returns in a short time. If there is a tool by which four screws can be fastened temporarily as for any posture of a target, [ER-III(N)] (or [ER-III(F)]) becomes a top-priority process.

(viii) Operator skill

Priority order: A process of forte of teaching operator is given priority.

The recovery process in which the teaching operator is used is the most suitable. There is a possibility that a suitable process for an expert operator is different from that of an unskilled operator.

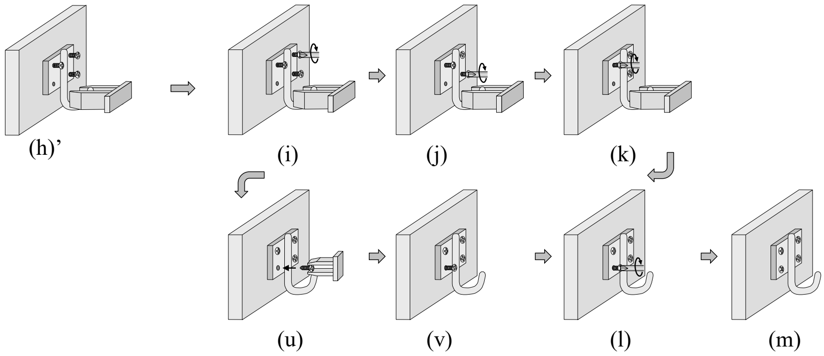


Fig. 11 [ER-III] Error Recovery III

We considered a suitable process derived by selecting one evaluation standard at a time from eight. However, in actuality, there is a possibility that the optimum process is derived using multiple evaluation standards.

6. Conclusion

When an error occurs in automated plants, the process advances to the recovery part. In general, many types of recovery processes can be considered. In this study, we have shown a method to derive a suitable process using only one evaluation standard at a time selected from various standards. The preferential order of recovery processes can be obtained considering each evaluation standard.

However, we have used only one standard, although we did consider eight types of evaluation standards. Therefore, in future studies, we will include techniques to derive a suitable recovery process using several evaluation standards simultaneously and determine the applicability of our method to actual systems.

Acknowledgements

This work was supported by New Energy and Industrial Technology Development Organization (NEDO) in Japan.

References

1. B. R. Donald, Planning multi-step error detection and recovery strategies, Int. J. Robot. Res., 9(1) (1990) 3-60.
2. T. Niemueller, G. Lakemeyer and S. S. Srinivasa, A Generic Robot Database and its Application in Fault Analysis and Performance Evaluation, in Proc. IEEE/RSJ Int. Conf. on Intell. Robots Syst., (Vilamoura, Portugal, 2012), 364-369.
3. E. D. Lello, M. Klotzbucher, T. D. Laet and H. Bruyninckx, Bayesian Time-Series Models for Continuous Fault Detection and Recognition in Industrial Robotics Tasks, in Proc. IEEE/RSJ Int. Conf. on Intell. Robots Syst., (Tokyo, Japan, 2013), 5827-5833.
4. E. Krabbe, E. Kristiansen, L. Hansen and D. Bourne, Autonomous Optimization of Fine Motions for Robotic Assembly, in Proc. IEEE Int. Conf. Robot. Autom., (Hong Kong, China, 2014), 4168-4175.
5. A. S. Wang and O. Kroemer, Learning Robust Manipulation Strategies with Multimodal State Transition Models and Recovery Heuristics, in Proc. IEEE Int. Conf. Robot. Autom., (Montreal, Canada, 2019), 1309-1315.
6. A. Nakamura, K. Nagata, K. Harada, N. Yamanobe, T. Tsuji, T. Foissotte and Y. Kawai, Error recovery using task stratification and error classification for manipulation robots in various fields, in Proc IEEE/RSJ Int. Conf. on Intell. Robots Syst., (Tokyo, Japan, 2013), 3535-3542.
7. A. Nakamura, K. Nagata, K. Harada and N. Yamanobe, Technique of Recovery Process and Application of AI in Error Recovery Using Task Stratification and Error Classification, J. Robotics, Networking and Artificial Life, 5(1), 2018, pp. 56-62.
8. A. Nakamura, N. Yamanobe, I. R. Alpizar, K. Harada and Y. Domae, Cost-oriented Planning for Error Recovery in an Automation Plant, J. Robotics, Networking and Artificial Life, 6(4), 2020, pp. 225-230.
9. T. Hasegawa, T. Suehiro and K, Takase, A robot system for unstructured environments based on an environment model and manipulation skills, in Proc. IEEE Int. Conf. Robot. Autom., (Sacramento, USA, 1991), 916-923.
10. A. Nakamura, T. Ogasawara, T. Suehiro and H. Tsukune, Fine motion strategy for skill-based manipulation, Artificial Life and Robotics, Springer, 1(3), 1997, pp. 147-150.
11. A. Nakamura, K. Kitagaki and T. Suehiro, Using simplified geometric models in skill-based manipulation, Advanced Robotics, 18(8), 2004, pp. 835-858.

Robotic Picking for Piled Sushi Topping

Kenta Matsuura

Graduate School of Engineering Science, Osaka University
1-3 Machikaneyama, Toyonaka 560-8531, Japan
Currently with Yaskawa Electric Co. Ltd., Japan

Keisuke Koyama, Weiwei Wan, Kensuke Harada

Graduate School of Engineering Science, Osaka University
1-3 Machikaneyama, Toyonaka 560-8531, Japan
E-mail: {koyama, wan, harada}@sys.es.osaka-u.ac.jp

Abstract

This paper proposes a method for picking piled sushi topping. By observing a human motion picking a sushi topping, we propose two picking strategies where one is to insert a finger into the separation among two toppings with shaking the finger, and the other is to insert a finger into the separation between a topping and a table. Along with two segmentation method, i.e., plane segmentation and LCCP segmentation methods, we experimentally verify the effectiveness of the proposed approach.

Keywords: Robot, Picking, Food Elastic Object.

1. Introduction

In recent years, labor shortage has become a serious problem in the food manufacturing industry, and the demand for automation of food manufacturing has been increasing. For example, taking food out of containers is a monotonous and heavy work for workers, so automation using robots is expected. However, food products are often flexible and irregularly shaped that do not retain a specific shape, such as fish fillets and noodles. This makes difficult to recognize their pose with vision sensors or to grasp them with a robot hand. In this study, we propose a recognition and grasping method focusing on the characteristics of flexible irregular objects, and grasp a sushi topping as an example (Fig.1).

In research on picking industrial parts without deformation, model-based methods[1], visibility-based methods[2], and machine learning-based methods[3] have been proposed. However, none of them assume flexible food as the object to be grasped. In contrast, in this study, we assume a situation where sushi toppings are stacked, and we consider grasping one of them. When a human grasps a sushi topping, he or she moves his or her fingers from side to side and puts his or her fingers in between the sushi seeds, and grasps one of them by

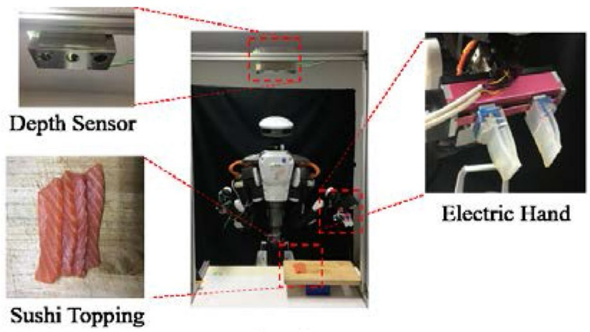


Figure 1 Overview of robotic picking for piled sushi toppings

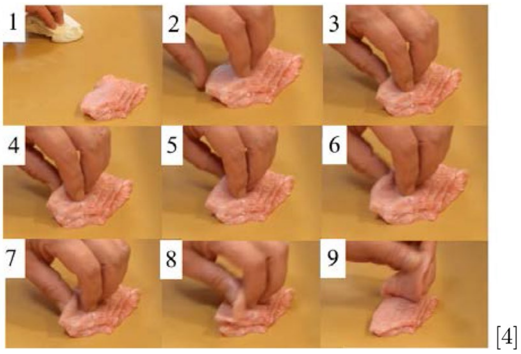


Figure 2 Human strategy for picking piled sushi topping

separating it (Fig. 2). This research examines the possibility of using a robot to perform this kind of grasping task for a flexible object. We consider two segmentation methods of point clouds and two grasping methods based on the human behavior. We evaluate the success rate of each segmentation method and each grasping method by conducting experiments using real sushi toppings.
Section 2 of this paper describes the proposed method, Section 3 describes the experiments and their results, and Section 4 summarizes and discusses future plans.

2. Proposed Method

In this study, we use a depth sensor to obtain a point cloud and segment it. To pick a sushi topping, we use a two-fingered parallel gripper where each finger is made of soft material. Section 2.1 describes the segmentation method for the point cloud of each sushi species and the calculation of the grasping pose, and Section 2.2 describes the method to grasp one sushi species based on the center of gravity of each segment.

2.1 Segmentation

We consider two segmentation methods of captured point cloud. The first method is to detect planes in a 3D point cloud dataset using the RANSAC algorithm and to segment the point cloud into planes [5] (Plane Model Segmentation). The second method is to divide similar voxels into groups of small convex regions, which are separated by concave boundaries[6] (LCCP Segmentation). In this study, after segmentation using Plane Model Segmentation[5] or LCCP Segmentation[6], the center of gravity of a cluster whose size is greater than a threshold is calculated for the purpose of determining the grasping pose. The density of the object was assumed to be uniform, and the center of gravity was calculated by dividing the sum of the x, y, and z coordinates of the points belonging to the cluster by the number of points in the cluster for each cluster. The grasped object is selected based on the obtained center of gravity.

2.2 Grasping method

In this study, we propose two grasping methods based on the human behavior of grasping sushi toppings. The first method is to insert the finger into the trough between two sushi toppings while shaking the finger to separate them (Fig. 3). The fingertip of the hand is first placed vertically above the separation between the two sushi pieces, and the hand is tilted slightly about the x-axis shown in Fig. 3, and then moves in the vertically downward direction until the hand hits the sushi piece. Then, the finger is inserted into the trough between the two sushi toppings with shaking the finger. While inserting the finger, the hand is gradually tilted back about the x-axis. The hand grasps the sushi topping after sliding the finger into the sushi sopping.
The second method is to place a finger between the sushi topping and the cutting board (Fig.4). In the coordinate system shown in Fig. 4, the fingertip is placed vertically above the point with the largest x-coordinate of the cluster of the largest z-coordinate of the center of gravity. Then, the hand is tilted about the y-axis where its rotation center is located at the fingertip. Then, the hand is moved vertically downward direction until it touches the cutting board. Then, the finger is inserted into the trough between a sushi topping and the cutting board. While inserting the finger, the hand is tilted back about the y-axis. The hand grasps the sushi topping after sliding the finger into the sushi sopping.

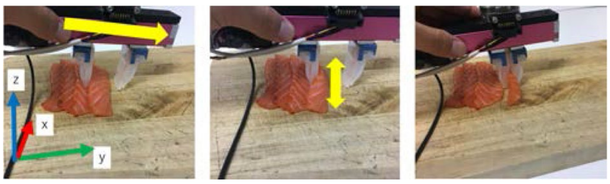


Figure 3 A finger is inserted into the separation between two sushi toppings (Picking strategy 1)

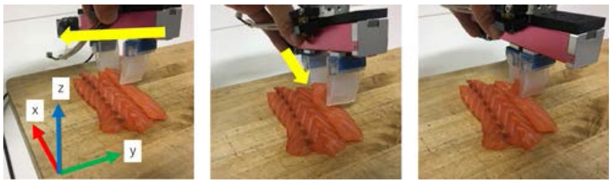


Figure 4 A finger is inserted between a sushi topping and the chopping board (Picking strategy 2)

3. Experiment

In our experiment, we use an industrial manipulator to grasp salmon and scallops, which positions and sizes are unknown as shown in Figure 5(a).

The robot grasps the object based on the information about the position and shape of the object calculated from the obtained point cloud. After acquiring the point cloud, we segment it by using the two methods described in Section 2.1. After segmentation, we calculate the center of gravity of the cluster with more than 100 points. The robot picks the object for 10 times by using two segmentation methods and two grasping methods.

The hand used in the experiment is an electric hand manufactured by THK Corporation where its fingers are made of silicon that allows the fingertips to maintain a moderate stiffness for grasping food (Fig. 5). The dimensions of a finger are 45 mm in length, 20 mm in width, 30 mm in opening and closing width, and 6 mm in thickness of the fingertip. We used salmon and scallops, and the scallops were opened and laid out on the table in front of the robot.

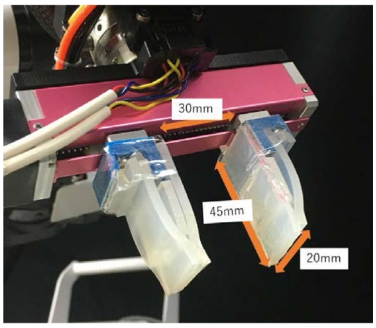
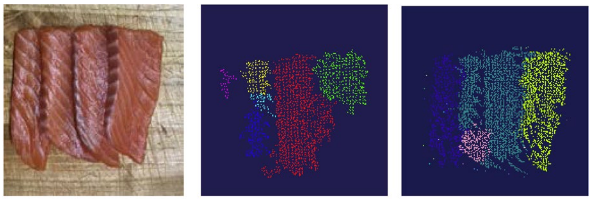


Figure 5 Electric hand used in our experiment

3.1. *Experimental Results and Discussions*

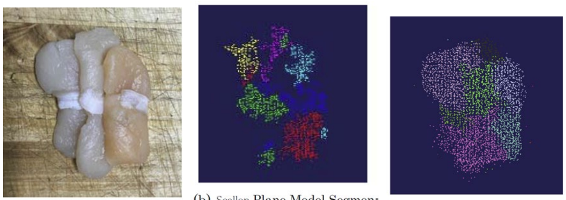
In Figs. 6 and 7, (a) shows the sushi toppings arranged on the cutting board, (b) shows the result of segmenting (a) into clusters by using the Plane Model Segmentation where points with the same color belong to the same cluster, and (c) shows the result of segmenting (a) into clusters by using the LCCP Segmentation. Fig. 8 shows the experimental result using the grasping method 1, and Fig. 9 shows the result of using the grasping method 2. The number of successful grasps out of 10 trials with two different segmentation methods is shown in Table 1.

In Plane Model Segmentation, as the distance between two bumps became small, it often becomes impossible to segment each sushi piece accurately. This is due to the unevenness and curvature of the sushi species since this method divides the point cloud into planes. Therefore, we could not successfully grasp the object when the separation between two pieces was 10 mm or 15 mm. On the other hand, LCCP Segmentation succeeded in segmenting the region with high accuracy. In the Plane Model Segmentation, we adjusted the thickness of each cluster. On the other hand, in LCCP Segmentation, we adjusted the angle at which the object was considered to be concave. The grasping method 1 was able to grasp all the sushi species at all distances between two pieces if they were divided into appropriate clusters. Even if there was almost no gap in the trough between two sushi species, it was sometimes possible to grasp them by inserting the fingers while moving it back and forth. Grasping method 2 was successful when the distance between two pieces was 15mm and 20mm. When the distance was 10mm, the finger interfered with the neighboring sushi topping and could not insert the finger between the sushi topping and the cutting board. In both grasping methods 1 and 2, the weight of the scallop was larger than that of the salmon, so the scallop was sometimes dropped even after grasping it once if it was grasped too shallowly.



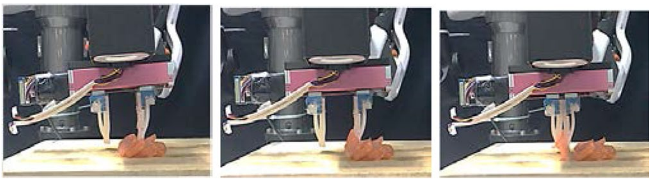
(a) Sushi toppings (b) Plane Model Segmentation (c) LCCP Segmentation

Figure 6 Result of segmentation (Salmon)



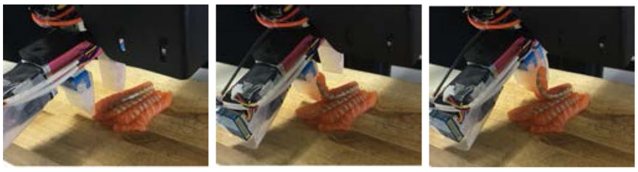
(a) Scallops (b) Scallop Plane Model Segmentation (c) Scallop LCCP Segmentation

Figure 7 Result of segmentation (Scallop)



(a) Initial position (b) Insert (c) Grasp

Figure 8 Result of picking strategy 1



(a) Initial position (b) Insert (c) Grasp

Figure 9 Result of picking strategy 2

Table 1 Number of successful picking

strategy	segmentation	salmon 10[mm]	salmon 15[mm]	salmon 20[mm]	scallop 15[mm]
1	PMS	0	0	5	0
1	LS	6	6	7	4
2	PMS	0	0	4	0
2	LS	0	6	5	4

PMS : Plane Model Segmentation LS : LCCP Segmentation

4. Conclusions

In this study, we proposed a unique grasping method for grasping flexible irregular objects by using two different methods to recognize objects from point cloud information and to find the center of gravity of clusters. From the experiments, we confirmed that when the objects are lined up with flexible irregular shapes, such as the sushi species used in this study, LCCP Segmentation can provide robust recognition, and the grasping method1 is relatively flexible regardless of the spacing of the objects.

In the future, we will develop the segmentation method used in this study to further improve the accuracy of segmentation and to estimate not only the center of gravity but also the approximate size and angle of the objects from the point cloud. In addition, we are investigating the construction of a system that can flexibly grasp various types and arrangements of objects instead of a fixed placement method.

References

1. K. Harada, W. Wan, T. Tsuji, K. Kikuchi,K. Nagata, and H. Onda. Iterative visual recognition for learning based randomized bin-picking. Experimental Robotics, pp. 646–655. Springer, 2016.
2. Y. Domae, H. Okuda, Y. Taguchi, K. Sumi, and T. Hirai. Fast graspability evaluation on single depth maps for bin picking with general grippers. IEEE Int. Conf. on Robotics and Automation (ICRA), pp. 1997–2004. 2014.
3. R. Matsumura, Y. Domae, W. Wan, and K. Harada, Learning Based Robotic Bin-picking for Potentially Tangled Objects, IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 7984-7991, 2019.
4. Sushi-master Toshi, https://youtu.be/QgcxLLZ7KJs, Jan., 2019.
5. A. Hast, J. Nysjo, and A. Marchetti. Optimal ransac-towards a repeatable algorithm for finding the optimal set. 2013.
6. S. Christoph Stein, M. Schoeler, J. Papon, and F. Worgotter. Object partitioning using local convexity, IEEE Conference on Computer Vision and Pattern Recognition, pp. 304–311, 2014.
7.

Motion Generation by Learning Relationship between Object Shapes and Human Motions

Tokuo Tsuji*[1], Sho Tajima[2], Yosuke Suzuki[1], Tetsuyou Watanabe[1], Shoko Miyauchi[3], Ken'ichi Morooka[3], Kensuke Harada[4], and Hiroaki Seki[1]

[1]Institute of Science and Engineering, Kanazawa University, Kakuma-machi, Kanazawa-shi, Ishikawa 920-1192, Japan.

[2]Graduate School of Natural Science and Technology, Kanazawa University, Kakuma-machi, Kanazawa-shi, Ishikawa, 920-1192, Japan.

[3]Faculty of Informatqion Science and Electrical Engineering, Kyushu University, 744, Motooka, Nishi-ku, Fukuoka-shi, Fukuoka 819-0395, Japan.

[4]Department of Systems Innovation, Graduate School of Engineering Science, Osaka University, Machikaneyama-chou, Toyonaka-shi , Osaka 560-8531, Japan

E-mail: {tokuo-tsuji, suzuki, twata, hseki}@se.kanazawa-u.ac.jp

Abstract

This paper presents a method for planning a robot motion of daily tasks by learning the relationship between object shapes and human motions. Robots are required to be able to deal with multifarious objects in various categories. However, it is difficult for robots to plan motions automatically for performing a task because objects even in the same category have different shapes. In our method, the motions are estimated by learning the relationship between object shapes and human motions using linear regression analysis.

Keywords: Motion Planning, Learning by demonstration, Machine Learning, Linear Regression Analysis, Robot hand

1. Introduction

Robots are expected to handle various objects in the field such as variable volume production, sorting of logistics, and daily tasks of life support. Since industrial objects and daily-life objects have different shapes even in the same category, robots have to adjust motions according to the difference in object shapes. For example, it is considered that the motion for pouring water with a bottle is generated by human teaching. Even though the motion can be used for the bottles which have the same shape, the motion may not be used for the bottles which have different shapes. The trajectory and execution time should be different according to object shapes. Since it is impossible to teach motions of every shapes manually, the method for planning motions automatically according to object shapes is needed. Most motion planners of previous works are used for pick-and-place task and assembly task. However, it is difficult to plan a complicated motion like pouring water by using them.

We propose a method for planning motions by learning the relationship between shapes and human

Fig. 1. Overview.

motions. We applied the proposed method to the task of pouring water from plastic bottles (Fig.1).

Our main contributions in this paper are as follows.

(1) Estimating shapes from measurement data using an RGB-D camera with one viewpoint

(2) Learning the relationship between human motions and object shapes for the motion generation in advance

(3) Acquisition object shapes and human motions by using an RGB-D camera and motion capture system

(4) Planning a motion automatically according to shape parameters acquired automatically

(5) Generating a motion for novel objects which has different shapes in the same category without a trial

Motions of pouring water from bottles into a cup are correlated with shapes of plastic bottles and cups. However, we assume the following things to verify the effectiveness of our proposed method. The object shapes except the manipulated object are the same. The initial water mass in the plastic bottles are known and constant.

2. Related works

In daily tasks, a task of folding cloth has been learned by robots based on working images [1]. Ochi et al. [2] proposed for scooping powder with a specific spoon by using deep learning. Rozo et al. [3] performed the pouring motion by learning time series data of joint angles and force sensor from pouring motion by a robot which human manipulated. Schenck et al. [4] poured a specific amount of liquids for different containers by using real-time visual feedback in a PID controller and determining the amount of liquid in a container from vision sensor during a pouring action. Lopez et al. [5] optimize pouring motion for novel containers as a receiver efficiently by learning the motion to mitigate spillage from the simulation of the pour. Kroemer et al. [6] learned from the information of object shapes and human demonstrations and generalize the pouring motion for novel containers by using a trial-and-error approach quickly. Tamosiunaite et al. [7] learned motion primitives by reinforcement learning. The pouring motion was performed efficiently because the learned motion was adapted to novel objects and the motion was relearned. Although the shapes of containers as a receiver were considered, the shapes of pouring containers were not considered in these researches or they needed some trials to perform pouring motions.

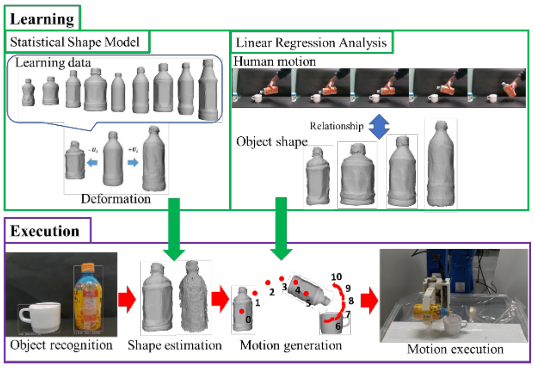


Fig. 2. Overview of motion generation.

It is necessary to handle the different shapes of objects so that motions can be planned with consideration for tasks, objects, and environments simultaneously. However, the previous researches uses simple rotation for pouring motions and is not adapted to arbitrary trajectory which depends on the shape of bottles. In this study, we propose a motion trajectory generation to handle different shapes of manipulated objects.

3. Overview of motion generation

A flow chart of the overview is shown in Fig.2. The proposed method consists of two phases: learning and execution.

In the learning phase, two types of learning are performed. The first type constructs a statistical shape model (SSM) [8] from learning data collected from objects in the same category with different shapes. The second type derives relational equations between shapes and time series of human motions via linear regression analysis. The learning phase is completed prior to the execution phase.

In the execution phase, the shape of an object is estimated using an SSM. Motions are estimated by the linear regression equations based on the parameters of the shape obtained by the SSM.

4. Modeling and estimation of shapes by the SSM

SSMs of main objects are prepared for each category in the learning phase. In the execution phase, after recognizing the category of an object, its shape is estimated by using an RGB-D camera and expressed by various parameters. Shape estimation is performed by

using point clouds from depth images and contours from RGB images. An SSM describes an average shape and variations of multiple shapes that vary between different objects. A three-dimensional point cloud is resampled by mapping it onto an arbitrary target surface [8].
The shape vector $X_i (i = 1, \dots, n)$ of an SSM is represented as follows:

$$X_i = [x_{i0}, y_{i0}, z_{i0}, \dots, x_{im}, y_{im}, z_{im}] \quad (1)$$

where the number of points is m and the number of shape models is n. A deformed shape vector X_i for the SSM is obtained by summing the calculated average shape vector $\bar{X}$ and vectors U_i as follows:

$$x = \bar{X} + \Sigma_{i=1}^{n} U_i b_i \quad (2)$$

where the values of the shape parameters $\beta = [b_1, b_2, \dots, b_n]$ represent the weight vector for each vector. The deformed shapes for each category are expressed by changing β.

Main objects are measured from one viewpoint by an RGB-D camera for shape estimation. A point cloud of the surface of a main object is extracted from a depth image. The contours of a main object are extracted from an RGB image. The iterative closest point algorithm (ICP) is used for registration between the point cloud of a depth image and that of an SSM. 50 SSMs were generated via random deformation for finding the parameters which have the minimum error. This procedure repeats until the shape parameters β are constant for ten iterations. Some examples of shape estimation via SSM are presented in Fig. 3.

5. Motion generation

The shape of a main object is represented by shape parameters β. The relationships between the shape parameters β and a human motion are learned and modeled by using linear regression analysis during the learning phase. The motion of a main object is estimated by using linear regression analysis during the execution phase.

5.1. *Learning data acquisition*

The speeds of human motions are different in each iteration. Therefore, we divide a motion into several smaller motions at various key frames and acquire learning data for an execution time t. The divided motions are defined as motion1, motion2, ..., and motionN_t.

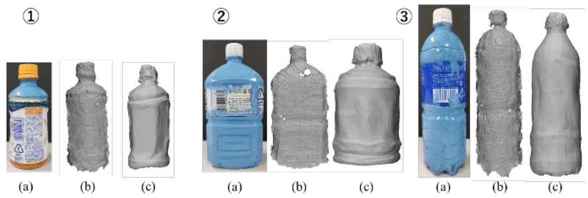


Fig. 3. Shape estimation (a) RGB Image, (b) Measurement point cloud, (c) SSM.

5.2. *Linear regression*

Linear regression is applied to the data for the shape parameters and motions and a regression coefficient matrix is calculated. Let n be the number of learning models for the construction of an SSM.
Let N be the number of learning data for linear regression. Only motion in a two-dimensional plane is considered. The position in the horizontal direction is defined as y and the position in the vertical direction is defined as z. The rotation angle in the yz plane is defined as θ.

A linear equation is formulated from the shape parameters $B = [\beta_1, \beta_2, \dots \beta_N]^T$ and the position in the horizontal direction at a particular time $Y = [y_1, y_2, \dots, y_N]^T$. The normal equation for the linear equation $Y = BP$ is solved as follows:

$$P = (B^T B)^{-1} B^T Y \quad (3)$$

A numerical solution $(B^T B)^{-1}$ is derived via singular value decomposition. The posture and the position in the vertical direction at a particular time are also calculated. The regression coefficient matrices for the execution time are obtained in the same manner for motion1, motion2, ..., and motionN_t. The products of the shape parameters β from novel objects and regression coefficient matrices P are calculated and the relative position and posture of the main object for each section (time series), as well as the execution time t for each motion are estimated.

6. Evaluation

We evaluated the proposed method for pouring water into a cup from a plastic bottle. The generated motions were evaluated experimentally using an actual robot. Shape models of nine plastic bottles were used for constructing an SSM with n=9. Ten real plastic bottles that were different from the nine shape models for the SSM were used in our experiments. Their point cloud data were measured and shape parameters β were estimated 10 times for each bottle. Human motions of pouring water into the cup were performed 10 times for each bottle. The initial water mass in each bottle was 0.2 kg and 0.1 kg of water was poured into the cup. The number of the dataset is 100 sets in total.

The data from nine out of the ten plastic bottles were used as learning data (N=90) and the data from the remaining bottle were used as testing data. Robot motions were generated by estimating the positions and postures of main objects for each section, as well as the execution times for each motion based on shape estimation of the testing data. We used five sequences per object and performed 50 experiments using an actual robot.

In these experiments, the motions generated based on the prepared dataset were verified by using an actual robot. The experimental conditions were as follows. The initial position and posture of the cup and plastic bottles were determined manually. The grasping points of the plastic bottles were also determined manually. The initial water mass in the plastic bottles was 0.2 kg. The experiment is shown in Fig. 4. The success rates of avoiding spilling water is 92% (46/50).

7. Conclusions

We proposed a method for generating motions by estimating the positions and postures of main objects and the execution time for each motion based on shape estimation. We performed position and posture estimations and obtained results very close to the variation of positions and postures in human motions. In experiments using an actual robot, the success rate for pouring water into a cup was 92% in terms of avoiding spilling water. In the future, we plan to generate motions based on shape changes in target objects, the initial water masses in plastic bottles for generalization of the proposed method.

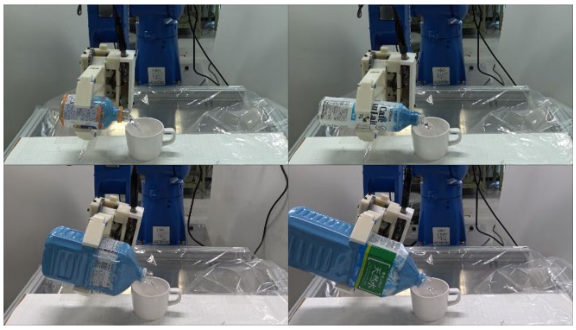


Fig. 4. Experiment.

References

1. P.-C. Yang, K. Sasaki, K. Kase, S. Sugano and T. Ogata, Repetable Folding Task by Humanoid Robot Worker using Deep Learning, IEEE Robotics and Automation Letters(RA-L), pp. 397-403, 2017.
2. H. Ochi, W. Wan, Y. Yang, N. Yamanobe, J. Pan and K. Harada, Deep Learning Scooping Motion using Bilateral Teleoperations, Proc. of IEEE International Conference on Advanced Robotics and Mechatronics, 188, 2018.
3. L. Rozo, P. Jim´enez, and C. Torras, Force-based robot learning of pouring skills using parametric hidden Markov models, Robot Motion and Control, 2013 9th Workshop on. IEEE, 2013, pp. 227-232.
4. C. Schenck and D. Fox,Visual closed-loop control for pouring liquids.,In 2017 IEEE International Conference on Robotics and Automation, pp. 2629-2636. IEEE, 2017.
5. T. Lopez-Guevara, N. K. Taylor, M. U. Gutmann, S.Ramamoorthy, and K. Subr,Adaptable Pouring, Teaching Robots Not to Spill using Fast but Approximate Fluid Simulation, Proceedings of the 1st Annual Conference on Robot Learning, vol.78, pp. 77-86, 13-15 Nov 2017.
6. O. Kroemer, E. Ugur, E. Oztop, and J. Peters, A kernel-based approach to direct action perception, in 2012 IEEE International Conference on Robotics and Automation, may 2012, pp. 2605-2610.
7. M. Tamosiunaite, B.Nemec, A. Ude, and F. Worgotter, Learning to pour with a robot arm combining goal and shape learning for dynamic movement primitives, Robotics and Autonomous Systems, vol. 59, no. 11, pp. 910-922, 2011.
8. S. Miyauchi, K. Morooka, Y. Miyagi, T. Fukuda, T. Tsuji and R. Kurazume, Tissue Surface Model Mapping onto Arbitrary Target Surface Based on Self-organizing Deformable Model, Workshop on Computer Vision at the 4th International Conference on Emerging Security Technologies，pp.79-82，2013.

Simulation and Regression Testing for Behavior of Software Models Based on Extended Place/Transition Net with Attributed Tokens

Tomohiko Takagi
Faculty of Engineering and Design, Kagawa University
2217-20 Hayashi-cho, Takamatsu-shi, Kagawa 761-0396, Japan

Ryo Kurozumi
Graduate School of Engineering, Kagawa University
2217-20 Hayashi-cho, Takamatsu-shi, Kagawa 761-0396, Japan
E-mail: takagi@eng.kagawa-u.ac.jp, s19g457@stu.kagawa-u.ac.jp

Abstract

We propose a technique of simulation and regression testing for Extended Place/transition Net with Attributed Tokens (EPNAT) models, and then show an early prototype tool to partially support it. In the technique, the information about a current marking (a current distribution of attributed tokens, including current values of attributes), current values of global variables, and current fireable transitions is indicated for the simulation, and also the good execution traces in the simulation are recorded as test cases for the regression testing. When an EPNAT model is modified, the test cases can be applied to it in order to reveal regression failures.

Keywords: software modeling, place/transition net, VDM, simulation, regression testing

1. Introduction

Formal specifications that represent abstracted software requirements in unambiguous and executable forms play an important role in the development of high-quality software. Extended Place/transition Net with Attributed Tokens (EPNAT)[1] is a formal specification description language for modeling the expected behavior of state transition-based software that consists of multiple objects, such as modules and subsystems. In an EPNAT model, each attributed token corresponds to an object and has variables to characterize the object, which are called attributes. The firing of transitions leads to the increase, decrease, and move of attributed tokens, the change of values of attributes and global variables, and so on. The behavior is constrained by invariants, pre-conditions, post-conditions, and type constraints. Engineers need to understand such complex aspects of the EPNAT model when constructing, validating and refining it.

In order to address this problem, we propose a technique of simulation and regression testing for EPNAT models, and then show an early prototype tool to partially support it. In the technique, the information about a current marking (that is, a current distribution of attributed tokens, including current values of attributes), current values of global variables, and current fireable transitions is indicated for the simulation, and also the good execution traces in the simulation are recorded as test cases for the regression testing. When an EPNAT model is modified, the test cases can be applied to it in order to reveal regression failures.

This paper is organized as follows. Section 2 describes our technique of simulation and regression testing for EPNAT models, and then section 3 illustrates our early prototype tool. Finally, we show discussion and future work in section 4.

2. Technique Overview

In this section, we propose the technique of simulation and regression testing for EPNAT models.

2.1. *Simulation of EPNAT models*

The simulation is intended to help engineers to construct, validate, refine, and understand EPNAT models, and it is applicable to both completed EPNAT models and EPNAT models under construction. The simulation of a given EPNAT model is performed according to the following procedure. The overview of the procedure is shown in Fig. 1. Step 2, 3, 5, and 7 can be automatically executed, and should be supported by a tool.

Step 1. An engineer specifies a starting state (that is, a state to start the simulation) according to the aim of the simulation. In this study, a state is characterized by a marking and values of all the global variables in an EPNAT model. Note that a marking in an EPNAT model includes values of attributes. The starting state does not necessarily need to be an initial state (that is, a state established just after software is invoked) or reachable state (that is, a state that can be reached from the initial state).

Step 2. If the starting state violates any type constraints and invariants, the procedure returns to Step 1. Otherwise, the current state of the given EPNAT model is initialized to the starting state.

Step 3. The pre-conditions of all the transitions are evaluated in order to determine fireable transitions in the current state. If there are no fireable transitions, the simulation is terminated.

Step 4. The engineer confirms the determined fireable transitions and acceptable values of their arguments. If he/she finds any faults, the simulation will be stopped. Otherwise, the engineer selects one arbitrary fireable transition and specifies values of its arguments according to the aim of the simulation.

Step 5. If the specified values of the arguments violate any type constraints and invariants, the procedure returns to Step 4. Otherwise, the selected transition is fired, and then its actions are executed by using the specified values of arguments. As a result of them, the current state is changed. If any invariants, and the post-condition of the fired transition are violated due to some sort of failures included in the given EPNAT model, the simulation will be stopped.

Step 6. The engineer confirms the current state. If he/she finds any faults, or if he/she achieves the aim of the simulation, the simulation will be terminated.

Step 7. This procedure returns to Step 3.

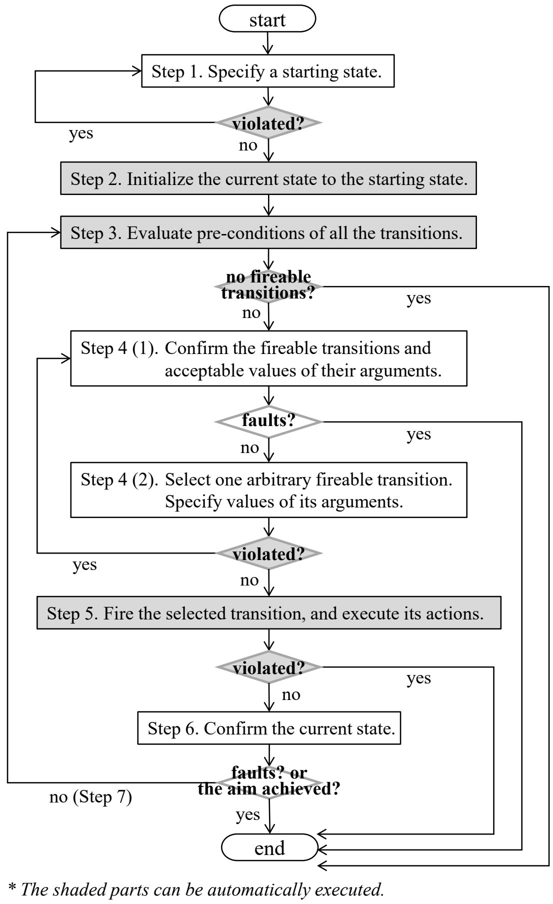


Fig. 1. Procedure of simulation (overview).

2.2. *Regression Testing of EPNAT models*

A good execution trace in the simulation is recorded as a test case for the regression testing. In this context, the word "good" means that (a) the execution trace does not include the occurrence of faults, (b) the execution trace is useful for the growth of test coverage, (c) the execution trace corresponds to the typical use of software, and (d) the execution trace is useful to cover fault-prone parts[2]. The item (a) is particularly essential for a test case.

```
<test-case> ::= <state> (<fireable-transitions> <state-transitions>)
<state-transitions> ::= <transition> <state> (<fireable-transitions> <state-transitions>)
<fireable-transitions> ::= <event> (<fireable-transitions>)
<transition> ::= <event> <values-of-arguments>
<state> ::= <marking> <values-of-global-variables> | <violation-state>
<marking> ::= <place> <attributed-tokens-in-the-place> (<marking>)
<attributed-tokens-in-the-place> ::= <attributed-token> (<attributed-tokens-in-the-place>)
<attributed-token> ::= <object-type> <values-of-attributes>
```

Fig. 2. Abstracted structure of a test case for an EPNAT model (written in BNF).

When an EPNAT model is modified, the test cases are applied to it in order to reveal regression failures. Some model-based coverage criteria[3,4] can be introduced to evaluate the effectiveness of the regression testing. Additionally, EPNAT models can be converted into VDM++ specifications[1,5], and thus common code-based coverage criteria[6] also can be introduced. In general, multiple test cases need to be executed to satisfy coverage criteria. Also, regression testing is repeatedly performed, and thus it should be automated by a tool in order to save time and effort.

Fig. 2 shows the abstracted structure of a test case for an EPNAT model. It is written in BNF, and parentheses are used to represent an optional element. A test case is a sequence of successive state transitions of arbitrary length. Each transition is identified by an event, and can have values of arguments. Each transition should follow fireable transitions that include its event. Fireable transitions are determined by a state, and thus should follow a state in a test case. Each state is characterized by a marking and values of all the global variables. If an invalid value is given to an argument of a transition in order to test invariants and a post-condition, the following state should be a violation state, that is, a state in which some invariant or the post-condition has been violated. A marking is expressed as a sequence of attributed tokens in each place, and an attributed token is characterized by object type and values of attributes. Each place can hold one or more attributed tokens, but the place and its attributed tokens need to have the same object type. Therefore, the object type of each attributed token is important information in regression testing of EPNAT models.

Transitions and a starting state correspond to test data (also called test input). On the other hand, fireable transitions and states excepting the starting state correspond to expected output. When there are no differences between expected output and test output through the execution of a given test case, it is concluded that the test case has successfully passed. Otherwise, an engineer needs to find and fix a regression failure, or update the test case so as to reflect the latest true software specification. After that, the engineer should perform confirmation testing, that is, apply the failed test case to the fixed EPNAT model, or apply the updated test case to the EPNAT model.

3. Early Prototype Tool

This section shows our early prototype tool to partially support the simulation and regression testing of EPNAT models. The tool includes two functions. One is an EPNAT editor to support the construction of EPNAT models, which has been developed in our previous study[1]. Another is an EPNAT simulator we are developing in this study, and it can be invoked from the EPNAT editor. The EPNAT simulator interacts with an existing tool called VDMJ[5] in order to execute a given EPNAT model.

Fig. 3 shows a screen shot of the EPNAT simulator that executes the simulation of the EPNAT model that represents the behavior of a simple load balancer[1]. The EPNAT model under simulation is visualized on the right pane of the EPNAT simulator. Fireable transitions in the current state are highlighted in green. Therefore, an engineer as a user of the EPNAT simulator will be able to easily confirm them, and select a next transition in order to proceed with his/her simulation. Also, the current state and the simplified execution trace (described as a sequence of fired transitions) are indicated on the left pane. When an engineer selects a fireable transition and specifies its values of arguments on the right pane, the EPNAT simulator automatically executes the firing of the selected transition, and then updates the current state, the execution trace, and the graphical image of the EPNAT model. An engineer can stop and reset the current simulation at any time. When an engineer thinks that the execution trace is good for a test case in regression testing, he/she can save it.

An engineer can start automated regression testing by loading a test case and an EPNAT model to be tested.

4. Discussion and Future Work

In this paper, we have proposed a technique of simulation and regression testing for EPNAT models, and then have shown an early prototype tool to partially support it. In the technique, the information about a current marking (including current values of attributes), current values of global variables, and current fireable transitions is indicated for the simulation, and also the good execution traces in the simulation are recorded as test cases for the regression testing. When an EPNAT model is modified, the test cases can be applied to it in order to reveal regression failures.

When a failure that was caused in the earlier stage of software development is found in the later stage (typically, the processes of system testing and acceptance testing), the cost to fix it generally tends to become higher. Therefore, it is important to find and fix failures in the earlier stage, and it is expected that our technique can be applied to address this problem. However, our tool is under development, and the functions to support our technique are not completely implemented at present. Also, the following matters need to be tackled in future work in order to improve the technique:

- An engineer will often need to confirm that the set of reachable states in an EPNAT model under construction includes all the indispensable states to satisfy given software requirements and also the set does not include any invalid states. However, test cases are manually created in our technique and tool, and thus it will be difficult for most engineers to do such task systematically at small cost. Model checking[7] may be useful to address this problem.
- After an engineer has made a modification on an EPNAT model, he/she will often need to maintain some existing test cases, that is, to update some existing test cases so as to reflect the latest true software specification. A technique and tool to systematically support it should be constructed.
- Coverage criteria are useful to create good execution traces in simulation and to select good test cases in regression testing. Engineers will need guidelines for the effective use of coverage criteria, and will need a tool to automatically suggest execution traces and test cases according to coverage criteria.

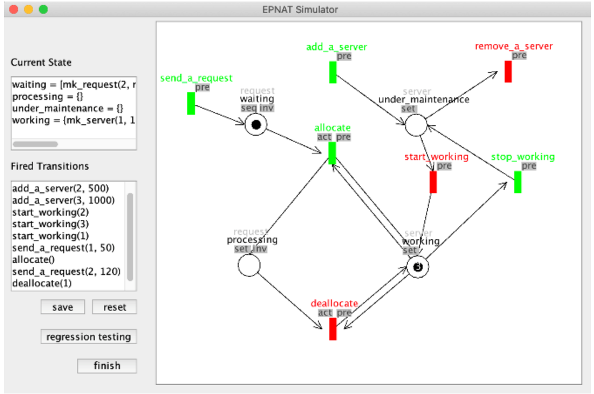


Fig. 3. Screen shot of an early prototype tool (EPNAT simulator).

Based on the above, we will develop the tool to support our extended technique, and apply it to pilot projects in order to evaluate its effectiveness.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP17K00103.

References

1. T. Takagi and R. Kurozumi, Software Modeling Technique and its Prototype Tool for Behavior of Multiple Objects Using Extended Place/Transition Nets with Attributed Tokens, *Journal of Robotics, Networking and Artificial Life*, Vol.7, No.3, pp.194-198, Dec. 2020.
2. J.A. Whittaker, J. Arbon and J. Carollo, *How Google Tests Software*, Addison-Wesley Professional, 2012.
3. T. Takagi, R. Kurozumi and T. Katayama, State Transition Tuple Coverage Criterion for Extended Place/Transition Net-Based Testing, *Proceedings of Pacific Rim International Symposium on Dependable Computing*, pp.29-30, Dec. 2019.
4. T. Takagi, N. Oyaizu and Z. Furukawa, Concurrent N-switch Coverage Criterion for Generating Test Cases from Place/Transition Nets, *Proceedings of 9th International Conference on Computer and Information Science*, pp.782-787, Aug. 2010.
5. J. Fitzgerald, P.G. Larsen, P. Mukherjee, N. Plat and M. Verhoef, *Validated Designs for Object-Oriented Systems*, Springer-Verlag London, 2005.
6. B. Beizer, *Software Testing Techniques*, Van Nostrand Reinhold, 2nd edition, 1990.
7. E.M. Clarke Jr., O. Grumberg and D. Peled, *Model Checking*, MIT Press, 1999.

Development of an Early Prototype Tool for Learning Software Modeling Using Extended Place/Transition Net

Tomohiko Takagi
Faculty of Engineering and Design, Kagawa University
2217-20 Hayashi-cho, Takamatsu-shi, Kagawa 761-0396, Japan

Akio Usuda
Faculty of Engineering, Kagawa University
2217-20 Hayashi-cho, Takamatsu-shi, Kagawa 761-0396, Japan
E-mail: takagi@eng.kagawa-u.ac.jp, s17t213@stu.kagawa-u.ac.jp

Abstract

This paper shows an early prototype tool for learning software modeling using Extended Place/transition Net (EPN), and then gives the discussion about its effectiveness and challenges. A user of the tool, that is, an engineer tries to construct his/her EPN model based on given software requirements by selecting and putting the given components of EPN. The EPN model is converted into a VDM++ specification for a user who is familiar with Vienna Development Method (VDM). Also, the behavior of software based on the EPN model is partially visualized by using animated graphics for a user who is a learner at the first stage. In the end, the correctness of the user's EPN model is automatically checked.

Keywords: software modeling, place/transition net, VDM, personal on-demand learning

1. Introduction

Extended Place/transition Net (EPN)[1] is Place/transition Net (PN) that includes some additional elements written in VDM++[2] to enhance its representation power, and can be used to formally model the state transition-based behavior of software in development processes. An EPN model, that is, a software model drawn by using EPN, can be executed on interpreters to understand and validate software specifications, and also can be converted to another formal software specification, source codes, and test cases. However, the use of EPN is based on technical knowledge and skills, and therefore engineers will need to learn them.

This paper shows an early prototype tool for learning software modeling using EPN. A user of the tool, that is, an engineer tries to construct his/her EPN model based on given software requirements by selecting and putting the given components of EPN. The EPN model is converted into a VDM++ specification[3] for a user who is familiar with Vienna Development Method (VDM). Also, the behavior of software based on the EPN model is partially visualized by using animated graphics[4,5] for a user who is a learner at the first stage. In the end, the correctness of the user's EPN model is automatically checked. This study is intended to support personal on-demand learning, and the key ideas of this study are (1) the construction using components, (2) the conversion into VDM++ specifications, and (3) the visualization using animated graphics.

This paper is organized as follows. Section 2 shows the steps to learn software modeling using EPN in this study. Section 3 illustrates the early prototype tool we are developing in this study, and then section 4 gives the discussion about its effectiveness and challenges.

2. Learning Software Modeling Using EPN

This section shows the three steps to learn software modeling using EPN in this study, that is, (1) creating exercises, (2) working on exercises, and (3) checking learner's answers. They have been developed based on our previous study[5].

2.1. *Creating exercises*

In the first step, instructors, that is, skilled engineers create exercises for learners. Each exercise consists of (i) software requirements, (ii) a completed EPN model, (iii) component candidates, (iv) animated graphics, and (v) hints to construct EPN models. They are generally created in this order.

2.1.1. *Software requirements*

The software requirements are written in natural languages. They should include enough information to construct a correct EPN model in the next step, such as the use case scenarios, state transitions, data processing, and constraints of the software.

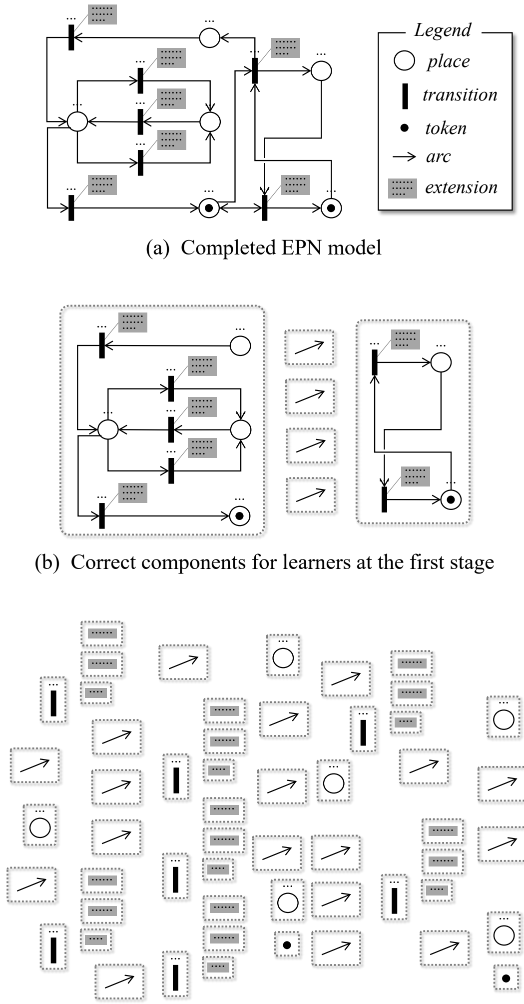
2.1.2. *Completed EPN model*

The completed EPN model is a correct answer in the exercise, and will be used to check learner's answers in the last step. It should be strictly based on the software requirements.

2.1.3. *Component candidates*

The component candidates that will be used by learners for constructing their EPN models in the next step are classified into the following two sets.

One is the set of correct components, and they are obtained by disassembling the completed EPN model, as shown in Fig. 1. When the exercise is intended for learners at the first stage, the size of each component may be made bigger in order to reduce the level of its difficulty. Another is the set of incorrect components, and they are created by mutating the correct components. Model-based mutation operators[6] can be applied to the elements of PN (that is, places, transitions, arcs, and tokens), and traditional mutation operators can be applied to the extension from PN to EPN (that is, actions and guards written in VDM++). When the exercise is intended for learners at the first stage, the set of incorrect components may be made smaller or empty in order to reduce the level of its difficulty.



(a) Completed EPN model

(b) Correct components for learners at the first stage

(c) Correct components for advanced learners

Fig. 1. Creation of the set of correct components (overview).

The instructors should confirm whether the component candidates lead to any other correct answers, that is, EPN models that do not have the same structure as the completed EPN model but satisfy the software requirements.

2.1.4. *Animated Graphics*

The animated graphics consist of several graphical parts, and visualize the behavior of software based on a given EPN model. Some of the graphical parts are programmed to move by trigger, such as the fire of specific transitions and the satisfaction of specific conditions in a given EPN model.

2.1.5. *Hints to construct EPN models*

The hints are written in natural languages, and will be used by learners as clues about how their EPN models can be correctly constructed. They are not always needed when the exercise is for advanced learners.

2.2. *Working on exercises*

In the second step, learners work on the exercises the instructors have created. The learners are given all the materials excepting the completed EPN model. They will firstly try to understand the given software requirements and hints to construct EPN models. After that, they will select arbitrary components from the given component candidates in order to add onto their EPN models. The learners' EPN models under construction are automatically converted into VDM++ specifications[3], and learners who are familiar with VDM will review the VDM++ specifications in order to confirm their EPN models from another viewpoint. Also, learners at the first stage will watch animated graphics in order to confirm their EPN models intuitively[4,5]. When learners finish constructing their EPN models, they move to the last step.

2.3. *Checking learner's answers*

In the third (that is, the last) step, an EPN model that has been constructed by a learner in the second step is compared to the completed EPN model that has been constructed by the instructors in the first step. If they are exactly the same, it is concluded that the learner has successfully constructed the correct EPN model. If they are not the same, the learner shall try to correct all the mistakes on his/her EPN model. If needed, the learner is given additional hints to construct the correct EPN model, such as the information about the mistaken parts in his/her EPN model and the unsatisfied items in the given software requirements.

3. Early Prototype Tool

In this study, we are developing an early prototype tool for learning software modeling using EPN. The tool does not fully support the steps that have been discussed in the previous section, but includes some essential functions.

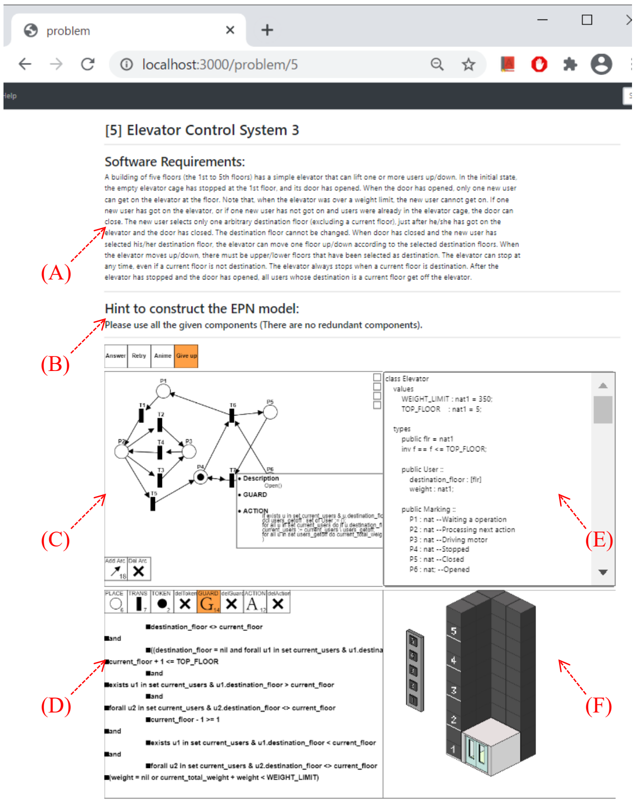


Fig. 2. Screen shot of our early prototype tool.

Fig. 2 is the screen shot of the tool that shows a sample exercise on the subject of a simple elevator control system. The tool is used on a Web browser. Its GUI chiefly consists of (A) the description of software requirements, (B) the description of hints to construct EPN models, (C) the pane to construct a learner's EPN model, (D) the pane to select component candidates, (E) the pane to show a VDM++ specification, and (F) the pane to show animated graphics.

In (D), component candidates are classified by the kinds of elements of EPN, that is, places, transitions, tokens, guards and actions. A learner can select an arbitrary one from (D), and can move it to (C). Components that have been moved from (D) can be returned to (D) by the learner. When a learner's EPN model is changed, its VDM++ specification shown in (E) is automatically updated in order to help a learner who is familiar with VDM to confirm his/her EPN model from another viewpoint. Also, in order to help a learner at the first stage to confirm his/her EPN model intuitively, (F) is intended to show the animated graphics of the elevator of which behavior is based on his/her EPN model. A learner can check the correctness of his/her EPN model, that is, can perform the automatic comparison between his/her EPN model and the correct EPN model at any time. The result of checking the correctness is indicated as "O" or "X" that are the symbols for a correct/incorrect answer, respectively. When a learner's EPN model is incorrect, the messages about mistakes may be given to the learner in order that he/she retries to the exercise. When the learner gives up constructing his/her EPN model, the completed EPN model is shown in (C).

4. Discussion

As shown in the previous sections, the key ideas of this study are (1) the construction using components, (2) the conversion into VDM++ specifications, and (3) the visualization using animated graphics. (1) will be useful to adjust the level of the difficulty of exercises. (2) will help learners who are familiar with VDM to confirm their EPN models from another viewpoint. (3) will help learners at the first stage to confirm their EPN models intuitively. However, through the development of the early prototype tool, we found the following challenges to be addressed in future:

- As with the method of Ref. 4, the creation of exercises, particularly the creation of animated graphics will require hard effort, and authoring tools should be introduced to support it.
- The animated graphics will not be so easy to quickly reveal learner's mistakes included in a large and complex EPN model. Optimized test case generation techniques may need to be introduced to solve it.
- Component candidates often lead to other correct answers. Instructors will require some techniques and tools to avoid the other correct answers or to confirm whether the set of their completed EPN models covers all the possible correct answers.

There are some closely related works. For example, a learning support technique for software modeling using PN was discussed in our previous study[5]. The steps discussed in section 2 have been developed based on it. However, unlike the previous study, this study is intended to support personal on-demand learning, and therefore does not include the steps of advising, review, and demonstration by instructors and other learners. Also, the previous study does not include the discussion about tools, and does not take EPN and VDM as objects of study. Ref. 4 shows a training support method and tool for bug fixing of EPN models. The characteristics of this previous study are to introduce animated graphics and to focus on bug fixing. The idea of the animated graphics is used also in this study.

5. Conclusion

In this paper, we showed an early prototype tool for learning software modeling using EPN, and then gave the discussion about its effectiveness and challenges. The key ideas of this study are (1) the construction using components, (2) the conversion into VDM++ specifications, and (3) the visualization using animated graphics. In a future study, we plan to develop the prototype tool and conduct preliminary experiments to improve our method and tool.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP17K00103.

References

1. T. Takagi, R. Kurozumi and T. Katayama, State Transition Tuple Coverage Criterion for Extended Place/Transition Net-Based Testing, *Proceedings of Pacific Rim International Symposium on Dependable Computing*, pp.29-30, Dec. 2019.
2. J. Fitzgerald, P.G. Larsen, P. Mukherjee, N. Plat and M. Verhoef, *Validated Designs for Object-Oriented Systems*, Springer-Verlag London, 2005.
3. T. Takagi and R. Kurozumi, Prototype of a Modeling Tool to Convert between Extended Place/Transition Nets and VDM++ Specifications, *Proceedings of International Conference on Artificial Life and Robotics*, pp.157-160, Jan. 2019.
4. T. Takagi, S. Morimoto, Y. Ue and Y. Imai, Animated Graphics-based Training Support Method and Prototype Tool for Bug Fixing of Extended Place/Transition Nets, *Journal of Robotics, Networking and Artificial Life*, Vol.5, No.4, pp.278-282, Mar. 2019.
5. Y. Ue and T. Takagi, Learning Support Technique of Software Visual Modeling Using Place/Transition Nets, *Proceedings of International Conference on Artificial Life and Robotics*, pp.751-754, Jan. 2020.
6. T. Takagi, R. Takata, Z. Furukawa, F. Belli and M. Beyazıt, Metrics for Model-Based Mutation Testing Based on Place/Transition Nets, *Proceedings of Joint Conference of International Workshop on Software Measurement and International Conference on Software Process and Product Measurement*, pp.7-10, Nov. 2011.

Test Suite Reusability Measurement Based on Uses Frequently and Coverage of Reused Test Cases

Mochamad Chandra Saputra[*], Tetsuro Katayama[*], Yoshihiro Kita[†], Hisaaki Yamaba[*], Kentaro Aburada[*], Naonobu Okazaki[*]

[*]*Interdisciplinary Graduate School of Agriculture and 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*

Email: chandra@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

Reusability of a test suite is one of important factors to decrease the cost in software testing. This research defines and measures the test suite reusability score expressed with the degree of reusability of the test suite. It is calculated by considering both frequency and code coverage of successful reused test cases in the test suite. Test suite reusability measurement provides useful information to improve the efficiency of software testing, especially in regression testing and automated testing.

Keywords: Test suite reusability, reused frequency, code coverage, reused test cases,

1. Introduction

White box testing knows as structural testing is one of the software testing approaches to test the program based on the internal structures of the program[1]. It is used to analyze internal logic and behavior, data structure, and code coverage. Test cases in the test suite is an important asset for software testing. The test cases are important on white box testing method to guarantee all independent paths or statements within a program executed at least once. Test cases should execute all true or false logical decisions and all loops at their or within the boundaries and ensure the validity of internal data structures[2].

Currently, reusing assets in the software development process have a great purpose. The background of reusing assets in software development is to reduce cost and effort[3]. A reusable test suite at the testing phase is decreasing testing time and cost. Generally, the test suite effectiveness is depended on the capability of the test cases on error detection related to coverage of the test cases on the program.

The research on test cases reusability measurement considers to code coverage is reported by modifying the program with reducing branches, the test suite is highly reusable but, modifying the program such as splitting, and loop, the test suite is less reusable[4]. The research has shown test suite reusability measurement is needed to consider the code coverage.

The use of code clones is increased day by day due to the growth of the use of open-source software and variants[5]. Code clones are efficient in reducing the cost and time on software development that have similar requirements. Testing code clones is needed a strategy to achieve efficiency on it. Test suite reusability is one of the best strategies to test the code clones.

This research proposes the formula for the test suite reusability score to measure the degree of test suite reusability based on reused test cases on another program. The formula is considered to test suite reusability frequency and distinct of code coverage. The test suite reusability score measurement is including good information to evaluate the efficiency of the test suite on reusability.

2. Principles of Reusability and Code Clones

The principle of software reusability is defined as the capability of an attribute to be reused in various objects[6]. The terminology of test suite reusability in this research is related to software reusability as the capability of test cases in the test suite to examine several or all paths of method should be tested on diverse objects. The test suite reusability additionally needs to consider code coverage on the reusability measurement. The case of test suite reusability measurement on white box testing, the capability of the test suite to examine different objects is not good enough. Consider the test suite can achieve 100% code coverage are one of criteria of a good test suite.

This research uses several types of code clone[7,8] such as code clone type 1 (exact clones) are identical clones except in white space and comments. The second type is code clone type 2, the differences from the original code are renamed identifiers, literals, types, layout, and comments but the structurally and syntactically is similar. The code clones type 3 are modified the statement such as statement insertions/deletions in addition to changes in identifiers, literals, types, and layouts. Code clone type 4 has been modified on code fragments to perform the same objective but different syntactic variants. The test suite reusability measurement is used on code clones because the test cases in the test suite need to use the same characters and by using the code clones we can have an identical or near identical program for uses of the test suite[9].

3. Test Suite Reusability Measurement

Reusing assets on software testing will help to reduce the cost of software testing. The asset is defined as any kind of product from any part of the software process[10]. The test suite is not very valuable if is not reusable in testing another program. The test suite reusability measurement is important to know how reusable the test suite.

The test suite reusability measurement is conducted by investigating the test case execution on the Java programs. First information used on this measurement is the number of successful reused test cases in the test suite. The successful test cases are the examination result of the test cases on Java program that could achieve the objective of the testing without any error. The next information is distinct code coverage. The test cases in the test suite are possible to execute similar lines of code.

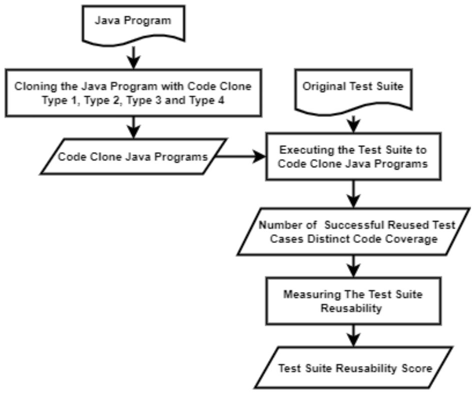


Figure 1. Test Suite Reusability Measurement Activity

The similar lines of code executed then count as one line of code from redundant lines of code or called distinct code coverage.

To simplify the formula, the research uses the following notation.

- SRTC: Successful reused test cases
- DCC: Distinct code coverage
- OT: Objects tested
- OLOC: Original line of code
- TC: Test Cases

By using the notation, the formula for test suite reusability score as follows.

$$TestSuiteReusabilityScore = \frac{\sum SRTC + \sum DCC}{(\sum OT \times \sum TC) + \sum OLOC} \quad (1)$$

The proposed formula for test suite reusability score has considering frequency and code coverage of successful reused test cases in the test suite.

4. Methodology

The test suite reusability measurement in this research is used on white box testing approach considered to the number of reused test cases in test suite to test the programs and code coverage. The test suite reusability measurement activity is shown in Figure 1. The activity of this experiment is using a Java program and given test suite to measure its reusability. Test suite reusability measurement based on code clones approach is used the number of successful reused test cases, and distinct code coverage from the test suite on calculation. The research uses code clones type 1, type 2, type 3, and type 4. Test suite reusability measurement results are informing the degree of test suite reusability.

5. Experiment

The research uses the *parallelogram* Java program as shown in Figure 2 with two given test suites. On the original *parallelogram* Java program, the test suites are

```
package parallelogram;
import java.io.BufferedReader;
import java.io.InputStreamReader;
public class Parallelogram {
  public static void main(String[] args) {
    double base,height;
    BufferedReader br = new BufferedReader(new
InputStreamReader(System.in));
    System.out.println("This program calculate value the
area of a trapezium");
    System.out.println("Enter the base and the height of
the parallelogram");
    try{
     base=Double.parseDouble(br.readLine());
     height=Double.parseDouble(br.readLine());
      }catch (Exception e){
        System.out.println("An error occurred");
        return;
        }
       if(base<=0 || height<=0){
        System.out.println("Wrong Input");
        return;}
      System.out.println("Area = " + base*height );
    }
}
```

Figure 2. Parallelogram Source Code

examining for collecting the coverage information. The test suite reusability measurement on this experiment is excluding the result of original test suite examination data. The experiment is focused on the test suite reusability score on another Java program that reuses the test suite to testing the program. The original Java program is cloning by code clone type 1, type 2, type 3, and type 4 approach.

This research has two given test suites. The first test suite contains three test cases and the second test suite contains four test cases are examined on code clones. The information of successful reused test cases and distinct code coverage from test cases examination on code clones are used for test suite reusability measurement.

6. Result and Discussion

The research is examined the given test suites on code clones type 1, type 2, type 3, and type 4 to collect information number of reused test cases in the test suite and code coverage of the test suite. Table 1 shows that the given test cases on the test suite successfully examine the code clones. The distinct code coverage on code clones shows that several lines of code on code clones type 1, 2, and 3 are not executed by test cases on the test suites during the testing.

The result of code coverage information shows in Table 1. The value 1 means this line is executed and 0 is not executed. Table 1 also shows the result of distinct code coverage in another column. The value 1 means this line is executed and 0 is not executed during the testing by using one or more test cases. The result of the number of successful reused test cases shows in Table 2. The result obtains from test cases examination on code clones. Table 3 shows the result of test suite reusability

Table 1. The result from code coverage information on code clones

	Code Clone Type-1								
	Test Suite-1				Test Suite-2				
No. LOC	TC-1	TC-2	TC-3	Disticnt Code Coverage	TC-1	TC-2	TC-3	TC-4	Disticnt Code Coverage
1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1
5	1	1	0	1	1	1	1	1	1
6	1	1	0	1	1	1	1	1	1
7	0	0	1	1	0	0	0	0	0
8	0	0	1	1	0	0	0	0	0
9	0	0	1	1	0	0	0	0	0
10	1	1	1	1	1	1	1	1	1
11	1	1	0	1	1	1	1	1	1
12	0	1	0	1	1	1	1	0	1
13	0	1	0	1	1	1	1	0	1
14	1	0	0	1	0	0	0	1	1
15	1	0	0	1	0	0	0	1	1

	Code Clone Type-2								
	Test Suite-1				Test Suite-2				
No. LOC	TC-1	TC-2	TC-3	Disticnt Code Coverage	TC-1	TC-2	TC-3	TC-4	Disticnt Code Coverage
1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1
5	1	1	0	1	1	1	1	1	1
6	1	1	0	1	1	1	1	1	1
7	0	0	1	1	0	0	0	0	0
8	0	0	1	1	0	0	0	0	0
9	0	0	1	1	0	0	0	0	0
10	1	1	1	1	1	1	1	1	1
11	1	1	0	1	1	1	1	1	1
12	0	1	0	1	1	1	1	0	1
13	0	1	0	1	1	1	1	0	1
14	1	0	0	1	0	0	0	1	1
15	1	0	0	1	0	0	0	1	1

	Code Clone Type-3								
	Test Suite-1				Test Suite-2				
No. LOC	TC-1	TC-2	TC-3	Disticnt Code Coverage	TC-1	TC-2	TC-3	TC-4	Disticnt Code Coverage
1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1
5	1	1	0	1	1	1	1	1	1
6	1	1	0	1	1	1	1	1	1
7	0	0	1	1	0	0	0	0	0
8	0	0	1	1	0	0	0	0	0
9	0	0	1	1	0	0	0	0	0
10	0	0	1	1	0	0	0	0	0
11	1	1	1	1	1	1	1	1	1
12	1	1	0	1	1	1	1	1	1
13	0	1	0	1	1	1	1	0	1
14	0	1	0	1	1	1	1	0	1
15	1	0	0	1	0	0	0	1	1
16	1	0	0	1	0	0	0	1	1

	Code Clone Type-4								
	Test Suite-1				Test Suite-2				
No. LOC	TC-1	TC-2	TC-3	Disticnt Code Coverage	TC-1	TC-2	TC-3	TC-4	Disticnt Code Coverage
1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1
3	1	1	0	1	1	1	1	1	1
4	1	1	0	1	1	1	1	1	1
5	1	1	0	1	1	1	1	1	1
6	1	1	0	1	1	1	1	1	1
7	1	1	0	1	1	1	1	1	1

measurement uses the formula(1). The test suite reusability measurement is very common information to reduce the cost of software testing. The different number of lines of code is possible for code clones type 3 and 4, for code clone type 1 and 2 usually the number lines of

Table 2. Result of the number of successful reused test cases

	Code Clones	Number of Test Cases	Number of Successful Test Cases Tested
Test Suite-1	Type-1	3	3
	Type-2		3
	Type-3		3
	Type-4		3
Test Suite-2	Type-1	4	4
	Type-2		4
	Type-3		4
	Type-4		4

Table 3. Result of test suite reusability measurement

	Code Clones	Test Suite Reusability Score	Average
Test Suite-1	Type-1	100%	100%
	Type-2	100%	
	Type-3	100%	
	Type-4	100%	
Test Suite-2	Type-1	84%	87%
	Type-2	84%	
	Type-3	80%	
	Type-4	100%	

code are the same. The distinct code coverage is used to selecting the redundant line of code executed by the test cases on the test suite. By using the distinct code coverage, the calculation of percentage code coverage is represented the actual condition. The test cases in the test suite are possible to execute similar lines of code. The number of successful reused test cases on the test suite as shown in Table 2 is important for test suite reusability measurement. The number of successfully reused test cases is one of the important parameters for the test suite reusability measurement.

The information of code coverages and number of successful reused test cases in the test suite is used for test suite reusability measurement. The result of test suite reusability measurement uses the formula (1) shown in Table 3. The result of test suite reusability score shows the different scores for several test suites because they have the different number distinct code coverage by the test cases as shown in Table 1. The number of distinct code coverage is important to ensure the capability of the test suite to achieve 100% code coverage. The score of test suite reusability 100% means the test suite has the perfect capability to achieve 100% code coverage and reused on another program. The test suite reusability measurement provides useful information to reduce the cost of software testing, especially in regression testing and automated testing.

7. Conclusion

This research confirms the test suite reusability measurement is calculated by combining the number of successful reused test cases and code coverage. The result shows the number of successful reused test cases in the test suite did not guarantee 100% code coverage on the test suite reusability measurement. The result of test suite reusability measurement is valuable information to reduce the cost of software testing, especially in regression testing and automated testing.

Our future works will focus on other measurement for test suite quality measurement.

References

1. Sommerville, I. *Software Engineering*. Addison-Wesley Publishing Company, 2010, pp. 133–170.
2. Roger S. Pressman, P. D. *Software Engineering: A Practitioner's Approach*. McGraw-Hill, 2009.
3. Sutcliffe, A. *The Domain Theory Patterns for Knowledge and Software Reuse*. Lawrence Erlbaum Associates, Inc.Publishers, 2002.
4. Dong, Y., Wang, Y., Lau, M. F. & Lin, S. *Experiments on Test Case Reuse of Test Coverage Criteria*. 7th International Conference on Ubiquitous Intelligence & Computing and 7th International Conference on Autonomic & Trusted Computing, 2010, pp. 277–281.
5. Saini, N., Singh, S. & Suman. *Code Clones: Detection and Management*. Procedia Computer. Science, 2018, pp. 718–727.
6. Younoussi, S. & Roudies, O. *All About Software Reusability: A Systematic Literature Review*. Journal of Theoretical and Applied Information. Technology, 2015, pp. 64–75.
7. Bellon, S., Koschke, R., Antoniol, G., Krinke, J. & Merlo, E. *Comparison and Evaluation of Clone Detection Tools*. IEEE Transactions on Software Engineering, 2007, pp. 577–591.
8. Roy, C. K. & Cordy, J. R. *Survey on Software Clone Detection Research*. in Technical Report No. 2007-541, 2007.
9. Asaithambi, S. P. R. & Jarzabek, S. *Towards Test Case Reuse: A Study of Redundancies in Android Platform Test Libraries*. in Lecture Notes in Computer Science vol. 7925, 2013, pp. 49–64.
10. Ezran, M., Morisio, M. & Tully, C. *Practical Software Reuse*. Springer London, 2002.

The Seven Information Features of Class for Blob and Feature Envy Smell Detection in a Class Diagram

Bayu Priyambadha*, Tetsuro Katayama*, Yoshihiro Kita †, Hisaaki Yamaba*, Kentaro Aburada*, Naonobu Okazaki*

**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: bayu@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

Measuring the quality of software design artifacts is difficult due to the limitation of information in the design phase. The class diagram is one of the design artifacts produced during the design phase. The syntactic and semantic information in the class is important to consider in the measurement process. The class-related information is used to detect the smell as an indicator of a lack of quality. All information related to the class is used by several classifiers to prove how informative it to be used to detect the smell. The smell types that are a concern in this research are Blob and Feature Envy. The experiment using three classifiers (j48, Multi-Layer Perceptron, and Naïve Bayes) confirms that the information can be used to detect Blob smell, on the other hand, Feature Envy, still needs more research. The average true positive rate of each classifier is about 80.67%.

Keywords: Smell Detection, Class Diagram Smell, Design Quality, Software Design

1. Introduction

Good quality of the software artifacts affects the final result of software products. It is important to measure the level of its quality. The measurement is not only to make a judgment about the best result of software product[1] but also for the process improvement in software development[2]. The metric quality is often used to measuring the quality of software artifacts.

The quality measurement is often done to the quality indicators of the software artifacts. For example, complexity, cohesion, and coupling[3]. Those three indicators are used to the quality measurement of the class. The software consists of a lot of classes as the template or blueprint of the objects. The object can work together with other objects to accomplished the specific functionality of the software. The inner class measurement means measure relationships between elements inner the class (cohesion). The cohesion is important to inform us about the compactness of the class. The level of compactness of the class is related to external quality attributes named maintainability and understandability. The more level of compactness of class then the class is easier to change and understand.

Nowadays, several researchers are extensively discussed and evaluate software design problems named smell, in the other term is “code smell”[4], or “design flaw”[5]. The existence of smell may affect the quality attributes of the software.

The metric and the existence of smells indicate the quality of software artifacts. The metric can be used as a tool to measure quality. And, the existence of smell indicates lack of quality. The relationship between software metrics and smells is an interesting thing to learn. Bigonha et. al., explain the usefulness of the software metric threshold for detection of bad smells and fault prediction[6]. All matrices used by Bigonha are the class level's metric. The measurement is done on the source code level.

The position of smell becomes a guide for the developer to improve the software artifacts. The smell can be overcome by refactoring activity. Fowler defines the term refactoring as the process of changing internal structure that is not changing the external behavior of software[4]. The code is modified over time, and the structure of the code will gradually disappear. The bad smell is emerging and the quality of code is decay as stated in the Lehman's Laws that said as the software is evolving then the structure is degraded[1]. Refactoring is the effort that we have to pay to manage the structure of code.

Mostly smell detection and refactoring process is done at the level of source code. There is the researcher that concern about the importance of shift the work to the design phase based on design artifacts. Start from the smell detection in design[7] then continue to the refactoring in design[8,9]. The challenge to do the smell detection on the design artifacts is the limitation of information. At the design level, there is only the model of the software that wants to build. One of the models as the product of the design is class diagram. Class diagram informs the structure of the software based on classes and their relationship. The possibility to collect the information from the class diagram is by understanding the syntactic and semantic information that may exist[10].

This research is aimed to collect the information based on the syntactic and semantic information that exists inner the class diagram. It uses the information to detect the existence of the smell in the class diagram. And, it uses the inner information of class to detect the smell. Two types of smell that are considered related to cohesion are Blob and Feature Envy. There is information that will be extracted from the class diagram and then combine with the smell dataset that is taken from the Landfill dataset[11]. That information will be classified using three classifiers, j48, Multi-Layer Perceptron (MLP), and Naïve Bayes, and the result will be compared. The focus of this research is how the usefulness of the information can be used to detect the smell.

The rest of the paper is organized as follows. In section 2, we present every data that will be used in this research and how to get those data. Section 3 describes how to label every data. Section 4 describes the whole process of

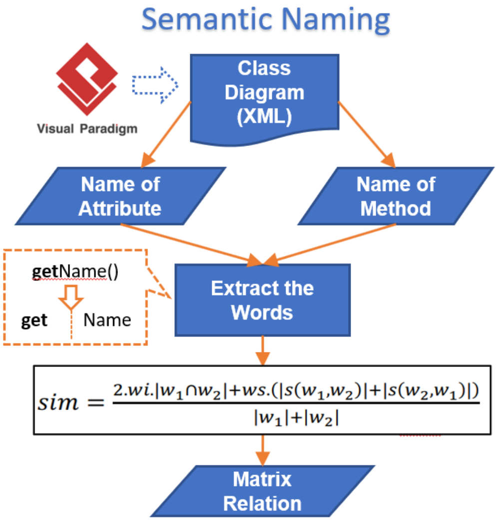


Fig. 1. Process of Semantic Similarity Analysis of the Label Name

classification. Section 5 describes the result and discussion. Then the last is the conclusion and future work in section 6.

2. Seven Information

There is seven information as the candidate data to use in smell detection. The data is related to one of the internal quality attributes called cohesion. The data are number of attributes, number of methods, number of the relation between method and attribute, number relation between method and method, number relation between attribute and attribute, the capacity of relation inner the class and cohesion.

In collecting the data it is important to consider not only the syntactic information but also the semantic information. Syntactic information is the information that we can directly read or extract from the class diagram. And, the semantic information is the information that needs a little processing (interpret the meaning) to extract. The object of study in this research is the class diagram that resulted by using the Visual Paradigm tools. Then, it converts to the XML file then called XML based class diagram (Fig. 1).

2.1. *Number of Attributes*

The first data is the attribute. The collection of the attributes in the class diagram is easy. The notation for the attributes name and the type of attribute is clearly described in the class diagram.

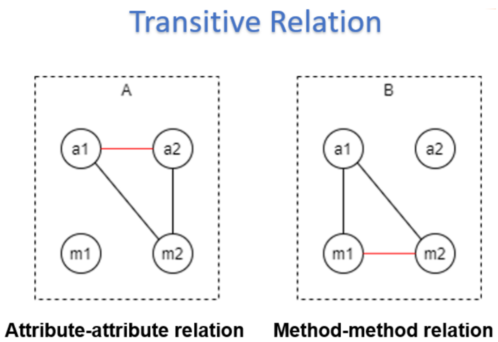


Fig. 2. The Description of Transitive Relation

The information about attributes is extracted from XML based class diagram.

2.2. *Number of Methods*

The second is the number of methods. The name, return value type and the parameters of the method are also clearly described in the class diagram. Collecting all information about the method is including in the syntactic information.

2.3. *Number of Relation between Method-Attribute*

The number of relations between Method and Attribute (MAR) can be count by considering not only syntactic but also semantic information. The relation between method and attribute is determined based on the similarity of type (syntactic) and the similarity of meaning (semantic) between method and attribute. Fig.1 describes the process to measure the closes meaning between the label name of method and attribute. All labels name are extracted from the XML's based class diagram. Then the labels are split into words. After that the similarity of meaning between the set of words is calculated[12]. If the label name between attribute and method is semantically similar then it will be considered as a relation. Matrix relation is generated to make it easier.

2.4. *Number of Relation between Method-Method and Attribute-Attribute*

Both the type of relation is called transitive relation or indirect relation. That relation is determined based on the direct relation between attribute and method. There are two definitions to determine method-method (MMR) and attribute-attribute relation (AAR). There will be a relation between methods if there are two methods that are related to the same attribute. Then, there will be a

Table 1. Representation of Dataset

The Seven Information Features

No.	Fields	Type
1.	MAR	Numeric
2.	MMR	Numeric
3.	AAR	Numeric
4.	Number of Attributes	Numeric
5.	Number of Methods	Numeric
6.	Capacity	Numeric
7.	Cohesion	Numeric
8.	Label	Blob, Feature Envy, No Smell

relation between attributes if there are two attributes that are related to the same method. This type of relation is included in the category of semantic information. Fig. 2 shows the analogy of both relations.

2.5. *Relation Capacity*

The maximum relation is the maximum number of relations that possible to exist in the class. The class is assumed as the area of the relation, that's has a capacity of relation. The maximum relation is calculated by using (1).

$$MaxRelation = \frac{(m+a)((m+a)-1)}{2} \tag{1}$$

Where m is the number of methods and a is the number of attributes.

2.6. *Cohesion Value*

The cohesion is calculated by considering MAR, MMR, and AAR. The relation is direct relation and transitive relation. All of the relations between method and attribute is divided by the maximum relation (1) that possibly exists in one class. The cohesion calculation is expressed as (2).

$$Cohesion = \frac{MAR+MMR+AAR}{MaxRelation} \tag{2}$$

3. Dataset

The Landfill dataset is consists of 243 instances of five types of code smells identified from 20 open-source software projects[11]. In the Landfill dataset, all classes that contain bad smell are labeled based on the type of smell.

In this research the dataset used to experiment is combine of the seven information and the label of smell provided by the Landfill dataset. The representation of the dataset is described in Table 1. The main dataset is extracted from six projects: jEdit, jHotDraw, FreeMind, HSQLDB, aTunes, and ArgoUML. The dataset consists of 300 data based on the smell class listed by the Landfill

Table 2. Recap of The Testing Result

No.	Classifier	Blob	Feature Envy	No Smell	True Positive Rate (TPR)
1.	J48	94.7%	25.5%	22.2%	51%
2.	MLP	94.7%	0%	38.9%	51%
3.	Naïve Bayes	52.6%	16.7%	55.6%	44.9%

dataset and the smell free classes, then used as training data. And, the other set of 49 data is generated for testing data.

4. Classification

In this experiment, the process of detecting the smell will use the classification method. This research uses three classifiers: j48, Multi-Layer Perceptron, and Naïve Bayes. The tool used in this experiment is Weka as a Machine learning software to solve data mining problems[13]. All classifiers are run by using a basic configuration. The use of a classifier is only the way to prove how the dataset can be distinctive to the bad smell (Blob and Feature Envy).

5. Result and Discussion

This section explains the result of the experiment by using the dataset and three classifiers. The recapitulation of the result is described in Table 2.

j48 has a TPR of about 51%. The correct classify for every label of data is, for Blob data is 94.7%, Feature Envy data is 25% and No data is 22.2%. MLP has a TPR value of about 51% of the data. Based on the label data, for Blob data is 94.7%, for Feature Envy data is 0% and for No data is 38.9%. Then, Naïve Bayes has a TPR value of about 44.9% where Blob data is 52.6%, Feature Envy data is 16.7% and No data is 55.6%.

Based on the result, two classifiers have a TPR value above 50%. The classifiers are j48 and MLP. But, the MLP classifier cannot identify the Feature Envy smell because the TPR for Feature Envy of MLP is 0%. The j48 and Naïve Bayes can detect the Feature Envy smell even though with the low rate. All classifiers can detect Blob smell with the TPR above 50%. The average TPR for Blob smell is 80.67%. The characteristics expressed in the dataset led to the identification of Blob smell. The existing information is not enough used to find both types of smells, especially for the Feature Envy smell. More detailed data is needed to improve the differentiation between the two types of smells.

6. Conclusion and Future Work

The use of seven information as a dataset of class to identify the Blob smell in the class diagram is very effective. The effectiveness is proven by the TPR above 50%. On the other hand, it is not very good to use to identify the Feature Envy smell. The dataset does not express the big differential for Blob and Feature Envy. It makes the classifier hard to identify each type of smell. Only j48 and Naïve Bayes can identify the Feature Envy but in the low of TPR.

To continue the research, it needs more exploration to increase the differentiation between both smells. The additional information of class maybe would be worth finding to increase the richness of data. The other, weighing every variable (MAR, MMR, AAR) is also considered important to find to sharpen the differences in the dataset.

References

1. Sommerville, I. *Software Engineering*. Software Engineering (Addison-Wesley Professional, 2010).
2. Chidamber, S. R. & Kemerer, C. F. *A Metrics Suite for Object Oriented Design*. IEEE Trans. Softw. Eng. **20**, 1994, pp. 476–493.
3. Chowdhury, I. & Zulkernine, M. *Using complexity, coupling, and cohesion metrics as early indicators of vulnerabilities*. J. Syst. Archit. **57**, 2011, pp. 294–313.
4. Fowler, M. *Refactoring Improving the Design of Existing Code*. (Pearson Education - Wesley, 2019).
5. Riel, A. J. *Object-oriented Design Heuristics*. (Addison-Wesley Publishing Company, 1996).
6. Bigonha, M. A. S. *et al. The usefulness of software metric thresholds for detection of bad smells and fault prediction*. Inf. Softw. Technol. **115**, 2019, pp. 79–92.
7. Sidhu, B. K., Singh, K. & Sharma, N. *A Catalogue of Model Smells and Refactoring Operations for Object-Oriented Software*. Proc. Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2018, 2018, pp. 313–319 doi:10.1109/ICICCT.2018.8473027.
8. Misbhauddin, M. & Alshayeb, M. *Model-driven refactoring approaches: A comparison criteria*. Proc. African Conf. Softw. Eng. Appl. Comput. ACSEAC 2012, 2012, pp. 34–39 doi:10.1109/ACSEAC.2012.20.
9. Dharmawan, T. & Rochimah, S. *Systematic literature review: Model refactoring*. Proc. 2017 4th Int. Conf. Comput. Appl. Inf. Process. Technol. CAIPT 2017. 2018, pp. 1–5.
10. Priyambadha, B. *et al. The Measurement of Class Cohesion using Semantic Approach*. Proc. Int. Conf. Artif. Life Robot. **25**, 2020, pp. 759–762.
11. Palomba, F. *et al. Landfill: An open dataset of code smells with public evaluation*. IEEE International Working Conference on Mining Software Repositories vols 2015-Augus. 2015, pp. 482–485.
12. Dijkman, R., Dumas, M., van Dongen, B., Käärik, R. & Mendling, J. *Similarity of business process models: Metrics and evaluation*. Inf. Syst. **36**, 2011, pp. 498–516.
13. Hall, M. *et al. The WEKA Data Mining Software : An Update*. SIGKDD Explor. **11**, 2009, pp. 10–18.

Improvement of RETUSS to Ensure Traceability between Sequence Diagram in UML and Java Source Code in Real Time

Kaoru Arima*, Tetsuro Katayama*, Yoshihiro Kita†,
Hisaaki Yamaba*, Kentaro Aburada*, 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
arima@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

Ensuring traceability of software deliverables is one of the methods to ensure software quality. RETUSS (Real-time Ensure Traceability between UML and Source-code System) is a tool that saves labor and time, and eliminates mistakes by human handling in ensuring traceability between UML and source codes. However, RETUSS is not useful due to its limited scope of application. This paper improves the usefulness of RETUSS by extending the scope of application of sequence diagram in UML and Java source code.

Keywords: software quality, traceability, UML, sequence diagram, Java

1. Introduction

The importance of software in society is increasing, and system failures and software bugs cause significant economic and social impact. Therefore, ensuring the quality of systems and software has become more important. Ensuring traceability of software deliverables is one of the methods to ensure software quality.[1] It can specify the scope of the impact due to the modification in the requirements and remove the gap between the documents and the source codes. However, it has the following two problems.

- Taking much labor and time to modify similarly other related deliverables in modifying a part of deliverables
- Having a risk that you cannot ensure traceability because of causing mistakes to ensure traceability by human handling

In order to solve them, our laboratory developed RETUSS (Real-time Ensure Traceability between UML and Source-code System).[2,3] RETUSS ensures traceability between UML (Unified Modeling Language)[4] and source codes by transforming them to each other in real time. Therefore, RETUSS can save labor and time, and eliminate mistakes by human handling in ensuring traceability between UML and source codes. RETUSS has the following functions.

- Description of class diagram
- Description of sequence diagram
- Description of Java source codes
- Description of C++ source codes
- Bidirectional transformation between class diagram and Java source codes
- Bidirectional transformation between class diagram and C++ source codes
- Bidirectional transformation between sequence diagram and Java source codes

However, RETUSS is not useful in ensuring traceability between sequence diagram and Java source codes due to its limited scope of application.

This paper improves the usefulness of RETUSS by extending the scope of application of sequence diagram and Java source codes.

2. The extended RETUSS

This paper extends the following two functions of the existing RETUSS.

- Description of sequence diagram
- Bidirectional transformation between sequence diagram and Java source codes

Fig. 1 shows the interface of the extended RETUSS. The extended RETUSS has the main window and the source code window. The main window is a window for describing UML. The source code window is a window for writing source codes.

Fig. 2 shows the structure of the extended RETUSS. The extended RETUSS consists of five parts: display part, correspondence part, description part, transformation part, and storage part. To extend the function of the description of sequence diagram, we mainly extend the display part, correspondence part, and description part. To extend the function of the bidirectional transformation between sequence diagram and Java source codes, we mainly extend the description part, transformation part, and storage part.

2.1. *Extending the function of the description of sequence diagram*

We add the following three functions to the function of the description of the sequence diagram.

- Adding a message
- Adding a combined fragment
- Deleting elements

By adding these functions, a user can edit the sequence diagram directly on RETUSS.

To add these three functions, we extend the display part, the correspondence part, and the description part. The extended display part displays three buttons in the main window: Message button, Combined Fragment button, and Delete button. The extended correspondence part has three new processes: displaying the dialog to add a message, displaying the dialog to add a combined fragment, and displaying the dialog to delete elements. The three processes are called from the event handlers of the three buttons displayed by the extended display part. The extended description part has four new processes: adding a message, adding a combined fragment, deleting a message, and deleting a combined fragment. These four processes are called by correspondence part and they change the sequence diagram information in the storage part.

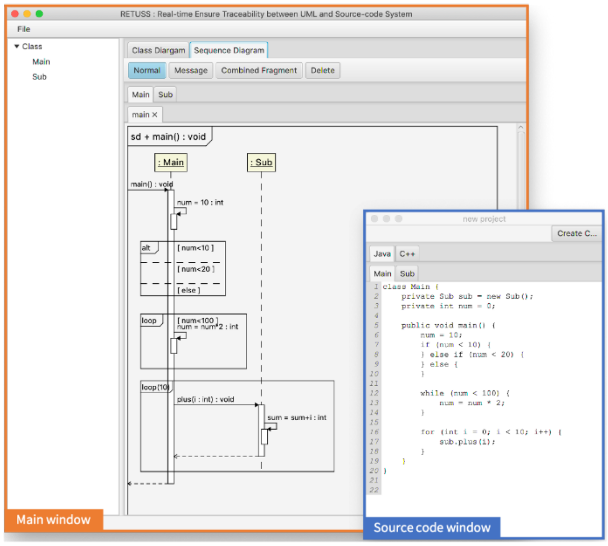


Fig. 1. The interface of the extended RETUSS

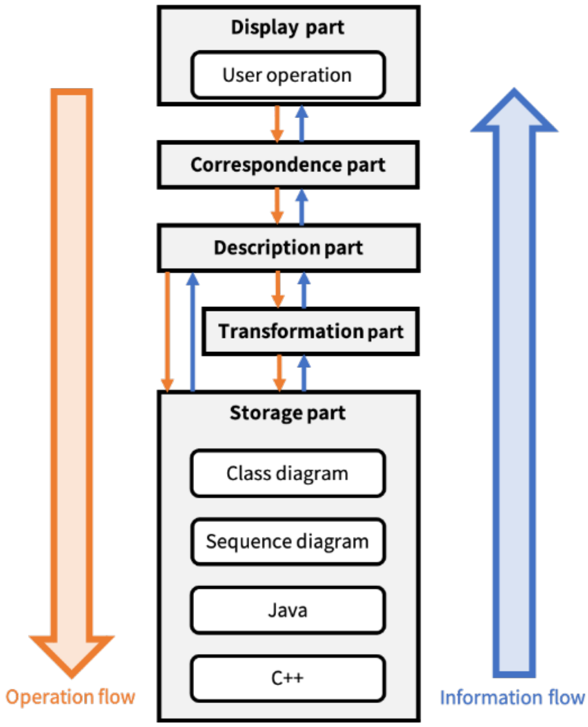
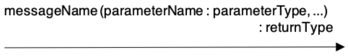


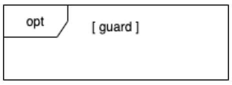
Fig. 2. The structure of the extended RETUSS

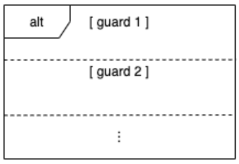
2.2. *Extending the function of the bidirectional transformation between sequence diagram and Java source codes*

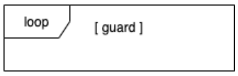
We define four new transformation rules to extend the scope of the bidirectional transformation between sequence diagram and Java source codes. Table 1 shows

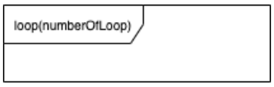
Table 1. The four new transformation rules for the bidirectional transformation between sequence diagram and Java source codes.

Name in sequence diagram	Notation in sequence diagram	Name in Java	Syntax in Java
Message		Method invocation expression	methodName(parameter, …);
		Method declaration	accessModifier returnType methodName(parameterType parameterName, …) { … }

Combined fragment opt		if-then statement	if (expression1) {
			…
			}

Combined fragment alt		if-then-else statement	if (expression1) {
			…
			} else if (expression2) {
			…
			} …

Combined fragment loop		while statement	while (expression) {
			…
			}

		for statement	for(int i=X; i<Y; i++) { … }
			for(int i=X; i<=Y; i++) { … }
			for(int i=X; i>Y; i--) { … }
			for(int i=X; i>=Y; i--) { … }

the four new transformation rules for the bidirectional transformation between sequence diagram and Java source codes. By the transformation rules in Table 1, the extended RETUSS can transform the following sequence diagram elements into Java source code syntaxes.

- Message of operation invocation with parameters
- Combined fragment opt
- Combined fragment alt
- Combined fragment loop

In addition, the extended RETUSS can transform the following Java source code syntaxes into sequence diagram elements.

- Method invocation expression with parameters
- Method declaration
- if-then statement
- if-then-else statement
- while statement
- for statement

Here, the extended RETUSS does not support nested structure of the above syntaxes.

To implement the transformation rules in Table 1, we extend the description part, transformation part, and storage part. The extended description part has a new process: transformation from Java source codes to Java information. The extended transformation part has a new process: transformation between sequence diagram information and Java information based on the transformation rules in Table 1. The extended storage part has three new classes in the sequence diagram information: interaction fragment class, combined fragment class, and interaction operand class. In addition, the extended storage part has three new classes in the Java information: If class, For class, and While class.

3. Application example

Fig. 3 shows the screenshot when the Java source codes are written in the extended RETUSS. It shows that the extended RETUSS can ensure traceability between sequence diagram and Java source code in writing if-then-else statement, while statement, and for statement of Java. In addition, we confirmed that the extended RETUSS can ensure traceability between sequence diagram and Java source codes in describing sequence diagram.

4. Evaluation

To evaluate the usefulness of the extended RETUSS, we experiment with four students of University of Miyazaki. The steps of the experiment are shown below.

(i) The experimenter prepares traceable sequence diagrams and Java source codes. We call these deliverables.
(ii) The experimenter instructs the participants to change the deliverables.
(iii) The participants change the deliverables as instructed

There are two types in the changes: sequence diagrams changes, Java source codes changes. There are two cases below for participants to change the deliverables.

- Case A: using the extended RETUSS.
- Case B: using EA (Enterprise Architect)[5] and a text editor.

Table 2 shows the times it took the participants to change in the two cases and the two changes. From Table 2, the time in case A was about 76.5% shorter than the time in case B, when the sequence diagrams changes was instructed. In addition, the time in case A was about 69.0% shorter than the time in case B, when the Java source code changes were instructed.

In summary, the extended RETUSS can save labor and time in ensuring traceability between sequence diagram and Java source codes. Therefore, the usefulness of RETUSS has improved by extending its scope of application while retaining the benefits of the existing RETUSS.

5. Conclusion

This paper has improved the usefulness of RETUSS by extending the scope of application of sequence diagram and Java source codes. The extended RETUSS allows you to edit sequence diagram directly on RETUSS, and also supports four new transformation rules for sequence diagram and Java source codes.

The experimental results showed that the extended RETUSS can save the time to ensure traceability between sequence diagram and Java source code by about 76.5% for sequence diagram changes, and about 69.0% for Java source code changes. Therefore, the usefulness of RETUSS has improved by extending its scope of application while retaining the benefits of the existing RETUSS.

The future works are as follows.

- Corresponding to other sequence diagram elements
- Corresponding to other Java source code syntaxes
- Corresponding to other UML diagrams
- Corresponding to other programming languages

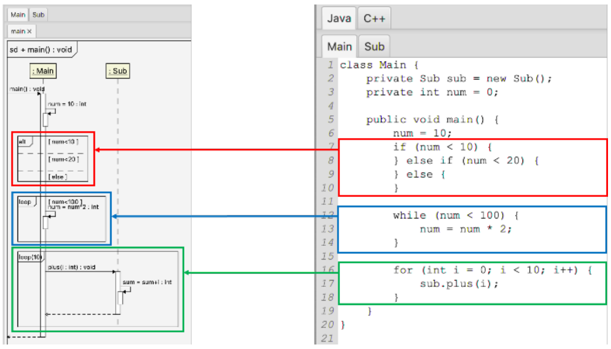


Fig. 3. The screenshot when the Java source codes are written in the extended RETUSS

Table 2. The times it took the participants to change (seconds)

Participants	Sequence diagrams changes		Java source codes changes	
	Case A	Case B	Case A	Case B
1	59	205	134	369
2	62	221	95	346
3	38	216	102	361
4	60	289	126	398
Average	54.75	232.75	114.25	368.50

References

1. SQuBOK Sakutei Bukai, *Guide to the Software Quality Body of Knowledge*, 2nd edn. Ohmsha, 2014 (in Japanese).
2. Tetsuro Katayama, Keisuke Mori, Yoshihiro Kita, Hisaaki Yamaba, Kentaro Aburada, Naonobu Okazaki: *RETUSS: Ensuring Traceability System between Class Diagram in UML and Java Source Code in Real Time*, Journal of Robotics, Networking and Artificial Life, Vol. 5(2), pp. 114–117, 2018.
3. GitHub, *RETUSS: Real-time Ensure Traceability between UML and Source-code System*, https://github.com/Morichan/Retuss (Accessed 2020-12-09)
4. The Object Management Group, *Welcome to UML Web Site!*, https://www.uml.org/ (Accessed 2020-12-09)
5. SPARX SYSTEMS, *UML modeling tools for Business, Software, Systems and Architecture*, https://www.sparxsystems.com/ (Accessed 2020-12-09)

Faster R-CNN Based Defect Detection of Micro-precision Glass Insulated Terminals

Qunpo Liu, Mengke Wang, Zonghui Liu, Bo Su
School of Electrical Engineering and Automation, Henan Polytechnic University, 2001 Century Avenue Jiaozuo, (454003), Henan, P.R.China,

Hanajima Naohiko
College of Information and Systems, Muroran Institute of Technology, 27-1 Mizumoto-cho,Hokkaido Muroran-shi, Hokkaido (050-8585), Japan
E-mail: lqpny@hpu.edu.cn, , hana@mondo.mech.muroran-it.ac.jp
www.hpu.edu.cn, www.muroran-it.ac.jp

(leave space here should be 4 lines between *www.university_name.edu* and **Abstract***)*

Abstract

Micro-precision glass insulated terminals (referred to as glass terminals) are the core components used in precision electronic equipment. As glass terminal, its quality has a huge impact on the performance of precision electronic equipment. Due to limitations in materials and production processes, some of the glass terminals produced have defects such as missing blocks, bubbles, and cracks. At present, it is difficult to ensure product quality and production efficiency with manual inspection methods. However, the defect characteristics of glass terminals are quite different, and it is difficult for traditional defect detection technology to design an ideal feature extractor for detection. Therefore, this paper proposes to use deep learning technology to detect missing blocks. First, preprocess the sample pictures of missing block defects of glass terminals, and then train the deep learning network based on Faster RCNN. According to the test results, the algorithm has an accuracy of over 90% in detecting missing defects in glass terminals.

Keywords: Micro-precision Glass Insulated Terminal; Faster R-CNN； Missing Block Detection

1. Introduction

Micro-precision glass insulated terminals are the core components used in precision electronic equipment. The glass terminal has a small size and requires high accuracy. Defects in the glass terminals would bringing serious losses and consequences[1]. The difficulties in defect detection are mainly three points[2]: (1) The complex imaging background of the defect contains a variety of interference noise; (2) The shape, size, and location of defects are diverse; (3) The difference in the characteristics of defects of the same type is large, and the difference in the characteristics of defects of different types is small. Therefore, this paper proposes to use deep learning technology to detect missing blocks[3].

2. Image acquisition and annotation

The equipment used for image acquisition and photography is the CCD 540TVL (High Resolution B / W Black and White Camera) carried by the MVP400CNC automatic image measuring instrument; the working distance is 86mm; The glass terminal image shooting method adopts the vertical shooting mode of the eight-zone light source; the collection condition is strong light; the magnification during shooting is 1.0*30.77. This article uses labeling software for labeling. After labeling, an xml file will be generated, which can be used for Faster RCNN[4] training.

3. Faster RCNN algorithm

3.1. *Feature extraction network*

The feature extraction network used in this article is ResNet50, and its structure is mainly composed of Conv Block and Identity Block[5,6]. Among them, the Conv Block cannot be directly connected to the network. Its main function is to change the dimension of the network, and the dimensions of the input and output results are different. The identity block can be connected to the network, and its main function is to deepen the network depth, and the input and output results have the same dimensions.

ResNet50
7×7 64,Conv layer,stride 2
3×3 Max pool, stride 2
$\begin{bmatrix}1\times1 & 64\\ 3\times3 & 64\\ 1\times1 & 256\end{bmatrix}\times3$
$\begin{bmatrix}1\times1 & 128\\ 3\times3 & 128\\ 1\times1 & 512\end{bmatrix}\times4$
$\begin{bmatrix}1\times1 & 256\\ 3\times3 & 256\\ 1\times1 & 1024\end{bmatrix}\times6$
$\begin{bmatrix}1\times1 & 512\\ 3\times3 & 512\\ 1\times1 & 2048\end{bmatrix}\times3$
Average pool

Fig.1 ResNet50 network structure

ResNet50 is mainly composed of residual blocks. Its structure is as shown in Figure 1. First, it performs a convolution operation with 64 convolution kernels of size 7*7 and step size of 2, and then performs a maximum pooling with a pooling kernel size of 3*3 Then, after going through 3, 4, 6, and 3 residual blocks, respectively, 1*1 and 3*3 represent the size of the convolution kernel of the residual block, and 64, 256 and so on represent the number of convolution kernels, respectively.

3.2. *RPN*

RPN is the region generation network. By using d n*n window sliding on the shared feature map extracted by the feature extraction network[7], it is mapped into a d-dimensional feature vector, and k Anchors are generated on each pixel. These anchors perform the regression (reg) and classification (cls) operations of the prior box respectively. Assuming that there are k anchors for each pixel on the shared feature map, it is necessary to predict these k pixels, and the classification operation can get 2k outputs , Indicates whether the k area contains the target. The regression operation refers to the calculation of continuously approaching the label frame with k anchors. This operation can get 4k outputs. The k anchors get the center point coordinates and width and height information of the suggested frame (***x,y,w,h***), the process of getting the suggestion box from anchor is also called anchor parameterization. Taking a pixel of the current shared feature map as the center, k anchors with different aspect ratios and sizes are generated, and their aspect ratios and sizes are represented by ratio and scale, respectively. For a feature vector of size w*h, there are a total of w*h*k anchors. The structure of the RPN network is shown in Figure 2.

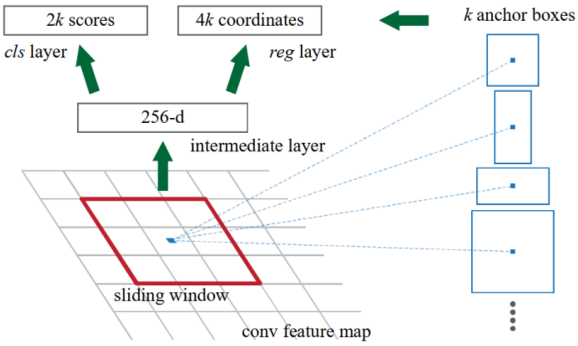


Fig.2 RPN network structure

3.3. *RPN loss function*

The overall loss function of RPN is

$$L(\{p_i\},\{p_i^*\}) = \frac{1}{N_{cls}}\sum_i L_{cls}(p_i,p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i,t_i^*) \quad (1)$$

where, i represents the index of the anchor, p_i and p_i^* respectively indicate that the i-th anchor contains the target predicted value and the anchor tag value. When the i-th anchor contains the target, $p_i^*=1$, and $p_i^*=0$ in other cases. t_i and t_i^* respectively represent the position and scale information of the suggestion box and the label box,

L_{cls} represents the classification loss, that is, the log loss of the anchor corresponding to the target category, and L_{reg} is the regression loss, that is, the loss of the suggestion box obtained from the candidate box, definition for

$$L_{\text{reg}}(t_i, t_i^*) = R(t_i - t_i^*) \quad (2)$$

where, $\boldsymbol{R}$ is the robust loss function (S_{L1}), that is, using Smooth_L1 to calculate the loss value.

It can be seen from $\boldsymbol{p_i^* L_{reg}}$ that the regression loss can be calculated when $\boldsymbol{p_i^*}$=1, and there is no regression loss in other cases. Among them, N_{cls} represents the selection of $\boldsymbol{N_{cls}}$ anchors for RPN training, $\boldsymbol{N_{reg}}$ represents the shape of the shared feature map, and λ is equivalent to a scale factor, so that the weights of the classification loss and the regression loss are basically the same (generally $\lambda = N_{reg}/N_{cls} \approx 10$).

The regression process uses 4 coordinates:

$$t_x = (x - x_a)/w_a, t_y = (y - y_a)/h_a \quad (3)$$

$$t_w = log(w/w_a), t_h = log(h/h_a) \quad (4)$$

$$t_x^* = (x^* - x_a)/w_a, t_y^* = (y^* - y_a)/h_a \quad (5)$$

$$t_w^* = log(w^*/w_a), t_h^* = log(h^*/h_a) \quad (6)$$

where, $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{w}$, and $\boldsymbol{h}$ respectively refer to the center coordinates, width, and height of the a priori box. The variables $\boldsymbol{x}$, $\boldsymbol{x_a}$, and $\boldsymbol{x^*}$ respectively refer to the $\boldsymbol{x}$ coordinates of the suggestion box, anchor a priori box, and label box (the same applies to $\boldsymbol{y}$, $\boldsymbol{w}$, and $\boldsymbol{h}$).

In the RPN network, the obtained prediction frames need to be screened. This process uses IOU and non-maximum suppression methods. The following is an analysis of these two methods.

3.4. *ROI Pooling*

Its main function is to merge the shared feature map and the ROI (region of interest) extracted through RPN, and then divide the prediction frame into a set number of grids, and pool each grid, which is equivalent to The prediction boxes of different sizes in the shared feature map are pooled into feature maps of the same size, which is convenient for inputting subsequent networks for classification and regression.

As shown in Figure 3, the size of the prediction box output by the RPN is fixed to pool_$\boldsymbol{w}$ and pool_$\boldsymbol{h}$ (where pool_$\boldsymbol{w}$ and pool_$\boldsymbol{h}$ are both 7), which is equivalent to

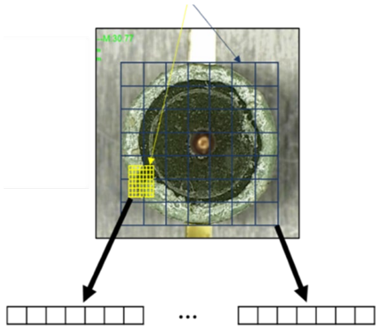


Fig.3 ROI Pooling working diagram

dividing it into 7*7 grids, that is, to achieve a fixed length output .

3.5. *Classification network*

The fully connected layer and the activation function are mainly used to judge and return the object category in the prediction frame to obtain a more accurate prediction frame, and its structure is shown in Figure 4.

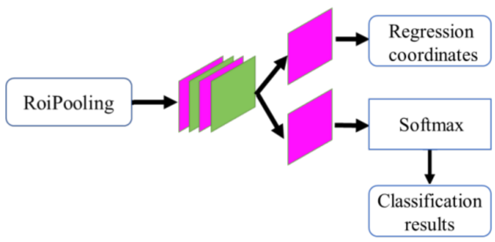


Fig.4 Classification network structure diagram

3.6. *Loss function of Faster RCNN*

Faster RCNN combines the loss functions of the classification model and the regression model, and a calculation formula can be used to calculate the overall loss of the network. The loss function expression after the network is merged is shown in equation (7).

$$L(p, u, t^u, v) = L_{cls}(p, u) + \lambda(u \geq 1) L_{loc}(t^u, v) \quad (7)$$

where, $\boldsymbol{p}$ represents the probability that the corresponding anchor output is a certain category, $\boldsymbol{u}$ is the actual category label corresponding to the anchor, $\boldsymbol{t^u}$ represents the regression of the candidate frame to the scale factor corresponding to the predicted frame, and $\boldsymbol{v}$ is the transformation parameter vector of the candidate frame regression.

$$L_{cls}(p,u) = -\log p_u \tag{8}$$

where, $L_{cls}(p, u)$ represents the classification loss, which is defined as the Equation (8), and the log base is e.

$$L_{loc}(t^u, v) = \sum i \in \{x,y,w,h\} s_{L1}(t_i^u - v_i) \tag{9}$$

where, $L_{loc}(t^u, v)$ represents the loss function of the prediction box obtained by regression of the candidate box, and S_{L1} represents the Smooth L1 loss function.

4. Experimental results

Using Faster RCNN to detect missing blocks and defects requires only 2 hours of training, and the number of training rounds is set to 50. It can detect whether more than 50 samples contain defects in 1 minute, with an accuracy rate of 98.03%, and with training as the number of rounds increases, there is still room for improvement in accuracy. Some results of this paper are shown in Figure 5.

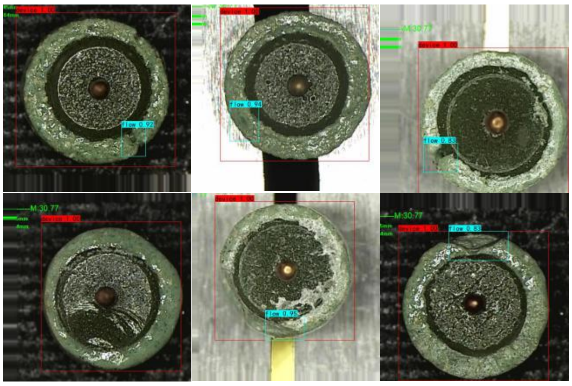


Fig.5 Example of Faster RCNN's Missing Block Defect Detection Results

5. Conclusion

The design of this article is supported by the working mechanism of CNN, and is designed for the implementation of the subsequent Faster RCNN and the detection of missing block defects. The detection process does not require manual feature extraction, but learns features based on the sample of the glass terminal to identify missing block defects and Perform calibration. With the expansion of the data set and the improvement of the model, there is still much room for improvement in the accuracy of missing block defect detection.

Acknowledgements

This work is partially supported by National Key Research and Development Project (2016YFC0600906), Innovation Scientists and Technicians Troop Construction Projects of Henan Province (CXTD2016054) and also supported by Innovative Scientists and Technicians Team of Henan Provincial High Education (20IRTSTHN019).

References

1. Q.Liu, M.Wang, G.Wang, R.Gao, N.Hanajima,Detection Algorithm of Porosity Defect on Surface of Micro-precision Glass,Encapsulated Electrical Connectors, Journal of Robotics, Networking and Artificial Life, 3 (2020) 212-216
2. Q.Liu, M.Wang, and N.Hanajima, Defect Sample Generation System Based On DCGAN for Glass Package Electrical Connectors, Proceedings of 2019 Chinese Intelligent Systems Conference, (2020)
3. WANG Z, ZHU D. An accurate detection method for surface defects of complex components based on support vector machine and spreading algorithm[J]. Measurement, 2019,147:106886.
4. Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J], Neural Information Processing Systems(NIPS). 2015.
5. Wang T, Chen Y, Qiao M, et al. A fast and robust convolutional neural network-based defect detection model in product quality control[J]. The International Journal of Advanced Manufacturing Technology, 2017.
6. He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[C],IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2016.
7. Ren S , He K , Girshick R , et al. Object Detection Networks on Convolutional Feature Maps[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(7):1476-1481.
8. Y.Jia, Alternative proofs for improved LMI representations for the analysis and the design of continuous-time systems with polytopic type uncertainty: a predictive approach, IEEE Transactions on Automatic Control, 48(2003) 1413-1416
9. Y.Jia, General solution to diagonal model matching control of multi-output-delay systems and its applications in adaptive scheme, Progress in Natural Science, 19(2009), 79-90.

Adaptive Sliding Mode Control for a Constant Tension Suspension System

Yuxin 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

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

Kai Gong
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: yuxinjia@buaa.edu.cn, ymjia@buaa.edu.cn, buaa_gk@buaa.edu.cn
www.buaa.edu.cn

Abstract

The constant tension suspension system (CTSS) can be used to counteract the gravity of experimental spacecraft, and is widely used in the ground test of spacecraft control schemes. In this paper, the mechanical model of a CTSS is established, and an adaptive sliding mode controller (ASMC) is designed for the CTSS. The simulation results show that the transient time and steady-state error of the ASMC system are small, which satisfies the requirements of microgravity simulation.

Keywords: constant tension suspension system, microgravity simulation, sliding mode control, adaptive law

1. Introduction

Space technology is a competitive focus of the cutting-edge technology in the world today. Many countries have successively carried out a series of major projects, such as space station construction, deep space exploration, and satellite navigation. These achievements indicate that mankind has entered a new stage of space exploration. On the other hand, higher requirements are also put forward for the reliability of the spacecraft control scheme.[1-4] Because spacecraft work in an area which is difficult for humans to reach, space technology is the research field that relies on ground experiments. The ground test of spacecraft control scheme can not only reduce the risk of actual operation in space, but also shorten the development cycle. Therefore, how to establish microgravity environment on the ground to simulate the real motion of spacecraft is a major demand.[5] Many research institutions have developed different kinds of devices to simulate the microgravity environment. The methods mainly include free fall, parabolic flight, suspension, liquid and air floatation. The

* Corresponding author.

suspension method balances the gravity of experimental spacecraft by adjusting the tension of hanging wire. The system has a simple structure and low cost. It not only overcomes the disturbance of air and liquid floatation on the motion of experimental spacecraft, but also can be used for long-time microgravity simulation.[6] Japan, the United States, Canada, China and other countries have applied the suspension method to microgravity simulation for the ground experiments of spacecraft. In 2019, China Academy of Space Technology (CAST) used the suspension method to counteract part of the gravity of the Mars probe, verifying the performance of the probe control scheme in the Martian gravity environment. The system consists of a 140-meter-high test tower, a follow-up system and a Mars surface simulation area, which satisfies the requirements of the Mars environment simulation experiment. The constant tension suspension system (CTSS) is the key component of suspended microgravity experimental device. It is of great significance to research how to guarantee that the output force of CTSS is constant under disturbance.

An adaptive sliding mode controller (ASMC) is proposed for the CTSS in this paper. Compared with the control method in Ref. 5, the stability of the system is improved.

2. System Modeling

The structure of the CTSS studied in this paper is shown in Fig. 1, including a micro delay servo module, a low stiffness structure, an unconstrained structure, a universal joint, a lifting unit, and an experimental spacecraft.[5]

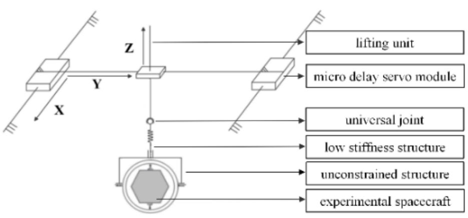


Fig. 1. Structure of the CTSS

The micro delay servo module is composed of X-direction unit and Y-direction unit. It can measure the inclination of low stiffness structure by tilt sensor, and drive the lifting unit to track the movement of experimental spacecraft quickly, reducing the disturbance of the horizontal component of output tension on the movement of experimental spacecraft. The universal joint and unconstrained structure can ensure that the freedom of experimental spacecraft is not restricted, so that the experimental spacecraft can perform approximate unconstrained pitching, yawing and rolling motion under a small frictional torque disturbance. The vertical movement of lifting unit ensures that the output tension of the CTSS is always equal to the gravity of experimenttal spacecraft, realizing the long-term simulation of microgravity environment. The physical meanings of the main symbols are as follows:

- g, Gravitational acceleration.
- R, Radius of the gear in lifting unit.
- k, Elastic coefficient of low stiffness structure.
- ω, Swing angle between lifting unit and low stiffness structure.
- L_0, Initial length of low stiffness structure.
- L, Variation length of low stiffness structure.
- M, The sum of the masses of unconstrained structure and experimental spacecraft.
- T_α, Driving torque of the motor in lifting unit.
- M_1, M_2, The loads of the motors in X-direction and Y-direction units.
- M_a, M_b, The masses of the gear and rack in lifting unit.
- F_x, F_y, Driving forces of the motors in X-direction and Y-direction units.
- δ_x, δ_y, δ_z, Disturbing forces.
- ω_x, ω_y, Orthogonal decomposition values of ω.

According to the operation principle of CTSS, the model can be obtained by mechanical analysis.

$$\boldsymbol{H}\ddot{\boldsymbol{q}} + \boldsymbol{J}\dot{\boldsymbol{q}} + \boldsymbol{N} = \boldsymbol{F} + \tilde{\boldsymbol{H}}\boldsymbol{\delta} \tag{1}$$

Where $\boldsymbol{F}$ is the control vector of the CTSS, $\boldsymbol{q}$ is the state vector. $\boldsymbol{\delta}$ is the external disturbance, and it is assumed that $\boldsymbol{\delta}$ is bounded. The matrices defined in the model is as follows:

$$\boldsymbol{H}=\begin{bmatrix} -\omega_x(\eta_1-1) & -d(\eta_1-1) & 0 \\ -\omega_y(\eta_2-1) & 0 & -d(\eta_2-1) \\ -R(M-M_c)/M & 0 & 0 \end{bmatrix}$$

$$\tilde{\boldsymbol{H}}=\begin{bmatrix} 0 & -\eta_1/d & 0 \\ 0 & 0 & -\eta_2/d \\ RM_c/M & -\omega_x RM_c/(Md) & -\omega_y RM_c/(Md) \end{bmatrix}$$

$$\boldsymbol{J}=\begin{bmatrix} -2\dot{\omega}_x(\eta_1-1) & 0 & 0 \\ -2\dot{\omega}_y(\eta_2-1) & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},\ \boldsymbol{F}=\frac{1}{M}\begin{bmatrix} F_x \\ F_y \\ T_\alpha \end{bmatrix}, \boldsymbol{\delta}=\frac{1}{M}\begin{bmatrix} \delta_z \\ \delta_x \\ \delta_y \end{bmatrix}$$

$$\eta_1=(M+M_1)/M\ ,\ \eta_2=(M+M_2)/M$$

$$N = \begin{bmatrix} 0 \\ 0 \\ R\left(g + \left(MM_b g + LM_c k\right)/M^2\right) \end{bmatrix}, q = \begin{bmatrix} L \\ \omega_x \\ \omega_y \end{bmatrix}$$
$$d = L + L_0 + Mg/k,\ M_c = M_a/2 + M_b + M$$

3. ASMC Design

In order to suppress the uncertain disturbance of experimental spacecraft during the operation of CTSS, a new ASMC is designed in this section. Define variables:

$$\boldsymbol{A}(\boldsymbol{q},\dot{\boldsymbol{q}}) = -\boldsymbol{H}^{-1}(\boldsymbol{J}\dot{\boldsymbol{q}} + \boldsymbol{N})$$
$$\boldsymbol{B}(\boldsymbol{q}) = \boldsymbol{H}^{-1},\ \tilde{\boldsymbol{\delta}} = \boldsymbol{H}^{-1}\tilde{\boldsymbol{H}}\boldsymbol{\delta} = \left[\tilde{\boldsymbol{\delta}}_1, \tilde{\boldsymbol{\delta}}_2, \tilde{\boldsymbol{\delta}}_3\right]^{\mathrm{T}}$$

The system model (1) is transformed into

$$\ddot{\boldsymbol{q}} = \boldsymbol{A}(\boldsymbol{q},\dot{\boldsymbol{q}}) + \boldsymbol{B}(\boldsymbol{q})\boldsymbol{F} + \tilde{\boldsymbol{\delta}} \tag{2}$$

The simulation of total weightlessness environment is studied in this paper, so the expected value of $\boldsymbol{q}$ is defined as $\boldsymbol{q}_d = [0, 0, 0]^{\mathrm{T}}$. The system error is $\boldsymbol{e} = \boldsymbol{q} - \boldsymbol{q}_d$, and the upper bound of $|\tilde{\boldsymbol{\delta}}|$ is $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3]^{\mathrm{T}}$. Thus, Eq. (2) is transformed into

$$\ddot{\boldsymbol{e}} = \boldsymbol{A}(\boldsymbol{q},\dot{\boldsymbol{q}}) + \boldsymbol{B}(\boldsymbol{q})\boldsymbol{F} + \tilde{\boldsymbol{\delta}} - \ddot{\boldsymbol{q}}_d \tag{3}$$

The switching function is defined as:

$$\begin{gathered} \boldsymbol{s} = \boldsymbol{D}\boldsymbol{e} + \boldsymbol{V}\mathrm{sgn}(\dot{\boldsymbol{e}})^z \\ \mathrm{sgn}(\dot{\boldsymbol{e}})^z = \left[|e_1|^z \mathrm{sign}(e_1), |e_2|^z \mathrm{sign}(e_2), |e_3|^z \mathrm{sign}(e_3)\right]^{\mathrm{T}} \end{gathered} \tag{4}$$

Where, $\boldsymbol{D} = \mathrm{diag}(d_1, d_2, d_3)$ and $\boldsymbol{V} = \mathrm{diag}(v_1, v_2, v_3)$ are positive definite matrices, $\boldsymbol{s} = [s_1, s_2, s_3]^{\mathrm{T}}$, $z = \mathrm{diag}(z, z, z)$ and satisfies $1 < z < 2$.

Define $\boldsymbol{\varphi} = [\varphi_1, \varphi_2, \varphi_3]^{\mathrm{T}}$ is the adaptive parameter of the ASMC, and the adaptive law is selected as:

$$\dot{\boldsymbol{\varphi}} = \left[\frac{|s_1|}{f_1}, \frac{|s_2|}{f_2}, \frac{|s_3|}{f_3}\right]^{\mathrm{T}} \tag{5}$$

Where, f_1, f_2, f_3 are all positive constants.

Assumption 1. The stable value $\hat{\boldsymbol{\varphi}} = [\hat{\varphi}_1, \hat{\varphi}_2, \hat{\varphi}_3]^{\mathrm{T}}$ of the adaptive parameter satisfies $\hat{\varphi}_p > \beta_p,\ p = 1, 2, 3$.

Theorem 1. Under Assumption 1, the system error $\boldsymbol{e}$ will converge to **0**, if the ASMC $\boldsymbol{F}$ is designed as:

$$\begin{aligned} &\boldsymbol{F} = \boldsymbol{F}_e + \boldsymbol{F}_d + \boldsymbol{F}_s \\ &\boldsymbol{F}_e = -\boldsymbol{B}(\boldsymbol{q})^{-1}\mathrm{diag}\left(\frac{1}{zv_1|\dot{\boldsymbol{e}}_1|^{z-1}}, \frac{1}{zv_2|\dot{\boldsymbol{e}}_2|^{z-1}}, \frac{1}{zv_3|\dot{\boldsymbol{e}}_3|^{z-1}}\right)\boldsymbol{D}\dot{\boldsymbol{e}} \\ &\boldsymbol{F}_d = \boldsymbol{B}(\boldsymbol{q})^{-1}\left(-\boldsymbol{A}(\boldsymbol{q},\dot{\boldsymbol{q}}) + \ddot{\boldsymbol{q}}_d\right) \\ &\boldsymbol{F}_s = -\boldsymbol{B}(\boldsymbol{q})^{-1}\left(\mathrm{diag}(\boldsymbol{\varphi})\mathrm{sign}(\boldsymbol{s})\right) \end{aligned} \tag{6}$$

When $\dot{\boldsymbol{e}}_p = 0$, define

$$1\Big/\left(zv_p|\dot{\boldsymbol{e}}_p|^{z-1}\right) = 0,\ p = 1, 2, 3$$

Proof. The error of the adaptive parameter $\boldsymbol{\varphi}$ is defined as $\tilde{\boldsymbol{\varphi}} = \boldsymbol{\varphi} - \hat{\boldsymbol{\varphi}}$, and the positive definite scalar function is selected as:

$$\begin{gathered} V_L = \frac{1}{2}s^{\mathrm{T}}s + \frac{1}{2}\tilde{\boldsymbol{\varphi}}\mathrm{diag}(f_1, f_2, f_3)\mathrm{diag}(\theta_1, \theta_2, \theta_3)\tilde{\boldsymbol{\varphi}} \\ \mathrm{diag}(\theta_1, \theta_2, \theta_3) = \boldsymbol{z}\boldsymbol{V}\mathrm{diag}\left(|\dot{\boldsymbol{e}}|^{z-1}\right) \end{gathered} \tag{7}$$

The differentiation of Eq. (4) and Eq. (7) is

$$\begin{gathered} \dot{\boldsymbol{s}} = \boldsymbol{D}\dot{\boldsymbol{e}} + \boldsymbol{z}\boldsymbol{V}\mathrm{diag}\left(|\dot{\boldsymbol{e}}|^{z-1}\right)\ddot{\boldsymbol{e}} \\ \dot{V}_L = s^{\mathrm{T}}\dot{s} + \tilde{\boldsymbol{\varphi}}^{\mathrm{T}}\mathrm{diag}(f_1, f_2, f_3)\mathrm{diag}(\theta_1, \theta_2, \theta_3)\dot{\tilde{\boldsymbol{\varphi}}} \end{gathered} \tag{8}$$

Substitute the ASMC Eq. (6) into Eq. (8)

$$\begin{aligned} \dot{V}_L &= \boldsymbol{s}^{\mathrm{T}}\mathrm{diag}(\theta_1, \theta_2, \theta_3)\left(-\mathrm{diag}(\boldsymbol{\varphi})\mathrm{sign}(\boldsymbol{s}) + \tilde{\boldsymbol{\delta}}\right) \\ &\quad + \tilde{\boldsymbol{\varphi}}^{\mathrm{T}}\mathrm{diag}(\theta_1, \theta_2, \theta_3)\left[|\boldsymbol{s}_1|, |\boldsymbol{s}_2|, |\boldsymbol{s}_3|\right]^{\mathrm{T}} \\ &= \theta_1\left(\tilde{\boldsymbol{\delta}}_1\boldsymbol{s}_1 - \hat{\varphi}_1|\boldsymbol{s}_1|\right) + \theta_2\left(\tilde{\boldsymbol{\delta}}_2\boldsymbol{s}_2 - \hat{\varphi}_2|\boldsymbol{s}_2|\right) \\ &\quad + \theta_3\left(\tilde{\boldsymbol{\delta}}_3\boldsymbol{s}_3 - \hat{\varphi}_3|\boldsymbol{s}_3|\right) \le 0 \end{aligned} \tag{9}$$

According to the analysis of Eq. (9), the controller Eq. (6) can make the system stable. Further, substituting the controller Eq. (6) into the system error model Eq. (3), it can be found that the solution $(\boldsymbol{e}, \dot{\boldsymbol{e}})$ of the system error model will not make $\dot{V}_L \equiv 0$ except for the origin $(\boldsymbol{0}, \boldsymbol{0})$. Therefore, the system error $\boldsymbol{e}$ will converge to **0**. The proof is completed.

4. Experimental Analysis

In this section, the ASMC is applied to the CTSS for simulation experiments. The adaptive parameter of the ASMC is dynamically adjusted according to the value of Eq. (4), and finally converge to a constant. In order to reduce the oscillation caused by the Sign($\boldsymbol{s}$), Eq. (10) is used to replace it in the simulation.

$$\boldsymbol{s}\big/\left(\left|\boldsymbol{s}\right|+0.05\right) \tag{10}$$

The parameters of the CTSS model Eq. (1) are R=0.02m, M=15.5kg, M_a=0.52kg, M_b=4kg, M_1=16kg, M_2=41.5kg, k=690N/m, L_0=0.45m, g=10m/s^2. The pulse disturbance shown in Fig. 2 is applied to the system.

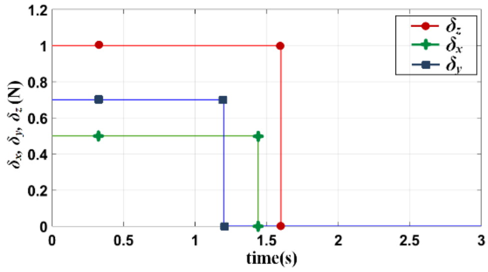


Fig. 2. External disturbance

The parameters of the ASMC are as follows:

$$f_1 = f_2 = f_3 = 1, \overline{\boldsymbol{\varphi}} = [1, 1, 1]^{\mathrm{T}}, z{=}1.1$$
$$\boldsymbol{D}{=}\mathrm{diag}(280, 510, 450),\ \boldsymbol{V}{=}\mathrm{diag}(7.5, 8, 9)$$

Where $\overline{\boldsymbol{\varphi}}$ is the initial value of the adaptive parameter.

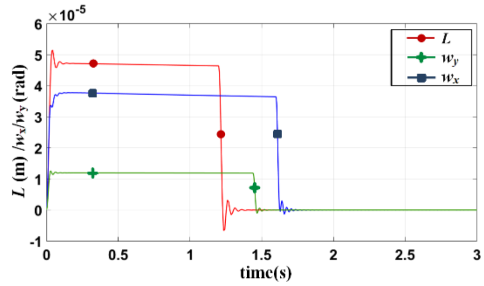


Fig. 3. State response of the ASMC system

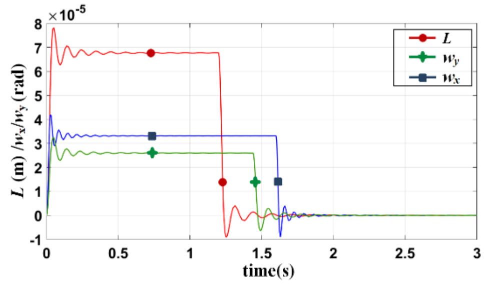


Fig. 4. State response of the SMC system in Ref. 5

The state response is shown in Fig. 3. It can be found that the maximum error of ω_x is 3.779×10^{-5} rad, which is 0.0022°, the maximum error of ω_y is 1.254×10^{-5} rad, which is 0.0007°, and the maximum error of L is 5.151×10^{-5} m. Therefore, the output tension error of the system is 0.036 N, which satisfies the requirements of microgravity simulation.

The SMC in Ref. 5 is also applied to the CTSS, and its state response is shown in Fig. 4. The results show that the convergence speed of the ASMC system is faster and the steady-state error is smaller. When the simulation time t >1.6 s, the state error of the ASMC system satisfies $\|\boldsymbol{e}\| < 4\times10^{-6}$, while the SMC system needs 1.7 s.

5. Conclusions

A new ASMC is designed for the model of CTSS in this paper. The adaptive parameter can be dynamically adjusted with the value of the switching function. The asymptotic stability of the ASMC system is proved, and the simulation experiment is carried out to compare with the controller in Ref. 5. Experiments show that the ASMC system has smaller steady-state error and faster convergence speed, which satisfies the requirements of microgravity simulation.

Acknowledgements

This work was supported by the NSFC (61520106010, 61327807, 61134005, 61521091) and the National Basic Research Program of China (973 Program: 2012CB 821200, 2012CB821201).

References

1. Z. Jiang, J. Xu, H. Li, et al, Stable Parking Control of a Robot Astronaut in a Space Station Based on Human Dynamics, *IEEE Trans. Robot.*, 36(2), 2020, pp. 399-413.
2. C. Wang, H. Nie, J. Chen, et al, The design and dynamic analysis of a lunar lander with semiactive control, *Acta Astronaut.*, 157, 2019, pp. 145-156.
3. X. Liu, G. Cai, M. Wang, et al. Contact control for grasping a non-cooperative satellite by a space robot, *Multibody Syst. Dyn*., (5), 2020, pp. 119-141.
4. A. Witze, NASA rethinks approach to Mars exploration, *Nature*, 538(7624), 2016, pp. 149-150.
5. J. Jia, Y. Jia, S. Sun, Continuous Non-singular Fast Terminal Sliding Mode Control for an Active Gravity Field Simulator, *Journal of Robotics Networking and Artificial Life*, 3(4), 2017, pp. 253-256.
6. V. Arakelian, Gravity compensation in robotics, *Adv. Robot.*, 30(2), 2016, pp. 79-96.

Weighted Multiple Model ADRC for Uncertain Linear Systems

Weicun Zhang[1*], Liuqiao Song[2], Jing Ge[1]
[1]School of Automation and Electrical Engineering, University of Science and Technology Beijing,

30 Xueyuan Road Haidian District, Beijing, China
E-mail: weicunzhang@ustb.edu.cn
www.ustb.edu.cn

[2]School of electrical engineering and computer science, Oregon State University,

Corvallis, Oregon OR 97331-5501, USA

Abstract

For uncertain linear systems such as parameter jumping, this paper presents a weighted multiple model adaptive control frame work, which uses fixed model set to cover the uncertainties of the real plant to be controlled, and for each local model with minor uncertainties, the corresponding local controller is designed based on active disturbance rejection controller (ADRC) approach. Some simulations have been conducted based on MATLAB to verify the effectiveness of the proposed weighted multiple model ADRC adaptive control algorithm.

Keywords: Multiple model adaptive control, ADRC, weighting algorithm.

1. Introduction

For systems with large uncertainties, such as, model structure change or parameter jumping, conventional adaptive control can no longer meet the performance needs of industrial control systems. Therefore, the control research of uncertain systems has important theoretical significance and practical significance. Weighted multiple model adaptive control (WMMAC) [1-12] can effectively control complex and uncertain systems with prescribed performance indexes. Aiming at the uncertain system, this paper proposes a weighted multiple model adaptive control approach with active disturbance rejection control (ADRC) as local control strategy, which was originally proposed by Professor Jingqing Han [13-20].

Aiming at a class of uncertain linear controlled plants with nonlinear disturbances, this paper combines the active disturbance rejection controller and the weighted multiple model adaptive control system to design a weighted multiple model ADRC control system for the second-order time-delay system and a weighted multiple model ADRC control system for the third-order uncertain system, so that to completed the control of these two uncertain linear systems with nonlinear disturbances. The simulation results show that compared with a single ADRC control system, the weighted multiple model ADRC control system can better meet the control

*Corresponding author：W. Zhang, weicunzhang@ustb.edu.cn

performance requirements and has stronger tracking ability and robustness when the parameters are uncertain in a large range or with multiple jumps.

The block diagram of the proposed system is shown in Fig. 1.

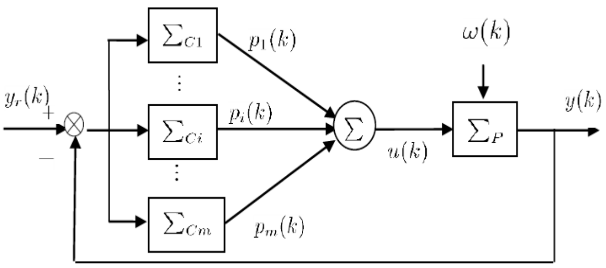


Figure 1 The block diagram of the proposed system

2. Local controller design based on ADRC

Consider a plant with small uncertainties

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_3 + b0u(t) \\ \dot{x}_3 = \dot{f}(x_1, x_2, \omega(t), t) = h(t) \\ y = x_1 \end{cases} \tag{1}$$

where $f(.)$ represents unknown unmodeled dynamics or 'total disturbance', $u(t)$ control input, $\omega(t)$ external disturbance, $b0$ control gain, $y(t)$ system output.

First, we need to design the extended state observer (ESO)

$$\begin{cases} \varepsilon_1 = z_1 - y \\ \dot{z}_1 = z_2 - \beta_{01}\varepsilon_1 \\ \dot{z}_2 = z_3 - \beta_{02}\varepsilon_1 + b_0 u \\ \dot{z}_3 = -\beta_{03}\varepsilon_1 \end{cases} \tag{2}$$

Its discrete-time version has the following form

$$\begin{cases} \varepsilon_1 = z_1(k) - y(k) \\ z_1(k+1) = z_1(k) + h(z_2(k) - \beta_{01}\varepsilon_1) \\ z_2(k+1) = z_2(k) + h(z_3(k) - \beta_{02}\varepsilon_1 + b_0 u) \\ z_3(k+1) = z_3(k) - h\beta_{03}\varepsilon_1 \end{cases} \tag{3}$$

Where ε_1 is observation error, β_{01}, β_{02}, and β_{03} are the gains of the observer, z_1, z_2, and z_3 are the estimations of y, $\dot{y}$, and $f(.)$, respectively.

Second, we need to design the control law based on the ESO

$$u_0 = k_p e_1 - k_d z_2 \tag{4}$$

where $e_1 = v_1 - z_1$, v is the reference input of the system, k_p and k_d are adjustable parameters.

Based on (4), the local control law is given by

$$u = \frac{u_0 - z_3}{b_0} \tag{5}$$

3. Multiple model ADRC

Consider time-varying uncertain plant

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = x_3 + b(t)u(t) \\ \dot{x}_3 = \dot{f}(x_1, x_2, d(t), \omega(t), t) = h(t) \\ y = x_1 \end{cases} \tag{6}$$

Then we need to design multiple model ADRC, which uses multiple models (model set) to cover the uncertainties of $b(t)$, for each (local) model, a local controller (ADRC controller) is designed off-line, and the global control law is weighted sum (on-line) of all the local controllers. That is

$$u(t) = \sum p_i(t) u_i(t)$$

where u_i(t) is calculated according to (5).

The weighting algorithm is described as follows (Suppose the number of local controllers is N)

where e_i is the output error of each local model, i.e., $e_i(k) = y(k) - y_i(k)$.

The calculation of weights values is carried out at each time instant k, and through a ZOH (Zero Order Hold), $p_i(t)$can be obtained from $p_i(k)$.

4. Simulation results

To verify the effectiveness of the proposed multiple model ADRC framework, many simulations have been conducted with $\omega(t)=0.1\sin t$ in equation (6) , based on MATLAB/ Simulink.

Suppose the plant will change from Model 1 (corresponding to $b(t) = 0.6$ in equation (6)) to Model 2 (corresponding to $b(t) = 1$ in equation (6)) at time 30s, and finally change to Model 3 (corresponding to $b(t) = 1.5$ in equation (6)) at time 60s. The model set include 4 models, i.e., Model1, Model 2, Model 3, and Model 4 (corresponding to $b(t) = 2$ in equation (6)). The step responses of multiple model ADRC system and single ADRC system are shown in Fig. 2.

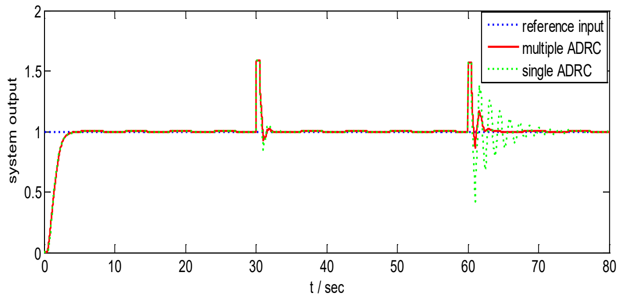


Figure 2. Step response comparison between multiple model ADRC and single ADRC

Weight values of multiple model ADRC are shown in Figure 3.

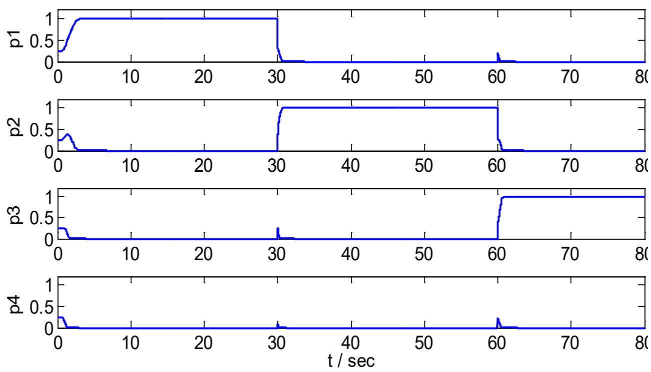


Figure 3. Weight values of multiple model ADRC

5. Conclusion and future work

This paper presented a multiple model ADRC framework for uncertain dynamic time-varying system. Simulation results verified the effectiveness of the proposed methods including weighting algorithm and local controller scheme. In the future research work, we will focus on model set selection for time-varying system with more than one variable.

Acknowledgements

This work was partly supported by National Natural Science Foundation of China (No. 61520106010; 61741302).

References

1. Fekri, M. Athans and A. Pascoal, Issues, progress and new results in robust adaptive control, International Journal of Adaptive Control and Signal Processing, 2006; 20(10): 519-579
2. D. T. Magill, Optimal adaptive estimation of sampled stochastic processes, IEEE Trans. Automat. Contr., 1965; 10: 434-439
3. D. G. Lainiotis, Partitioning: A unifying framework for adaptive systems-I: Estimation; II: Control, Proc. IEEE, 1976; 64:1126-1143 and 1182-1197
4. M. Athans et al., The stochastic control of the F-8C aircraft using a multiple model adaptive control (MMAC) method-Part I: Equilibrium flight, IEEE Trans. Automat. Contr., 1977, 22: 768-780
5. A. Badr, Z. Binder, D. Rey, Weighted multi-model control, International journal of systems science, 1992; 23(1): 145-149
6. M. Kuipers, P. Ioannou, Practical Robust Adaptive Control: Benchmark Example, Proc. of American Control Conference, Seattle, Washington, USA, 2008:5168-5173
7. Kuipers and P. Ioannou, Multiple model adaptive control with mixing, IEEE Transactions on Automatic Control, 2010; 55(8), 1822-1836
8. S. Baldi, P. A. Ioannou, and E. Mosca, Multiple Model Adaptive Mixing Control: the Discrete-time case, IEEE Trans. on Automatic Control, 2012; 57(4):1040-1045
9. N. Sadati, G. A. Dumont, and H. R. Feyz Mahdavian, Robust Multiple Model Adaptive Control Using Fuzzy Fusion, 42nd South Eastern Symposium on System Theory, Tyler, TX, USA, March 7-9, 2010
10. Z. Han, and K. S. Narendra, New Concepts in Adaptive Control Using Multiple Models, IEEE Transactions on Automatic Control, 2012; 57(1): 78-89
11. W. Zhang, Stable Weighted Multiple Model Adaptive Control: Discrete-Time Stochastic Plant, International Journal of Adaptive Control and Signal Processing, 2013; 27(7): 562-581

12. W. Zhang, Further Results on Stable Weighted Multiple Model Adaptive Control: Discrete-Time Stochastic Plant, International Journal of Adaptive Control and Signal Processing, 2015; 29(12): 1497-1514
13. J. Han. Active disturbance rejection control technique-the technique for estimating and compensating the uncertainties. Beijing: National Defense Industry Press; 2008.
14. J. Han. From PID to active disturbance rejection control. IEEE Transactions on Industrial Electronics 2009; 56(3): 900-906.
15. Aaron Radke, and Z. Gao. A survey of state and disturbance observers for practitioners, Porc. American Control Conference. 2006 5183-5188.
16. Y. Huang, W.C. Xue. Active disturbance rejection control: Methodology and theoretical analysis. ISA Transactions 2014; 53: 963-976.
17. Wenchao Xue, Wenyan Bai, Shen Yang, Kang Song, Yi Huang, and Hui Xie. ADRC with adaptive extended state observer and its application to air-fuel ratio control in gasoline engines. IEEE Transactions on Industrial Electronics 2015; 62: 5847-5857.
18. Z. Gao. Active disturbance rejection control: a paradigm shift in feedback control system design. Proc. American Control Conference. 2006 2399-2405.
19. Qing Zheng, Linda Q. Gao, & Zhiqiang Gao. On stability analysis of active disturbance rejection control for nonlinear time-varying plants with unknown dynamics. Proc. IEEE Conference on Decision and Control. 2007 3501-3506.
20. Qing Zheng, Zhongzhou Chen, & Zhiqiang Gao. A practical approach to disturbance decoupling control. Control Engineering Practice 2009; 17: 1016-1025.

Trajectory Tracking Control of Differential Wheeled Mobile Robots

Based on Rhombic Input Constraints

Kai Gong
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

Yuxin 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: buaa_gk@*buaa.edu.cn, ymjia@buaa.edu.cn, yuxinjia@buaa.edu.cn,*
www.buaa.edu.cn

Abstract

This paper focuses on the trajectory tracking control algorithm for differential wheeled mobile robots (DWMRs) based on rhombic input constraints. The kinematics and dynamics model of DWMRs are Established, and vector analysis method is used to design the controller when the linear velocity and angular velocity of DWMRs were not independent of each other. Through the trajectory tracking simulation of the 8-shaped curve, a good control performance is obtained.

Keywords: differential wheeled mobile robots, rhombic input constraints, trajectory tracking, vector analysis

1. Introduction

The tracking control of DWMRs has a very broad application background. There are many methods have been used in controller design for trajectory tracking. Sliding mode control,[1] backstepping control,[2] robust control,[3] fuzzy control,[4] active disturbance rejection control[5] etc. are used to solve this problem. From a practical perspective, the input constraints must be considered when designing controller, however most existing studies are assume that the input constraints of the robots' linear velocity v and angular velocity ω are independent of each other, that is, $|v| \le m_1$, $|\omega| \le m_2$, where m_1 and m_2 are positive constants. (Fig.1). A proof will be given later, the actual situation is that the input field of DWMRs is the rhombic area defined by $|v/m| + |\omega\, l/m| \le 1$, where m is the maximum velocity of the two driving wheels and l is half of the distance

*Corresponding author.

between them. If the controller in Ref.6 is applied to a differential drive robots, the rectangle estimation needs to be extracted in the rhombic input field. The largest rectangle estimate that can be calculated is determined by $|v| \le m/2$ and $|\omega| \le m/2l$, it is only half of the rhombic input field, which leads to the robots' mobility cannot be fully utilized. Rhombic input constraints are considered first time in Ref.7, it proposed a geometric analysis method to design time-varying feedback parameters.

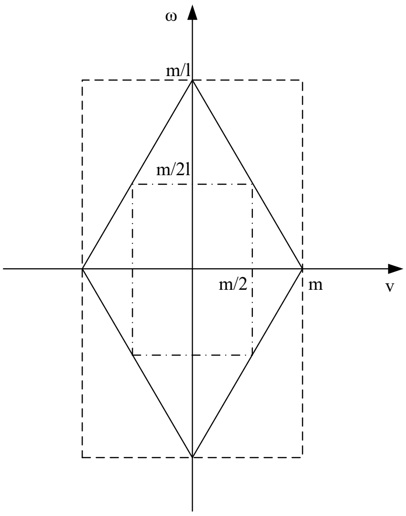


Fig. 1. Rectangular and diamond constraints

2. Problem Statement

2.1. *Rhombic Input Constraints*

Consider DWMRs shown in Fig.2, the driving wheels' velocities of the robots are v_l and v_r respectively. Assuming that the two driving wheels have the same performance, there maximum velocities are both m, that is $v_l \le m$ and $v_r \le m$. Usually v and ω of DWMRs are used as control inputs, and their relationship with the speed of the driving wheel is

$$v = (v_l + v_r)/2 \tag{1}$$

$$\omega = (v_r - v_l)/2l \tag{2}$$

Thus v and ω are constrained by

$$\begin{cases} -(m+v)/l \le \omega \le (m+v)/l, v \in [-m, 0) \\ -(m-v)/l \le \omega \le (m-v)/l, v \in [0, m] \end{cases} \tag{3}$$

The above is collated into one expression:
Equation (3) can be sorted into one expression:

$$|v/m| + |\omega\, l/m| \le 1 \tag{4}$$

This constraints is shown in the Fig.1 as a rhombus with black solid line.

2.2. *Tracking Control Based on Rhombic Input Constraints*

The kinematics and dynamics equations of two-wheel differential mobile robots is

$$\dot{x} = v\cos\theta,\ \ \dot{y} = v\sin\theta,\ \ \dot{\theta} = \omega \tag{5}$$

(x, y) is the center point coordinates of DWMRs and θ is used to indicate its azimuth angle (see Fig.2).

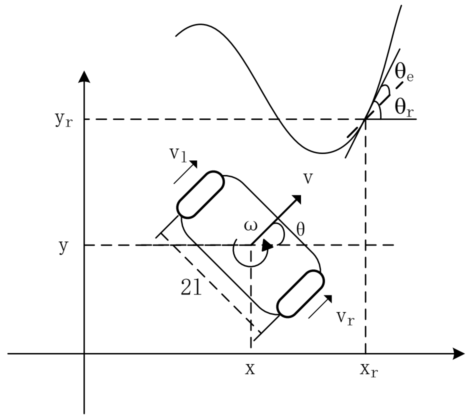


Fig. 2. Trajectory tracking of DWMRs

Assumption 1. The input constraint of DWMRs is (4), and its reference trajectory satisfies:

$$\dot{x}_r = v_r\cos\theta_r,\ \ \dot{y}_r = v_r\sin\theta_r,\ \ \dot{\theta}_r = \omega_r \tag{6}$$

and

$$|v_r/m| + |\omega_r l/m| \le 1 - l\varepsilon/m \tag{7}$$

Among them, $(x_r, y_r, \theta_r, v_r, \omega_r)$ is the target values of $(x, y, \theta, v, \omega)$, where ε is a constant satisfies $0 < \varepsilon < m/l$.

Remark 1. We introduce a constant ε in equation (7) to ensure that the reference trajectory can be tracked by DWMRs under the input constraints (4)

Fig.2 shows that system errors of DWMRs are defined as:

$$\begin{bmatrix} x_e \\ y_e \\ \theta_e \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_r - x \\ y_r - y \\ \theta_r - \theta \end{bmatrix} \tag{8}$$

The tracking errors system can be obtained by deriving the two sides of the above formula (8)

$$\begin{aligned} \dot{x}_e &= v_r\cos\theta_e - v + \omega y_e \\ \dot{y}_e &= v_r\sin\theta_e - \omega x_e \\ \dot{\theta}_e &= v_r\omega_r - \omega \end{aligned} \tag{9}$$

Now our goal is to design the controller to make the system errors x_e, y_e and θ_e tend to zero, and the control variables v and ω must meet the constraints (4).

3. Controller Design Based on Rhombic Input Constraints

Before our controller design work begins, we first understand two important lemmas

Lemma 1.[8] $f:[0,\infty)\to R$ is first-order continuous differentiable and $\lim_{t\to\infty} f(t)$ is a finite value, if $\dot{f}(t), t\in[0,\infty)$ is uniformly continuous, then $\lim_{t\to\infty}\dot{f}(t)=0$.

Lemma 2.[7] There is a scalar function $\rho(x), x\in[0,\infty)$, which satisfies the following properties:

1. $\rho(x)$ is a continuous and non-decreasing function;
2. $\rho(0)=0$, and $0<\rho(x)\le 1$ for $x\in(0,\infty)$;
3. $\lim_{x\to 0^+}\rho'(x)=\rho_0$, which ρ_0 is a positive constant.

Define $\psi(x)$ as

$$\psi(x)=\begin{cases}\rho(x)/x & x\in(0,\infty)\\ \rho_0 & x=0\end{cases} \tag{10}$$

Then, for $\forall\sigma\in(0,\infty)$, there always exist α and β, such that $\alpha\le\varphi(x)\le\beta$ holds for $x\in[0,\sigma]$, where both α and β are positive constants.

$\rho(x)=\tanh(x)$ is a function that satisfies the above conditions.

There are many results about the design of the tracking controller of the differential drive robots in the existing literature. In this paper we choose the controller in Ref.9.

$$\begin{aligned} v &= v_r\cos\theta_e + k_x x_e \\ \omega &= \omega_r + k_y v_r y_e \frac{\sin\theta_e}{\theta_e} + k_\theta\theta_e \end{aligned} \tag{11}$$

where k_x, k_y and k_θ are positive constants. Through formula (11), we can easily see that too large errors will cause the control variables v and ω to be too large, and then the control variables will exceed the constraints. In this way, the control commands cannot be executed well.

Lemma 3.[7] For controller (11), if following conditions are met:

1) $\underline{k}_x\le k_x\le\overline{k}_x, \underline{k}_y\le k_y\le\overline{k}_y, \underline{k}_\theta\le k_\theta\le\overline{k}_\theta$

2) k_y is differentiable and $\dot{k}_y\ge 0$.

where $\underline{k}_x,\overline{k}_x,\underline{k}_y,\overline{k}_y,\underline{k}_\theta,\overline{k}_\theta$ are positive constant values, Then, trajectory tracking errors of DWMRs will converge to zero, that is x_e, y_e, θ_e will converge to zero.

Controller (11) can be designed by using the vector analysis method, define the controller v,ω as a vector

$$\overrightarrow{OD}=\begin{bmatrix} v\\ \omega\end{bmatrix}$$

then by defining other vectors as:

$$\overrightarrow{OA}=\begin{bmatrix} v_r\cos\theta_e\\ \omega_r\end{bmatrix}, \overrightarrow{AB}=\begin{bmatrix} 0\\ k_y v_r y_e\dfrac{\sin\theta_e}{\theta_e}\end{bmatrix}$$
$$\overrightarrow{BC}=\begin{bmatrix} k_x x_e\\ 0\end{bmatrix}, \overrightarrow{CD}=\begin{bmatrix} 0\\ k_\theta\theta_e\end{bmatrix} \tag{12}$$

Then we can get the vector representation of the controller

$$\overrightarrow{OD}=\overrightarrow{OA}+\overrightarrow{AB}+\overrightarrow{BC}+\overrightarrow{CD} \tag{13}$$

In order to make the controller meet the rhombic input constraints, we need to analyze each vector in turn.

$$\left|\frac{v_r\cos\theta_e}{m}\right|+\left|\frac{\omega_r l}{m}\right|\le\left|\frac{v_r}{m}\right|+\left|\frac{\omega_r l}{m}\right|\le 1-\frac{l\varepsilon}{m} \tag{14}$$

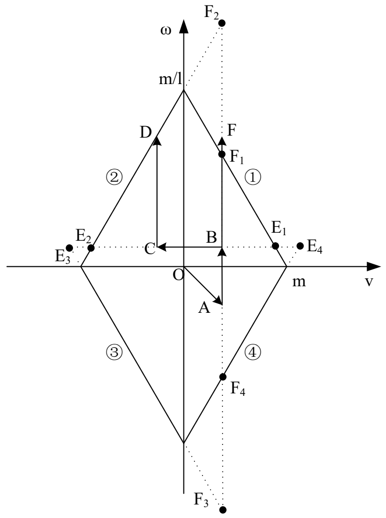


Fig. 3. Vector method design controller

Equation (14) shows that $\overrightarrow{OA}$ satisfy the rhombic input constraints, without loss of generality, we represent $\overrightarrow{OA}$ as shown in Fig.3, and because the length of $\overrightarrow{AB}$ is proportional to k_y, we can definitely find a k_y to make $\overrightarrow{AB}$ within the rhombic input constraints. Similarly, we can also find suitable k_x and k_θ, so $\overrightarrow{BC}$ and $\overrightarrow{CD}$ can meet the rhombic constraints respectively. Obviously, the controller $\overrightarrow{OD}$ will definitely meet the rhombic input constraints.

Since the requirement for k_y is $\dot{k}_y\ge 0$, we intuitively thought of designing if from Lyapunov function $V(t)$

$$V(t)=\frac{1}{2}\left(x_e^2+y_e^2+\frac{\theta_e^2}{k_y}\right) \tag{15}$$

Let k_y be

$$k_y = \frac{\lambda\varepsilon}{m\sqrt{2V(t)+\mu^2}} \tag{16}$$

Where λ and μ are constants, $0<\lambda<1$ and $\mu>0$.
According to (15) (16), we can get

$$k_y = \frac{-m\theta_e^2 + \sqrt{m^2\theta_e^4 + 4\lambda^2\varepsilon^2(x_e^2+y_e^2+\mu^2)}}{2m(x_e^2+y_e^2+\mu^2)} > 0 \tag{17}$$

$$\dot{k}_y = \frac{2k_x k_y^2 x_e^2 + 2k_\theta k_y \theta_e^2}{2k_y(x_e^2+y_e^2+\mu^2)+\theta_e^2} \tag{18}$$

If $k_x>0$ and $k_\theta>0$, then according to equations (17) and (18), $\dot{k}_y<0$ can be derived, and further from (15) we can get

$$\underline{k}_y = \frac{\lambda\varepsilon}{m\sqrt{2V(0)+\mu^2}} \le k_y \le \frac{\lambda\varepsilon}{m\mu} = \overline{k}_y \tag{19}$$

In this way, the vector $\overrightarrow{OB}$ can be expressed as:

$$\overrightarrow{OB} = \overrightarrow{OA} + \overrightarrow{AB} = (v_r\cos\theta_e, \omega_r + k_y v_r y_e \frac{\sin\theta_e}{\theta_e})^{\mathrm{T}} \tag{20}$$

Similar to $\overrightarrow{OA}$ satisfy the rhombic input constraints, we can easily verify that $\overrightarrow{OB}$ satisfies the rhombic input constraints through formulas (14), (15) and (16). Since $k_x, k_\theta>0$, so the directions of the vectors $\overrightarrow{BC}$ and $\overrightarrow{CD}$ are determined by the signs of x_e and θ_e, In order to expand the input field as much as possible while meeting the rhombic input constraints. First, we need to determine the triangle area $\triangle\mathrm{BEF}$ where the points of C and D are located, as shown in Fig.3, when $x_e<0$ and $\theta_e>0$, we take the constraint segment ② to determine the reference triangle $\triangle\mathrm{BE_2F_2}$, similarly, when $x_e>0$ and $\theta_e>0$, we take the constraint segment ① to determine the reference triangle $\triangle\mathrm{BE_1F_1}$, when $x_e<0$ and $\theta_e<0$, we get the reference triangle $\triangle\mathrm{BE_3F_3}$, and when $x_e<0$ and $\theta_e<0$, we get the reference triangle $\triangle\mathrm{BE_4F_4}$. Through the equations of the four constraint lines and the coordinates of point B, we can easily obtain the coordinates of point E as:

$$E:(\mathrm{sgn}(x_e)(m-\mathrm{sgn}(\theta_e)(\omega_r + k_y v_r y_e\frac{\sin\theta_e}{\theta_e})l), \quad w_r + k_y v_r y_e \frac{\sin\theta_e}{\theta_e}) \tag{21}$$

Similarly, we can get the coordinates of F as:

$$F:(v_r\cos\theta_e, \frac{\mathrm{sgn}(\theta_e)(m-\mathrm{sgn}(x_e)v_r\cos\theta_e)}{l}) \tag{22}$$

Where $\mathrm{sgn}(\cdot)$ is sign function

$$\mathrm{sgn}(x)\begin{cases} |x|/x & x\ne 0 \\ 0 & x=0 \end{cases} \tag{23}$$

Further we can get the expressions of $\overrightarrow{BE}$ and $\overrightarrow{BF}$ as

$$\begin{aligned}\overrightarrow{BE} &= \overrightarrow{OE} - \overrightarrow{OB} \\ &= (\mathrm{sgn}(x_e)(m-\mathrm{sgn}(\theta_e)(\omega_r + k_y v_r y_e \frac{\sin\theta_e}{\theta_e})l) \\ &\quad -v_r\cos\theta_e, 0)^{\mathrm{T}}\end{aligned} \tag{24}$$

$$\begin{aligned}\overrightarrow{BF} &= \overrightarrow{OF} - \overrightarrow{OB} \\ &= (0, \frac{\mathrm{sgn}(\theta_e)(m-\mathrm{sgn}(x_e)v_r\cos\theta_e)}{l} - \omega_r \\ &\quad -k_y v_r y_e \frac{\sin\theta_e}{\theta_e})^{\mathrm{T}}\end{aligned} \tag{25}$$

To design k_x and k_θ, let

$$\begin{aligned}\overrightarrow{BC} &= \frac{\rho(|x_e|)}{2}\overrightarrow{BE} \\ \overrightarrow{CD} &= \frac{\rho(|\theta_e|)}{2}\overrightarrow{BF}\end{aligned} \tag{26}$$

Then, we get from (12)(24)(26), that

$$\begin{aligned}k_x &= \frac{\psi(|x_e|)}{2}(m-\mathrm{sgn}(\theta_e)(\omega_r + k_y v_r y_e\frac{\sin\theta_e}{\theta_e})l \\ &\quad -\mathrm{sgn}(x_e)v_r\cos\theta_e) \\ k_\theta &= \frac{\psi(|\theta_e|)}{2}(\frac{m-\mathrm{sgn}(x_e)v_r\cos\theta_e}{l} - \mathrm{sgn}(\theta_e) \\ &\quad (\omega_r + k_y v_r y_e \frac{\sin\theta_e}{\theta_e}))\end{aligned} \tag{27}$$

By formula (15)(16)(20)and *Lemma2* we can easily get

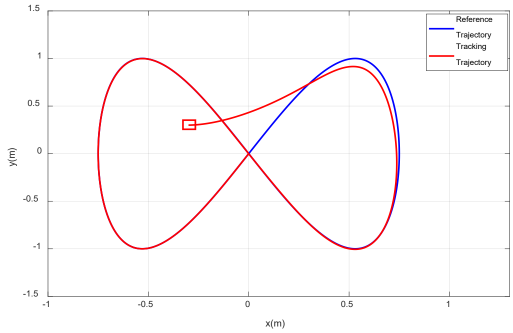
$$\begin{aligned}\underline{k}_x &\le \frac{\alpha(1-\lambda)\varepsilon l}{2} \le k_x \le \beta m \le \overline{k}_x \\ \underline{k}_\theta &\le \frac{\alpha(1-\lambda)\varepsilon}{2} \le k_\theta \le \frac{\beta m}{2} \le \overline{k}_\theta\end{aligned} \tag{28}$$

At this point, the k_x, k_y *and* k_θ meet the two conditions in *Lemma 3*, so the system error will converge to zero. And because of our vector method design the parameters ensure that the parameter control variables v and ω meet the rhombic input constraints too.

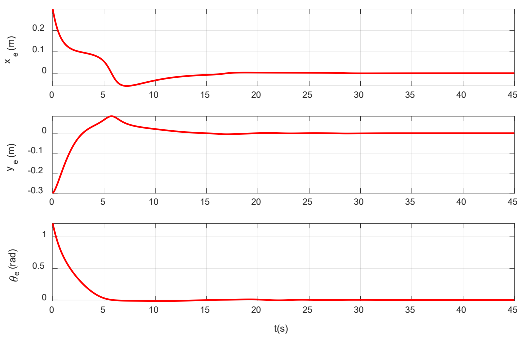
4. Simulation Results

In this section, we simulate and verify the effect of the controller. The maximum speed of the drive wheels is set to $m=0.4m/s$, the wheel spacing is set to $l=0.16m$, For setting some parameters of the controller, we choose $\rho(x) = \tanh(x),\ \varepsilon = 0.1,\ \lambda = 0.99,\ \mu = 0.01$.

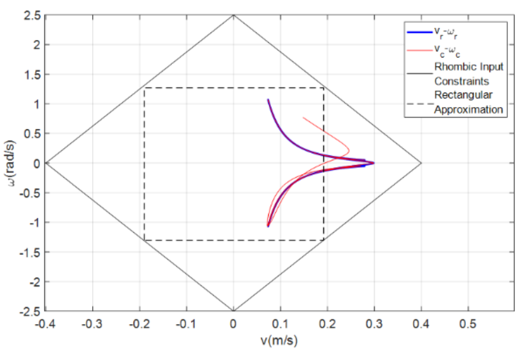
Fig.4(a) shows that DWMRs gradually tracks on the reference trajectory. The tracking errors x_e, y_e, θ_e are gradually converge to zero as shown in Fig.4(b), also we can guarantee the control variables v, ω satisfy the rhombic input constraints through Fig.4(c), and sometimes v can basically reach the bounds of rhombic input constraints.



(a) Tracking reference trajectory



(b) Tracking errors



(c) Input and constraints

Fig.4. Simulation results.

5. Conclusion

The trajectory tracking problem of DWMRs with rhombic input constraints is solved in this paper, compared with the controller based on rectangular input constraints. It can better play the robots' mobility and the controller not only solves the tracking problem, but also solves the stability problem, the simulation shows the controller is effective. Future work will focus on the controller design with uncertainty based on a more complex application environment.

Acknowledgements

This work was supported by the NSFC (61520106010, 61327807, 61134005, 61521091) and the National Basic Research Program of China (973 Program: 2012CB 821200, 2012CB821201).

References

1. Jinhua. Ye , Y. Suzhen , and J. Xiao . Trajectory Tracking Control of WMR Based on Sliding Mode Disturbance Observer with Unknown Skidding and Slipping. *2017 2nd International Conference on Cybernetics, Robotics and Control (CRC) IEEE*, 2018.
2. Hadi, Nabil H , and K. K. Younus . Path tracking and backstepping control for a wheeled mobile robots (WMR) in a slipping environment. *IOP Conference Series Materials ence and Engineering* 671(2020):012005.
3. Inoue, Roberto S. , et al. Robust recursive linear quadratic regulator for wheeled mobile robots based on optical motion capture cameras. *Asian Journal of Control (2019)*.
4. Chen, Yung Hsiang . A Novel Fuzzy Trajectory Tracking Control Design for Wheeled Mobile Robots. *Advances In Robotics & Mechanical Engineering 2(2020)*.
5. J. Han, From PID to Active Disturbance Rejection Control, *IEEE Transactions on Industrial Electronics*, vol. 56, no. 3, pp. 900-906, March 2009.
6. Y. X. Su and C. H. Zheng, Global asymptotic stabilization and tracking of wheeled mobile robots with actuator saturation, in *Proc. IEEE Int. Conf. Robots. Biomim.*, Dec. 2010, pp. 345–350..
7. X. Chen, Y. Jia and F. Matsuno, Tracking Control for Differential-Drive Mobile Robots With Diamond-Shaped Input Constraints, *IEEE Transactions on Control Systems Technology*, vol. 22, no. 5, pp. 1999-2006, Sept. 2014, doi: 10.1109/TCST.2013.2296900.
8. H. K. Khalil, Nonlinear Systems, 3rd ed. *Upper Saddle River, NJ, USA: Prentice-Hall*, 2002.
9. Blazic S.. A Novel Trajectory-Tracking Control Law for Wheeled Mobile Robots[J]. *Robotics and Autonomous Systems*, 2011, 59(11): 1001-1007

Maneuvering target tracking with improved interactive multiple model algorithm

Meiyu Zhu[1], Guanyu Chen[2], Weicun Zhang[1*]

[1]School of Automation and Electrical Engineering, University of Science and Technology Beijing,

30 Xueyuan Road Haidian District, Beijing, China
E-mail: weicunzhang@ustb.edu.cn
www.ustb.edu.cn

[2]University of Nottingham, NG7 2RD, UK

E-mail: eexgc10@nottingham.ac.uk

Abstract

With the rapid development of artificial intelligence, maneuvering target tracking technology has become a major aspect of scientific and technological research with increased performance requirements. Maneuvering target tracking technology is widely used in military and civilian fields. This paper presents a new maneuvering target tracking framework, which adopts an improved interactive multiple model (IMM) with new weighting algorithm. Some target tracking simulations have been conducted based on MATLAB to verify the effectiveness of the proposed maneuvering target tracking approach.

Keywords: Maneuvering target tracking, interactive multiple model, weighting algorithm.

Introduction

In recent decades, with the rapid development of artificial intelligence, target tracking technology has become a major aspect of scientific and technological research. Maneuvering target tracking technology is widely used in military and civilian fields. For example, in air defense and air traffic control, reliable and accurate tracking of targets is always the main purpose of target tracking system design. However, with the rapid development of modern military weapons and the continuous advancement of aerospace technology, the maneuverability of various maneuvering targets has become more complicated, the motion state is more difficult to estimate, and the target maneuvering methods are complex and diversified, making it more and more difficult to track maneuvering targets. The original maneuvering target tracking technology can no longer meet the needs of modern maneuvering target tracking requirements.

The two key parts of maneuvering target tracking are the maneuvering target tracking model and the adaptive filtering algorithm. Considering that the motion model of

*Corresponding author：W. Zhang, weicunzhang@ustb.edu.cn

the target will also change when the target is maneuvered, any single model will produce larger errors. Also, the tracking process is often accompanied by external interference, system parameter jumps and other problems, these problems will cause the tracking performances to deteriorate. Kalman filtering or multiple model adaptive estimation (MMAE) or [1-6] method is a good idea to solve such kind of problem. This article will focus on weighted multiple model adaptive filtering with applications in maneuvering target tracking.

This article first introduces several commonly used maneuvering target tracking models, and analyzes their advantages, disadvantages, and scope of application through simulation, and proposes directions for improvement. Then introduce and study the filtering algorithm of target tracking, mainly including Kalman filtering algorithm, extended Kalman filtering algorithm, unscented Kalman filtering algorithm, and analysis and summary through simulation. Then this paper proposes a new weighting algorithm based on the traditional MMAE algorithm, and proves its convergence. The simulation proves that the new weighting algorithm has the characteristics of fast convergence speed and small calculation amount. Finally, the new weighting algorithm is combined with the interactive multi-model filtering algorithm to track the maneuvering target. The experimental results show that the new algorithm can track the target well, and compared with the traditional algorithm, the tracking speed is faster and the error is smaller.

1. Interactive multiple model estimation

Suppose target motion can be described by r model, model set $M = \{M_1, M_2, \cdots, M_r\}$

Mj:

$$X(k+1) = F_j(k)X(k) + G_j(k)w_j(k) \tag{1}$$

$$Z(k) = F_j(k)X(k) + v_j(k) \tag{2}$$

$w_j(k)$ and $v_j(k)$ are independent Gaussian noises. The covariance matrixes of $w_j(k$ and $v_j(k)$ are R_j and Q_j, respectively. $u_j(k-1)$ represents the probability of model Mj, the model transition probabilities can be described by

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1r} \\ \vdots & \ddots & \vdots \\ p_{r1} & \cdots & p_{rr} \end{bmatrix} \tag{3}$$

IMM filtering has the following 4 steps:

Step 1: Interaction

$$u_{i|j}(k-1|k-1) = \frac{p_{ij}u_i(k-1)}{c_j} \tag{4}$$

$$c_j = \sum_{i=1}^{r} p_{ij} u_i\,(k-1) \tag{5}$$

$$\hat{X}_{0j}(k-1|k-1) = \sum_{i=1}^{r} \hat{X}_i(k-1|k-1)u_{i|j}(k-1|k-1) \tag{6}$$

$$P_{0j}(k-1|k-1) = \sum_{i=1}^{r}\{P_i(k-1|k-1) + [\hat{X}_i(k-1|k-1) - \hat{X}_{0i}(k-1|k-1)][\hat{X}_i(k-1|k-1) - \hat{X}_{0i}(k-1|k-1)]^T]\}u_{i|j}(k-1|k-1) \tag{7}$$

Step 2: Filtering

$$\hat{X}_j(k|k-1) = F_j(k-1)\hat{X}_{0j}(k-1|k-1) \tag{8}$$

$$P_j(k|k-1) = F_j(k-1)P_{0j}(k-1|k-1)F_j^T(k-1) + \Gamma_j(k-1)Q_j(k-1)\Gamma_j^T(k-1) \tag{9}$$

$$\varepsilon_j = Z(k) - H_j\hat{X}_j(k-1|k-1) \tag{10}$$

$$S_j(k) = H_j(k)P_j(k|k-1)H_j^T(k) + R_j(k) \tag{11}$$

$$K_j(k) = P_j(k|k-1)H_j(k)S_j^{-1}(k) \tag{12}$$

$$\hat{X}_j(k|k) = \hat{X}_j(k|k-1) + K_j(k)\varepsilon_j(k) \tag{13}$$

$$P_j(k|k) = [I - K_j(k)H_j(k)]P_j(k|k-1) \tag{14}$$

Step 3: Model probability updates

$$l_j(0) = \frac{1}{m}; u_j(0) = l_j(0) \tag{15}$$

$$l_j^{'}(k) = 1 + \frac{1}{r}\sum_{q=1}^{r}\left\|\varepsilon_j(q)\right\|^2 \tag{16}$$

$$l_{min}^{'} = \min_i\{l_j^{'}(k)\} \tag{17}$$

$$l_j(k) = \frac{l_{min}^{'}}{l_j^{'}(k)} l_j(k-1) \tag{18}$$

$$u_j(k) = \frac{1}{c} l_j(k)c_j \tag{19}$$

$$c = \sum_{j=1}^{r} l_j\,(k)c_j \tag{20}$$

Step 4: Combination

$$\hat{X}(k|k) = \textstyle\sum_{j=1}^{r} \hat{X}_j(k|k)u_j(k) \tag{21}$$

$$P(k|k) = \textstyle\sum_{j=1}^{r}\{P_j(k|k) + [\hat{X}_j(k|k) - \hat{X}(k|k)]\,[\hat{X}_j(k|k) - \hat{X}(k|k)]^T]\}u_j(k) \tag{22}$$

2. Simulation results

To verify the effectiveness of the proposed maneuvering target tracking based on improved IMM algorithm, many simulations have been conducted with MATLAB platform. Simulation results are satisfactory, details are omitted here due to space limit, just to illustrate some pictures of simulation results. Figure 1 shows plane tracking performance comparisons among Current Statistical (CS) model, IMM algorithm, and improved IMM algorithm; Figure 2 and Figure 3 show tracking performance comparisons in X-direction and Y-direction; Figure 4 shows the weights values of IMM algorithm; Figure 5 shows the weights values of improved IMM algorithm.

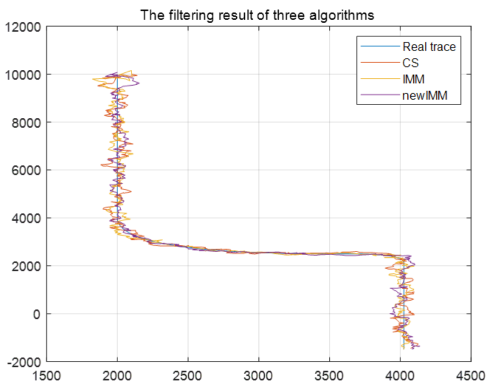


Figure 1 The filtering results of three algoritms

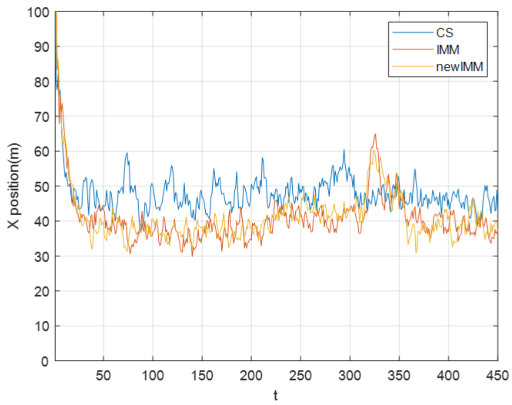


Figure 2 Positions: X direction

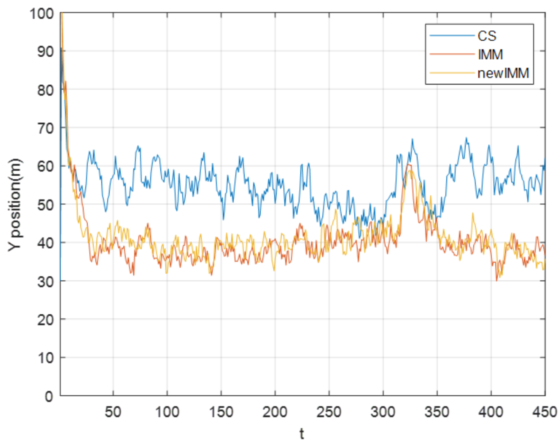
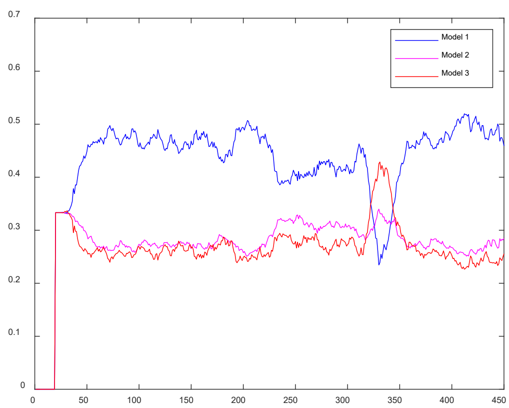


Figure 3 Positions: Y direction

Figure 4 Weight values of traditional IMM algorithm



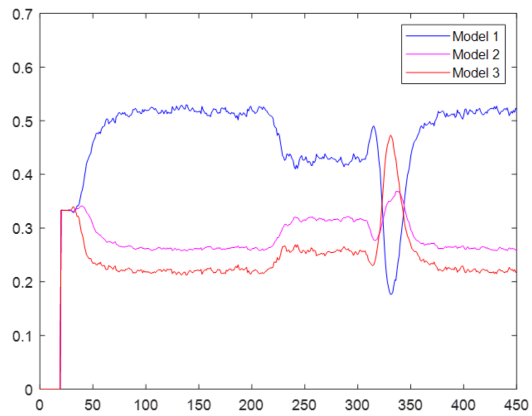


Figure 5 Weight values of improved IMM algorithm

3. Conclusion and future work

This paper presented an improved IMM maneuvering target tracking approach. Simulation results verified the effectiveness of the proposed method including new weighting algorithm. In the future research work, we will focus on model set selection with prescribed performance requirements.

Acknowledgements

This work was partly supported by National Natural Science Foundation of China (No. 61520106010; 61741302).

References

1. D. T. Magill, Optimal adaptive estimation of sampled stochastic processes, IEEE Trans. Automat. Contr., 1965; 10: 434-439
2. D. G. Lainiotis, Partitioning: A unifying framework for adaptive systems-I: Estimation; II: Control, Proc. IEEE, 1976; 64:1126-1143 and 1182-1197
3. R. E. Kalman, "A new approach to linear filtering and prediction problems," Journal of Basic Engineering, vol. 82, no. 1, pp. 35–45, 1960.
4. P. Zarchan and H. Musoff, Fundamentals of Kalman Filtering: a Practical Approach, American Institute of Aeronautics and Astronautics, 4th edition, 2015.
5. X.-R. Li and Y. Bar-Shalom, "Multiple-model estimation with variable structure," IEEE Transactions on Automatic Control, vol. 41, no. 4, pp. 478–493, 1996.
6. W. Zhang, S. Wang, and Y. Zhang, Multiple-model adaptive estimation with a new weighting algorithm, Complexity, 2018:1-11,2018

Encapsulated Agents of Hybrid Order Discrete Dynamics

Yunzhong Song†
School of Electrical Engineering and Automation, Henan Polytechnic University, 2001 Century Avenue Jiaozuo, 454003, P.R.China

Weicun Zhang
School of Automation and Electrical Engineering, University of Science and Technology Beijing, 30Xueyuan Road Beijing, 100083, P.R. China

Fengzhi Dai
School of Electronic Information and Automation, Tianjin University of Science and Technology, 1038 Dagu Nanlu Tianjin, 300222, P.R. China

Huimin Xiao
School of Computer and Information Engineering, Henan University of Economics and Law, 180 Jinshui Donglu Zhengzhou, 450046, P.R.China

Shumin Fei
School of Automation, South East University, 2 Sipai Lou Nanjing, 210096, P.R.China
E-mail:songhpu@126.com,weichunzhang@263.net, daifz@tust.edu.cn, xiaohm@huel.edu.cn, smfei@seu.edu.cn

Abstract

This paper is targeting to investigate the encapsulated cell realization of the hybrid order agents in discrete time, where agents are composed by two different dynamic order ones, one is the first order and the other is the second order, the work is the counter of its continuous ones of the encapsulated agents. To be further, the first order agents are assigned at low speed, and the second order agents are selected at comparatively higher speed. To overcome the obstacle brought by difference of the sampling rates, the lifting technology is taken into use to analyze the encapsulated system. Theoretical analysis and simulation results are made available for the further reference.

Keywords: encapsulated agents, hybrid order, discrete dynamics, lifting techniques, consensus

1. Introduction

Multi-agent systems have been extensively explored these days for their theoretical values and potential prospects in real engineering applications [1-9]. Among them, the heterogeneous agents were paid much more attention, partially because of their widely distributed in industries, social activities and biological groups; one of the striking point of heterogeneities of the multi-agents is their difference in dynamics, especially of dynamic orders. The hybrid order dynamics of multi-agents were firstly put forward and solved by match principle or

† Corresponding author

match equivalence, i.e. protocols should be different for agents with different dynamic orders [3,6].

However, it is not so wise for us just to stop at the low level of the multi-agents, something new should be added to push this research further. Is it possible to pack the agents with hybrid order together? Does that bring us much more promising solutions to the real engineering problems? These problems are full of the meaning. One of the work of this direction has been done, it is for hybrid order agents packed in an encapsulated cell. The dynamics for them is described in continues time [3]. In this note, discrete time agents of the hybrid order packed in an encapsulated cell will be touched upon here.

The following of this note will be arranged as follows: Problem description will be provided at first, then followed by theoretical analysis and simulations, some comments will also be available in this part; the last part will be conclusions. Acknowledgements will be acknowledged for the fund support.

2. Problem Description

The dynamics of the encapsulated agents are described in the following equations:

$$\begin{cases} X_j^i(k+1) = f(X_j^i(k), P) + U_j^i(k); \\ Y^i(k) = x_1^i(k); \end{cases} \tag{1}$$

Here, in Equation (1), index i is used to mark the layer of the agents, and index j is used to mark the order of the agents of the layer i; $u(k)$ is the controlled input for the agents, it comes from three different places, one is from the agent itself, the other is from the other agents of the same layer, and the third is from the other layer agents; the output signal is extracted from the first order agents; the capital x and u is employed to express the packed state variables, i.e. to the agent systems combined by the first and second order dynamics, both position sub-states and velocity sub-states are covered by the capital x, the capital letter p is used to denote the parameters of the agent systems; and the controlled input can be inferred correspondingly.

In case of simplicity and without loss of generality, we assume the multi-agents system is composed by two layers; each layer includes two different agents, among them one is the first order and the other is the second order. So, the Equation (1) can be refined as follows:

$$\begin{cases} \dfrac{x_1^i(k+1) - x_1^i(k)}{n_0 T_0} = u_1^i(k); \\ \dfrac{x_2^i(k+1) - x_2^i(k)}{T_0} = v_2^i(k); \\ \dfrac{v_2^i(k+1) - v_2^i(k)}{T_0} = u_2^i(k); \\ y^i(k) = x_1^i(k); \end{cases} \tag{2}$$

Corresponding to Equation (1),

$$\begin{cases} X_1^i(k) = x_1^i(k); \\ X_2^i(k) = \begin{bmatrix} x_2^i(k) \\ v_2^i(k) \end{bmatrix}; \end{cases} \tag{3}$$

As to the controlled inputs, both for the first order agents and the second agents, are enforced by three different parts, one is from the individual agent itself, the another is from the same layer of the different order agents, and the third is from the other layer agents. And in this way, we have system level of the whole agents, and sub-level of the layer, followed by the atom-level agents with different dynamic orders.

Notation 1: Parameters of n_0 and T_0 is assigned to represent the times of the sampling period and standard sampling period, respectively; since the first order agent system only has one integrator, and the second order agent system has two integrators, the information updating rate for the second is much more slower than the first agent system; and in this kind of sense, we chose T_0 as 1 and n_o as 4.

3. Theoretical Analysis and Simulation Results

This part will contribute for theoretical analysis and simulation results, some comments are also offered to trigger the prospective topics.

3.1. *Theoretical analysis*

Take two homogeneous encapsulated cell agent systems as an example, each one of them is composed of one of the first order agent and one of the second order agent, to be specific, for the first cell encapsulated agent system, it can be expressed as:

$$\begin{cases} x_1^1(k+1) = x_1^1(k) + 4u_1^1(k); \\ x_2^1(k+1) = x_2^1(k) + v_2^1(k); \\ v_2^1(k+1) = v_2^1(k) + u_2^1(k); \\ y^1(k) = x_1^1(k); \end{cases} \tag{4}$$

Correspondingly, the second cell encapsulated agent system can be expressed as:

$$\begin{cases} x_1^2(k+1) = x_1^2(k) + 4u_1^2(k); \\ x_2^2(k+1) = x_2^2(k) + v_2^2(k); \\ v_2^2(k+1) = v_2^2(k) + u_2^2(k); \\ y^2(k) = x_1^2(k); \end{cases} \tag{5}$$

Notation 2: The dynamic of the first order agent in cell encapsulated agent system is only influenced by its own dynamics and the dynamics of the second order agent of the same cell. To be different, the dynamic of the second order agent in cell encapsulated agent system can be altered by its own dynamic and also the dynamic of the first order agent, which is assigned to the same cell, besides that, the output of the other cell agent system can also contribute to the dynamics of the second order agent.

Notation 3: Consider the different sampling speed, the signal of the second order agent feed to the first order agent is the 4 times period delayed signal, that embodies the lifting technique.

For the first cell encapsulated agent system, the controlled input can be denoted as follows:

$$\begin{cases} x_1^1(k+1) = x_1^1(k) + 4u_1^1(k); \\ u_1^1(k) = u_{11}^1(k) + u_{12}^1(k); \\ u_{11}^1(k) = -\alpha_1 x_1^1(k); \\ u_{12}^1(k) = \beta_1[x_2^1(k-4) - x_1^1(k)]; \\ x_2^1(k+1) = x_2^1(k) + v_2^1(k); \\ v_2^1(k+1) = v_2^1(k) + u_2^1(k); \\ u_2^1(k) = u_{21}^1(k) + u_{22}^1(k) + u_{23}^1(k); \\ u_{21}^1(k) = -\alpha_2 x_2^1(k) - \alpha_3 v_2^1(k); \\ u_{22}^1(k) = \beta_2[x_1^1(k) - x_2^1(k-4)]; \\ u_{23}^1(k) = \beta_3[x_1^2(k) - x_1^1(k)]; \\ y^1(k) = x_1^1(k); \end{cases} \tag{6}$$

And the second cell encapsulated agent system, the controlled input can be denoted as follows:

$$\begin{cases} x_1^2(k+1) = x_1^2(k) + 4u_1^2(k); \\ u_1^2(k) = u_{11}^2(k) + u_{12}^2(k); \\ u_{11}^2(k) = -\alpha_1 x_1^2(k); \\ u_{12}^2(k) = \beta_1[x_2^2(k-4) - x_1^2(k)]; \\ x_2^2(k+1) = x_2^2(k) + v_2^2(k); \\ v_2^2(k+1) = v_2^2(k) + u_2^2(k); \\ u_2^2(k) = u_{21}^2(k) + u_{22}^2(k) + u_{23}^2(k); \\ u_{21}^2(k) = -\alpha_2 x_2^2(k) - \alpha_3 v_2^2(k); \\ u_{22}^2(k) = \beta_2[x_1^2(k) - x_2^2(k-4)]; \\ u_{23}^2(k) = \beta_3[x_1^1(k) - x_1^2(k)]; \\ y^2(k) = x_1^2(k); \end{cases} \tag{7}$$

Parameters of the whole system can be listed as follows:

$$\begin{cases} \alpha_1 = \dfrac{1}{8}; \alpha_2 = \dfrac{1}{2}; \alpha_3 = \dfrac{5}{3}; \\ \beta_1 = \dfrac{1}{8}; \beta_2 = \dfrac{1}{2}; \beta_3 = \dfrac{1}{16}; \end{cases} \tag{8}$$

Notation 4: The parameters are selected to guarantee the stability of the whole system.

3.2. *Simulation results*

When the initial values are provided, under the simulation environment like SIMULINK, the Simulink results are easily to be made possible. Figure 1 is the simulation results of the encapsulated system:

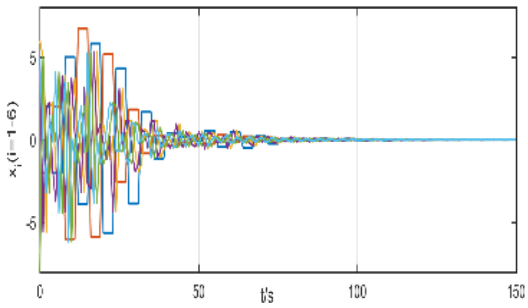


Fig.1 State curves of the system

From the simulation results, it could be seen clearly that all of the states of the system are driven to the stable equilibrium.

Notation 5: The different equilibrium values can be made possible if we assign different working modes of the sub-system.

3.3. *Some comments*

In order to make the work tight and complete, two comments are offered here.

Comment 1: Agent system is kicked off by studying the same order ones, especially of the first order agents and the second order agents, after then the hybrid order agent systems are initiated [3,6] and references there in followed by colleagues of the control field. Is it possible for us to come back to the same agents but with much more higher viewpoints?

Comment 2: Encapsulated cell agent systems are the right way for us to dig deep on the road of the same at low level, the different to deepen of the same, and come to the same in a higher level.

4. Conclusion

This note contributes to the background introduction, theoretical analysis and simulation research of the encapsulated cell agents with discrete dynamics. The value of this note is obvious in several aspects:

(1) It reiterates the importance of the layer of the system, now it is clear that the encapsulated cell agents are divided into three different levels, from the top to the down, they are the system level, the cell level, and the agent level;

(2) According to our knowledge, it is first discussion about the sampling speed of the hybrid order agents, lifting techniques are borrowed to design the controller;

(3) For the complex engineering system, the sub-systems are always intertwined with each other, hierarchy should be considered of course, and in this sense, our work is meaningful;

Besides, the complexity is the essential characteristics, time-delay, coupling, fragility, robust property et al, all of them should be taken into count [10-12], it is a long way for us to explore.

Acknowledgements

This work is partially supported by NSFC Grant (61340041 and 61374079) and the Project-sponsored by SRF for ROCS, SEM to Yunzhong Song as well as Natural Science Fund of Henan Province (182300410112).

References

1. R. Olfati, R. Murray, Consensus problems in networks of agents with switching topology and time delays, *IEEE Transactions on Automatic Control*, **49**(2004) 1520-1533.
2. Y.Song, F. Dai, H. Xiao and S. Fei, No free lunch principle in agent swarm systems: one Case Study, *Journal of Advances in Artificial Life Robotics,* **1** (2020) 33-37.
3. Y.Song, F. Dai, H. Xiao, *On the complexity of encapsulated agent cell*, Proceedings of 2019 Chinese Intelligent Systems Conference, (2019) 24-30
4. Y. Song, Z. Fu, F. Wang, Socialized multi-agent system rendezvous via networks of networks, *Journal of Robotics, Networking and Artificial Life*, **2** (2016) 136-139.
5. Y. Song, Z. Fu, F. Wang, Flock guiding of hybrid agents via root block, *Journal of Robotics, Networking and Artificial Life*, **5** (2019) 245-248
6. Y. Song, Consensus of agents with mixed linear discrete dynamics, *International Journal of Control, Automation and Systems*, **14** (2016) 1139-1143.
7. Y. Song, W. Zhao, Multi-agent system rendezvous via refined social system and individual roles, *WSEAS Transactions on Systems and Control*, **9**(2014),526-532
8. Y. Hong, J. Hu, L. Gao, Tracking control for multi-agent consensus with an active leader and variable topology, *Automatica*, **42**(2006) 1177-1182.
9. Y. Hong, G. Chen, L. Bushenell, Distributed observers design for leader-following control of multi-agent networks, *Automatica*, **44**(2008) 846-850.
10. Y. Jia, Robust control with decoupling performance for steering and traction of 4WS vehicles under velocity-varying motion, *IEEE Transactions on Control Systems Technology*, **8**(2000) 554-569.
11. Y. Jia, Alternative proofs for improved LMI representations for the analysis and the design of continuous-time systems with polytopic type uncertainty: a predictive approach, *IEEE Transactions on Automatic Control*, **48**(2003) 1413-1416
12. Y. Jia, General solution to diagonal model matching control of multi-output-delay systems and its applications in adaptive scheme, *Progress in Natural Science*, **19**(2009), 79-90.

Kinematic modeling and Simulation of humanoid dual-arm robot

Jiwu Wang
School of Mechanical and Electronic Engineering, Beijing jiaotong University
Beijing, Haidian District, China

Junxiang Xu
School of Mechanical and Electronic Engineering, Beijing Jiaotong University
Beijing, Haidian District, China
E-mail: jwwang@bjtu.edu.cn, 19121251@bjtu.edu.cn
www.bjtu.edu.cn

Abstract

In recent years, dual-arm robots have attracted more and more attention due to their advantages such as strong cooperation ability and high flexibility. With the improvement of real-time requirement of dual-arm cooperation, the inverse kinematics solution of robot becomes a key problem to be solved urgently. In order to solve the time-consuming problem of inverse kinematics of robot arm, a closed inverse kinematics solution algorithm for humanoid dual-arm robot was proposed. The effectiveness of the algorithm was verified by simulation.

Keywords: dual-arm robots ; real-time ; inverse kinematics; simulation.

1. Introduction

Humanoid dual-arm robot is a kind of bionic robot designed by imitating the shape, structure and function of human body. It can work with both hands cooperatively just like human. The inverse kinematics of manipulator mainly includes geometric method, analytical method and numerical method. The geometric method is a special case of analytic method in some cases, and its applicability is weak[1,2]. Analytical inverse kinematics of the manipulator can efficiently obtain all the inverse solutions of the manipulator in the desired position, but the manipulator must satisfy the Piper criterion[3]. The numerical method has no special requirements for the joint number and structure of the manipulator, but it needs to be solved through continuous iteration, which not only takes a long time, but the average calculation error is also 10 times that of the analytical method[4,5].

Because it takes a long time to meet the accuracy requirement and solve the problem of manipulator inverse kinematics solution, the analytical solution is adopted. A closed inverse kinematics algorithm is proposed for the humanoid robot arms. The algorithm is used for simulation analysis. Simulation results show that the algorithm is effective.

2. Kinematic modeling of dual-arm robot

2.1. *Forward kinematics*

The robot model in this paper is shown in Figure 1. Each arm of the robot has six degrees of freedom. The three axes of the shoulder joint intersect at one point. Because this case conforms to the Piper criterion, there are analytic solutions. Taking the shoulder joint as the basic coordinate frame[6], figure 1 also shows the link coordinate frames of the right arm and its D-H parameters. The left arm and the right arm are identical.

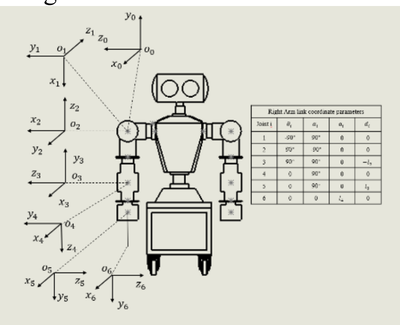


Fig. 1. Link coordinate frames of the right arm and its D-H parameters.

The position and orientation of the end-effector can be obtained by chain-multiplying the 6 link-transformation matrices together to obtain the spatial displacement of the 6th coordinate frame with respect to the base coordinate frame:

$$
\begin{aligned}
&{}^{0}\mathrm{T}_6=\prod_{i=1}^{6}{}^{i-1}\mathrm{A}_i={}^{0}\mathrm{A}_1\,{}^{1}\mathrm{A}_2\,{}^{2}\mathrm{A}_3\,{}^{3}\mathrm{A}_4\,{}^{4}\mathrm{A}_5\,{}^{5}\mathrm{A}_6\\
&=\begin{bmatrix} \mathrm{x}_6 & \mathrm{y}_6 & \mathrm{z}_6 & \mathrm{p}_6 \\ 0 & 0 & 0 & 1 \end{bmatrix}=\begin{bmatrix} \mathrm{n} & \mathrm{s} & \mathrm{a} & \mathrm{p} \\ 0 & 0 & 0 & 1 \end{bmatrix}
\end{aligned}
\tag{1}
$$

where x_i, y_i and z_i represent the unit vectors along the principal axes of the coordinate frame i, ${}^{i-1}\mathrm{A}_i$ is a general link transformation matrix, relating the ith coordinate frame to the (i−1)th coordinate frame, and [n, s, a, p] represents the normal vector, the sliding vector, the approach vector and the position vector of the hand respectively.

2.2. *Inverse kinematics*

We take the inverse of both sides of Eq. (1). The new matrix T' is

$$
\begin{aligned}
&\mathrm{T}'=\begin{bmatrix} \mathrm{n} & \mathrm{s} & \mathrm{a} & \mathrm{p} \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1}=\begin{bmatrix} \mathrm{n}' & \mathrm{s}' & \mathrm{a}' & \mathrm{p}' \\ 0 & 0 & 0 & 1 \end{bmatrix}\\
&={}^{6}\mathrm{A}_5\,{}^{5}\mathrm{A}_4\,{}^{4}\mathrm{A}_3\,{}^{3}\mathrm{A}_2\,{}^{2}\mathrm{A}_1\,{}^{1}\mathrm{A}_0={}^{6}\mathrm{A}_0
\end{aligned}
\tag{2}
$$

We can obtain an equation that we label as G_2 equation by moving the link transformation matrix ${}^{5}\mathrm{A}_6$ to the left-hand side of Eq.(2).

$$
{}^{5}\mathrm{A}_6\mathrm{T}' = {}^{5}\mathrm{A}_4\,{}^{4}\mathrm{A}_3\,{}^{3}\mathrm{A}_2\,{}^{2}\mathrm{A}_1\,{}^{1}\mathrm{A}_0 \tag{3}
$$

The left-hand side of E_2 is

$$
E_{L2}=\begin{bmatrix} \cdots & \cdots & \cdots & C_6\left(p_x'+l_{L4}\right)-S_6p_y' \\ \cdots & \cdots & \cdots & S_6\left(p_x'+l_{L4}\right)+C_6p_y' \\ \cdots & \cdots & \cdots & p_z' \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}
$$

And the right-hand side of E_2 is

$$
E_{R2}=\begin{bmatrix} \cdots & \cdots & \cdots & S_4C_5l_{L2} \\ \cdots & \cdots & \cdots & -C_4l_{L2}-l_{L3} \\ \cdots & \cdots & \cdots & S_4S_5l_{L2} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5}
$$

where $S_i\equiv\sin\theta_i, C_i\equiv\cos\theta_i, S_{ij}\equiv\sin\left(\theta_i+\theta_j\right), C_i\equiv\cos\left(\theta_i+\theta_j\right)$, and l_{Li} are geometric link parameters in Fig. 1. By comparing the elements (1,4), (2,4) and (3,4) of the E_{L2} and E_{R2}, we can obtain C_4 and then S_4 from C_4, and from which we can find the joint solution for θ_4. we can solve θ_5 when we have solved θ_4 .Similarly, we can solve θ_6 when we have solved θ_4 and θ_5.

$$
\begin{aligned}
&C_4=\frac{\left(p_x'+l_{L4}\right)^2+p_y'^2+p_z'^2-l_{L2}^2-l_{L3}^2}{2l_{A2}l_{A3}}\\
&\theta_4=a\tan2\left(\pm\sqrt{1-C_4^2},C_4\right)\\
&S_5=p_z'\big/\left(S_4l_{L2}\right)\\
&\theta_5=a\tan2\left(S_5,\pm\sqrt{1-S_5^2}\right)\\
&\theta_6=a\tan2\left(-\left(C_4l_{L2}+l_{L3}\right),S_4C_5l_{L2}\right)-a\tan2\left(p_y',p_x'+l_{L4}\right)
\end{aligned}
\tag{6}
$$

To solve for the remaining joint angles, we can move the link transformation matrix ${}^{3}\mathrm{A}_4\,{}^{4}\mathrm{A}_5$ to the left-hand side of Eq.(3). And we can obtain an equation that we label as E_4 equation.

$$
{}^{3}\mathrm{A}_4\,{}^{4}\mathrm{A}_5\,{}^{5}\mathrm{A}_6\mathrm{T}' = {}^{3}\mathrm{A}_2\,{}^{2}\mathrm{A}_1\,{}^{1}\mathrm{A}_0 \tag{7}
$$

The left-hand side of E_4 is

$$
E_{L4}=\begin{bmatrix} a11 & a12 & a13 & a14 \\ a21 & a22 & a23 & a24 \\ a31 & a32 & a33 & a34 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8}
$$

And the right-hand side of E_4 is

$$E_{R4}=\begin{bmatrix} C_1C_2C_3-S_1S_3 & C_1S_3+S_1C_2S_3 & S_2C_3 & 0 \\ -C_1S_2 & -S_1S_2 & C_2 & l_{L2} \\ S_1C_3+C_1C_2S_3 & S_1C_2S_3-C_1C_3 & S_2S_3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{9}$$

By comparing the element (2,3) of E_{L4} and E_{R4} ,we get C_2 and then S_2 , and the joint solution θ_2 from them,

$$\begin{aligned} &C_2=a23=a_z^{'}S_4S_5-a_y^{'}\left(C_4C_6+S_4C_5S_6\right)-a_x^{'}\left(C_4S_6+S_4C_5C_6\right) \\ &\theta_2=a\tan 2\left(\pm\sqrt{1-C_2^2},C_2\right) \end{aligned} \tag{10}$$

By comparing the elements [(1,3) (3,3)] and [(2,1) (2,2)], we can find the joint solution θ_3 and θ_1 ,

$$\begin{aligned} &a13=a_x^{'}\left(C_4C_5C_6+S_4S_6\right)+C_4S_5a_z^{'}+\left(S_4C_6-C_4C_5S_6\right)a_y^{'} \\ &a33=S_5C_6a_x^{'}-C_5a_z^{'}-S_5S_6a_y^{'} \\ &\theta_3=a\tan 2\left(a33,a13\right) \\ &a21=\left(S_4C_5C_6-C_4S_6\right)n_x^{'}+S_4S_5n_z^{'}-\left(S_4C_5S_6+C_4C_6\right)n_y^{'} \\ &a22=\left(S_4C_5C_6-C_4S_6\right)s_x^{'}+S_4S_5s_z^{'}-\left(S_4C_5S_6+C_4C_6\right)s_y^{'} \\ &\theta_1=a\tan 2\left(-a22,-a21\right) \end{aligned} \tag{11}$$

The solution of the six joints is obtained by the above procedure. The left arm is exactly the same as the right arm for the solution.

3. Dual-arm robot modeling based on Simscape

3.1. *Model building*

In this paper, the dual-arm robot simulation model mainly includes three parts, namely the trajectory planning joint Angle input module, the robot body module and the output measurement module. Its Simscape Multibody model is shown in Figure 2.

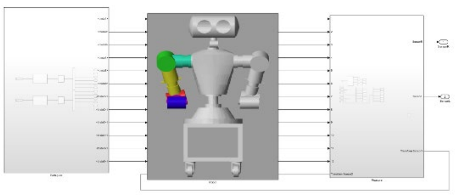


Fig. 2. Dual-arm robot simulation model

3.2. *Establishment of joint Angle input module for trajectory planning*

We plan the trajectory of space in Cartesian coordinates. After solving the inverse kinematics function, we can obtain the angles of each joint. Then we output these angles to the various joints of the arm. The joint angle input model for trajectory planning is shown in Figure 3.

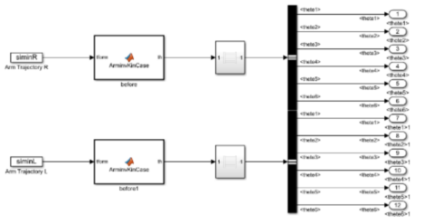


Fig. 3. The joint angle input model for trajectory.

3.3. *Establishment of robot ontology module*

Figure 4 shows the robot ontology module. It includes world coordinate frame and robot ontology. Each arm has six degrees of freedom, including three degrees of freedom for shoulder joint, one degree of freedom for elbow joint and two degrees of freedom for wrist joint.

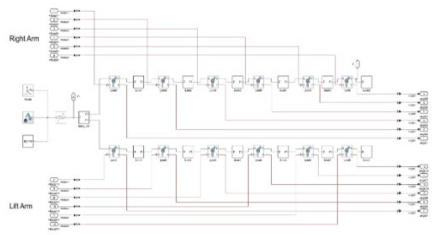


Fig. 4. The robot ontology module.

3.4. *Establishment of output measurement module*

The output measurement module shown in Figure 5 can measure the rotation Angle of each degree of freedom. Through this module, we can also obtain the displacement of the terminal coordinate frame of the manipulator relative to the base coordinate frame.

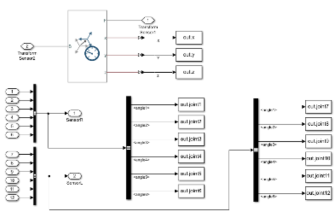
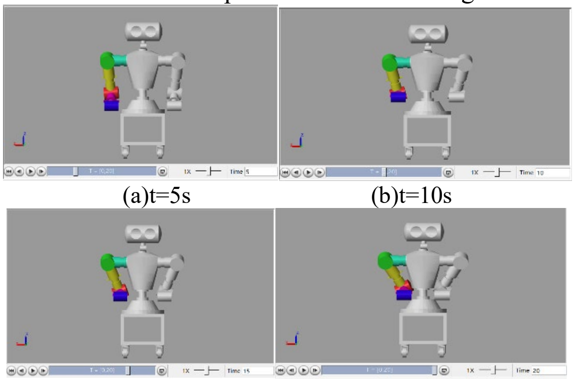


Fig. 5. The output measurement module.

4. Simulation verification and analysis

4.1. *Simulation process*

The robot first raised their arms and then approached each other. The entire process is shown in Figure 6.



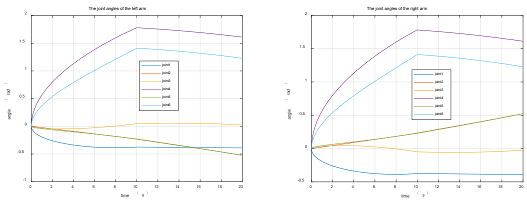
(a)t=5s (b)t=10s

(c)t=15s (d)t=20s

Fig. 6. Simulation diagram of each moment.

4.2. *The motion angles of each joint*

Figure 7 shows the motion angles of each joint of the manipulator. At the initial moment, the Angle of each joint is 0 radian, and at about 10 seconds, the robotic arms begin to approach each other. The Angle of each joint is continuous in the whole process, which proves that the inverse kinematics solution algorithm in this paper can solve the analytic solution effectively.



(a) Left arm's angles (b) Right arm's angles

Fig. 7. The motion angles of each joint of the left and right arms.

5. Conclusion

In this paper, the kinematics modeling of dual-arm robot is carried out. We present an analytical solution method for 6 DOF dual-arm robot for each arm. Simulation results show that the proposed method is effective. In the future, it will be applied to the arm trajectory planning of the robot.

Acknowledgements

This work was supported by the Key Research and Development Program (No.M18GY300021).

References

1. Ian S. Fischer. A geometric method for determining joint rotations in the inverse kinematics of robotic manipulators. 2000, 17(2):107-117.
2. Samer Tahya, M. Moghavvemi, Haider A.F.Mohamed. Geometrical approach of planar hyper-redundant manipulators: Inverse kinematics, path planning and workspace. Simulation Modelling Practice and Theory. 2010(1).
3. Buss S R. Introduction to inverse kinematics with Jacobian transpose, pseudoinverse and damped least squares methods, *J.* IEEE Transactions on Robotics and Automation, 2004.
4. Momani S,Abo-Hammour Z S,Alsmadi O M K. Solution of inverse kinematics problem using genetic algorithms. Applied Mathematics&Informaion Scienced . 2016
5. Mostafa Bayati. Using cuckoo optimization algorithm and imperialist competitive algorithm to solve inverse kinematics problem for numerical control of robotic manipulators. 2015, 229(5):375-387.
6. Ali MA, Park HA, Lee CSG. Closed-form inverse kinematic joint solution for humanoid robots. In: IEEE/RSJ 2010 international conference intelligent robotics system IROS 2010-coference proceeding.;2010. pp. 704-9

Authors Introduction

Dr. Jiwu Wang

He is an associate professor, Beijing Jiaotong University. His research interests are Intelligent Robot, Machine Vision, and Image Processing.

Mr. Junxiang xu

He is a postgraduate in Beijing Jiaotong University. His research interest is robot modeling and simulation.

Kinematics analysis and simulation of 6R Robot Based on Matlab/Simulink

Jiwu Wang
School of Mechanical and Electronic Engineering, Beijing jiaotong University
Beijing, Haidian District, China

Shuo Han
School of Mechanical and Electronic Engineering, Beijing jiaotong University
Beijing, Haidian District, China
E-mail: jwwang@bjtu.edu.cn, 19140566@bjtu.edu.cn
www.bjtu.edu.cn

Abstract

The 6R robot can imitate the human arm to complete some target grabbing tasks, so the kinematics analysis of the robot is significant in scientific research and practical application. In this paper, a kinematics solution method of 6R robot based on analytic method is introduced, which is faster and more accurate in solution than the numerical method. Then the trajectory of the end effector is planned by using the quintic polynomial method,in this way, there are no sudden changes in the speed of the end effector of the robot, and the operation is more stable.Furthermore, the accuracy of the kinematics solution method is verified and the motion trajectory of the manipulator is simulated by Matlab. At last ,the visualization of the robot kinematics model was realized based on the Simulink, and the kinematics simulation control system was established.

Keywords: 6-DOF robot ; forward kinematics ; inverse kinematics ; simulation.

1. Introduction

The industrial robot has played an important role in improving working conditions and production efficiency, and some algorithms and programs can control the robot to complete some specified tasks in a certain way.So that,the 6R robot is widely used in machining, electronic welding, industrial handling and other industries because of its flexibility and maneuver ability.

The basic characteristic of the 6R robot is the kinematics characteristic, so the kinematics analysis of the robot is significant in scientific research and practical application. In this paper,the mathematical model of the 6R robot is established by D-H algorithm, and the forward and inverse kinematics calculation is completed,so that the angle of each joint can be found out when the robot grabs the target. Then the motion trajectory of the robot was simulated in the joint space, and the smooth curves of angular displacement,angular velocity were obtained.

Furthermore, the rationality of the D-H parameters and Inverse kinematics results of the robot were verified by MATLAB, and the visualization of the robot kinematics model was realized based on the Simulink, and the kinematics simulation control system was established.

2. The Kinematics Analysis of the 6R Robot

2.1. Establish D-H coordinate system

In this paper,the ABB-irb120 robot that shown in figure 1 is used as an example of 6R robot in our discussion.

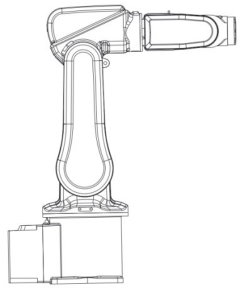


Fig. 1. The model of the 6R robot

We used the Denavit-Hartenberg (D-H) matrix representation for each link to describe the rotational and translational relationship between adjacent links. The coordinate system of each connected link is established , as shown in figure 2,and the D-H parameters were obtained, as shown in the table 1.

The coordinate system {0} in figure 2 is a defined reference coordinate system, which is fixed on the base and usually coincides with the origin of the coordinate system {1}.

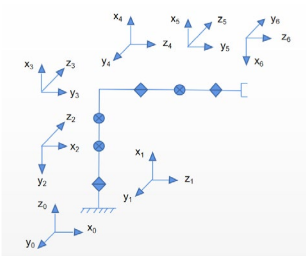


Fig. 2. Coordinate system of each link of the robot

Tab.1. D-H Parameter Table

Joint i	θ_i	α_i	a_i	d_i
1	θ_1	-90°	0	290
2	θ_2	0	270	0
3	θ_3	-90°	70	0
4	θ_4	90°	0	168
5	θ_5	-90°	0	0
6	θ_6	0	0	0

2.2. Forward Kinematics Solution

The purpose of the forward kinematics is to find out the position and posture of the end effector that relative to the coordinate system {0}. The position and posture between adjacent coordinate systems are represented by 4×4 homogeneous transformation matrix as follows[1]:

$$^{I-1}T_i=\begin{bmatrix}\cos\theta_i & -\sin\theta_i\cos\alpha_i & \sin\theta_i\sin\alpha_i & a_i\cos\theta_i\\ \sin\theta_i & \cos\theta_i\cos\alpha_i & -\cos\theta_i\sin\alpha_i & a_i\sin\theta_i\\ 0 & \sin\alpha_i & \cos\alpha_i & d_i\\ 0 & 0 & 0 & 1\end{bmatrix} \quad (1)$$

$^{i-1}T_i$ represent the transformation matrix from ‘the i-1 coordinate system to the i coordinate system. Through substituting the D-H parameters in Table 1 into formula (1), the homogeneous transformation matrix between each connecting rod can be obtained as follows:

$$^{0}T_1=\begin{bmatrix}\cos\theta_1 & 0 & -\sin\theta_1 & 0\\ \sin\theta_1 & 0 & \cos\theta_1 & 0\\ 0 & -1 & 0 & d_1\\ 0 & 0 & 0 & 1\end{bmatrix} \quad ^{1}T_2=\begin{bmatrix}\cos\theta_2 & -\sin\theta_2 & 0 & a_2\cos\theta_2\\ \sin\theta_2 & \cos\theta_2 & 0 & a_2\sin\theta_2\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$$

$$^{2}T_3=\begin{bmatrix}\cos\theta_3 & 0 & -\sin\theta_3 & a_3\cos\theta_3\\ \sin\theta_3 & 0 & \cos\theta_3 & a_3\sin\theta_3\\ 0 & -1 & 0 & 0\\ 0 & 0 & 0 & 1\end{bmatrix} \quad ^{3}T_4=\begin{bmatrix}\cos\theta_4 & 0 & \sin\theta_3 & 0\\ \sin\theta_4 & 0 & -\cos\theta_3 & 0\\ 0 & 1 & 0 & d_4\\ 0 & 0 & 0 & 1\end{bmatrix}$$

$$^{4}T_5=\begin{bmatrix}\cos\theta_5 & 0 & -\sin\theta_5 & 0\\ \sin\theta_5 & 0 & \cos\theta_5 & 0\\ 0 & -1 & 0 & 0\\ 0 & 0 & 0 & 1\end{bmatrix} \quad ^{5}T_6=\begin{bmatrix}\cos\theta_6 & -\sin\theta_6 & 0 & 0\\ \sin\theta_6 & \cos\theta_6 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$$

So that,we can get the the spatial displacement and posture transformation of the 6th coordinate frame relative to the base/reference coordinate frame{0}:

$$^{0}T_6={}^{0}T_1\,{}^{1}T_2\,{}^{2}T_3\,{}^{3}T_4\,{}^{4}T_5\,{}^{5}T_6 \quad (2)$$

2.3. Inverse Kinematics Solution

If the parameters of each joint and link of the robot and the relative position of the end effector relative to the fixed reference coordinate system are known, how to get the angle θ_i between the connected link of the robot , which is the inverse kinematics analysis solution.

Solution of θ_1, According to the analysis of forward kinematics, we can multiply $^{0}T_1^{-1}$ form the left on both side of the Eq.（2）:

$$^{0}T_1^{-1}\,{}^{0}T_6={}^{1}T_2\,{}^{2}T_3\,{}^{3}T_4\,{}^{4}T_5\,{}^{5}T_6 \quad (3)$$

By comparing the elements (3,4)of the Eq.(3),we can know that :

$$-p_y c_1+p_x s_1=0 \quad (4)$$

where $s_i \equiv \sin\theta_i$, $c_i \equiv \cos\theta_i$, $s_{ij} \equiv \sin(\theta_i+\theta_j)$, $c_{ij} \equiv \cos(\theta_i+\theta_j)$,so that :

$$\theta_1=\operatorname{atan2}(py-d_6a_y,px-d_6a_x) \qquad (5)$$

Solution of θ_2. We can multiply ${}^5T_6^{-1}$ form the left on both side of the Eq.（2）:

$${}^5T_6^{-1}\ {}^0T_6={}^0T_1\ {}^1T_2\ {}^2T_3\ {}^3T_4\ {}^4T_5 \qquad (6)$$

By comparing the elements (1,4),(2,4),(3,4) of the Eq.(6),we can know that :

$$p_{5x}=p_x-d_6a_x=c_1(a_2c_2+a_3c_{23}-s_{23}d_4)$$

$$p_{5y}=p_y-d_6a_y=s_1(a_2c_2+a_3c_{23}-s_{23}d_4)$$

$$p_{5z}=p_z-d_6a_z=d_1-a_2s_2-a_3s_{23}-d_4c_{23}$$

In there ,we let $K=\sqrt{p_{5x}^2+p_{5y}^2}$, $L=p_z-d_6*a_z-d_1$, $M=(L^2+K^2+a_2^2-a_3^2-d_4^2)/2a_2$, according to the trigonometric theorem, θ_2 can be obtained that[2]:

$$\theta_2=\operatorname{atan2}(K,L)-\operatorname{atan2}(-M,(K^2+L^2-M^2)^{0.5}) \qquad (7)$$

Solution of θ_3. We can know that:

$$K-a_{2s2}=a_3s_{23}+d_4c_{23};$$

$$L-a_2c_2=a_3c_{23}+d_4s_{23};$$

according to the trigonometric theorem, θ_{23} can be obtained that:

$$\theta_{23}=\operatorname{atan2}(a_3,d_4)-\operatorname{atan2}(E,(a_3^2+d_4^2-E^2)^{0.5})$$

$$E=K-a2*\cos\theta_2$$

$$\theta_3=\theta_{23}-\theta_2 \qquad (8)$$

Solution of $\theta 4$. Multiply ${}^0T_3^{-1}$ form the left on both side of the Eq.（2）:

$${}^0T_3^{-1}\ {}^0T_6={}^0T_1\ {}^1T_2\ {}^2T_3$$

$$W=-s_4s_5=a_xs_1-a_yc_1$$

$$Z=-c_4s_5=c_1c_{23}a_x-s_{23}a_z+s_1c_{23}a_y$$

So that,we can get the angle θ_4:

$$\theta_4=\operatorname{atan2}(W,Z); \qquad (10)$$

Solution of θ_5. Multiply ${}^0T_4^{-1}$ form the left on both side of the Eq.（2）

$${}^0T_4^{-1}\ {}^0T_6={}^4T_5\ {}^5T_6 \qquad (11)$$

By comparing the elements of the Eq.(11),we can know that :

$$s_5=-a_x(s_1s_4+c_1c_4c_{23})-a_y(c_1s_4-s_1c_4c_{23})-a_zc_4s_{23}$$

$$c_5=-a_xc_1s_{23}-a_ys_1s_{23}-a_zc_{23}$$

$$\theta_5=\operatorname{atan2}(s_5,c_5) \qquad (12)$$

Solution of θ_6. It is similar to the solution of θ_5,we just need to multiply ${}^0T_5^{-1}$ form the left on both side of the Eq.（2）

$${}^0T_5^{-1}\ {}^0T_6={}^5T_6 \qquad (13)$$

By comparing the elements of the Eq.(13),we can know that :

$$s_6=-n_x(s_1c_4+c_1s_4c_{23})-n_y(c_1c_4+s_1s_4c_{23})+n_zs_4s_{23}$$

$$c_6=o_x(c_4s_1-c_1s_4c_{23})-o_y(c_1c_4+s_1s_4c_{23})+o_zs_4s_{23}$$

$$\theta_6=\operatorname{atan2}(s_6,c_6) \qquad (14)$$

2.4. Trajectory planning of the 6R robot

In this paper, we use the method of quintic polynomial interpolation to plan the trajectory of the 6R robot.

$$\theta(t)=a_0+a_1t+a_2t^2+a_3t^3+a_4t^4+a_5t^5$$

So that,the function expression of the velocity is:

$$\theta(t)'=a_1+2a_2t+3a_3t^2+4a_4t^3+5a_5t^4$$

And the function expression of the acceleration is:

$$\theta(t)''=2a_2+6_{a3}t+12a_4t^2+20a_5t^3$$

By making constraints about the position, angular velocity and angular acceleration of each joint in space at the starting and ending positions of the path point:

$$\begin{cases} \theta(t_0)=a_0 \\ \theta(t_f)=a_0+a_1t_f+a_2t_f^2+a_3t_f^3+a_4t_f^4+a_5t_f^5 \\ \dot{\theta}(t_0)=a_1 \\ \dot{\theta}(t_f)=a_1+2a_2t_f+3a_3t_f^2+4a_4t_f^3+5a_5t_f^4 \\ \ddot{\theta}(t_0)=2a_2 \\ \ddot{\theta}(t_f)=2a_2+6a_3t_f+12a_4t_f^2+20a_5t_f^3 \end{cases}$$

So this is obtained from the above formula:

$$\begin{cases} a_0=\theta_0 \\ a_1=\dot{\theta}_0 \\ a_2=\dfrac{\ddot{\theta}_0}{2} \\ a_3=\dfrac{20\theta_f-20\theta_0-(8\dot{\theta}_f+12\dot{\theta}_0)t_f-(3\ddot{\theta}_0-\ddot{\theta}_f)t_f^2}{2t_f^3} \\ a_4=\dfrac{30\theta_f-30\theta_0-(14\dot{\theta}_f+16\dot{\theta}_0)t_f+(3\ddot{\theta}_0-2\ddot{\theta}_f)t_f^2}{2t_f^3} \\ a_5=\dfrac{12\theta_f-12\theta_0-(6\dot{\theta}_f+6\dot{\theta}_0)t_f-(\ddot{\theta}_0-\ddot{\theta}_f)t_f^2}{2t_f^3} \end{cases}$$

By substituting the above coefficients into the basic expression, the function expression using quintic polynomial interpolation method can be obtained.

3. The Kinematics Simulation Verification based on MATLAB

We use the Matlab Robotics Toolbox to verify the accuracy of kinematics solution. By given a joint angle group 'q=[pi/3, pi/4, 3*pi/4, -pi/5, pi/5 ,pi/6]', we use the forward and inverse kinematics algorithm finished in chapter 2.2 and 2.3 to get the pose and posture matrix of end effector, and the solution of each joint angle in Matlab , as shown in figure 3.

Then by comparing with the result that obtained by using the 'Fkine' and 'Ikine' function of Matlab,we found that the two results are consistent,so that the kinematics solution is verified.

$$T = \begin{matrix} 0.6235 & 0.2091 & -0.7534 & 0.06081 \\ -0.5528 & -0.5636 & -0.6139 & 0.1053 \\ -0.5529 & 0.7992 & -0.2358 & -0.06921 \\ 0 & 0 & 0 & 1 \end{matrix}$$

1.0471	1.7033	-2.3521	-2.7622	-1.9400	-1.9437
1.0471	0.7854	2.3602	-0.3637	-1.8126	0.9640
-1.0471	1.7033	-2.3521	1.5907	-1.2855	-2.1235
-1.0471	0.7854	2.3602	1.7318	-1.8082	2.0203
1.0471	1.7033	-2.3521	0.3794	1.9400	1.1979
1.0471	0.7854	2.3602	2.7779	1.8126	4.1056
-1.0471	1.7033	-2.3521	3.1217	1.2855	1.0181
-1.0471	0.7854	2.3602	4.8734	1.8082	5.1619

Fig. 3. Kinematics solution result finished by Matlab

The trajectory planning of the robot is also simulated by Matlab ,the curve of the velocity and acceleration of each joint angle with time ,shown in the figure 4.

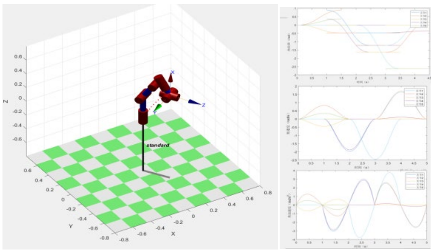


Fig. 4. Trajectory path and joint motion characteristic curve

4. The Visual simulation based on Simulink

The robot is modeled by SolidWorks 2016. First, the models of the robot is built and assembled. Then the assembly model is converted into the format of .urdf and imported into the Simulink. Then by using the modular simulation interface of the Simulink and referring with the previous solving function, the robotic visual simulation framework is finished,as shown in figure 5 and figure 6.

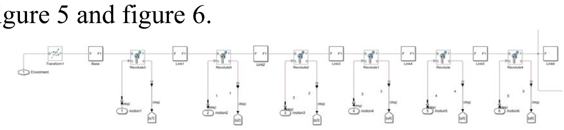


Fig. 5. The model of robot imported in Simulink

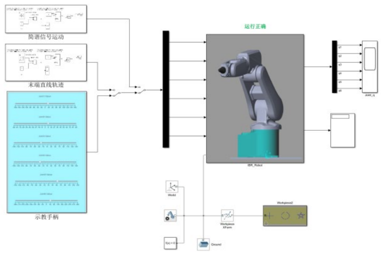


Fig. 6. Kinematics visual simulation by Simulink

5. Conclusion

In this paper,we take the IRB120 robot as the research object, using the D-H method to establish the connecting rod coordinate system, finish the solution of the forward and inverse kinematics analysis, get the 6R robot kinematics algorithm. Then the Matlab is used to model and simulate the robot. It can be seen from the simulation results that the model of the robot established in this study is correct, and the trajectory planning of the end effector is carried out, and the change of the angle of each joint with time is recorded, which verified the effectiveness of the kinematics algorithm.

Acknowledgements

This work was supported by the Key Research and Development Program (No.M18GY300021).

References

1. Guan Wei, "Research on obstacle avoidance path and trajectory planning foe six-dof manipulator," D. Shanghai Jiao Tong University, Jan.2016.
2. Hu Yaqing, Yu Jinpeng, Kinematics Analysis and Modeling Simulation Based on ABB-IRB120 Robot.J.Journal of Qingdao University(E&T) , Aug.2017,Vol.32.

Distance measurement system based on binocular stereo vision

Jiwu Wang
School of Mechanical and Electronic Engineering, Beijing Jiaotong University
Beijing, Haidian District, China

Xin Pei
School of Mechanical and Electronic Engineering, Beijing Jiaotong University
Beijing, Haidian District, China
E-mail: jwwang@bjtu.edu.cn, 19125998@bjtu.edu.cn
www.bjtu.edu.cn

Abstract

The distance measurement system based on binocular stereo vision is proposed. The binocular vision model is established, Matlab toolbox is used to calibrate and obtain the internal and external parameters of the binocular camera. Then stereo rectification and stereo matching are carried out with OpenCV. Finally, the depth information is obtained by the parallax measurement of image pairs, and the distance between any two points can be calculated. The simulation experiments have been done and the results demonstrated that the measuring method is feasible.

Keywords: stereo vision; camera calibration; stereo matching; distance measurement.

1. Introduction

Distance measurement is one of the most important tasks of a computer vision system, and it is also continues to be a popular research topic in computer and robot vision nowadays. Distance measurement plays an important role in robot navigation, obstacle detection, navigation of autonomous vehicles, surveillance monitoring, person localization and tracking, and many more. The distance measurement system based on binocular stereo vision is proposed in this paper. Firstly, the binocular vision model is established, a GUI interface is designed in MATLAB for calibration and the display of parameters. Then stereo rectification and stereo matching are carried out with OpenCV. Finally, the depth information is obtained by the parallax measurement of image pairs, and the distance between any two points can be calculated. The simulation experiments have been done and the results demonstrated that the measurement method is feasible.

The designed binocular stereo vision system consists of five modules: image acquisition, camera calibration, image rectification, stereo matching, and distance measurement. This system can be described by the flow diagram shown in Figure 1.

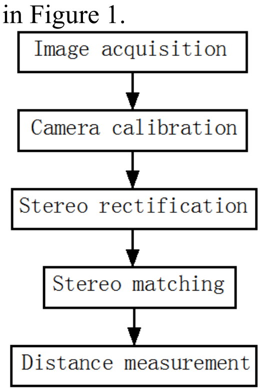


Fig. 1. System flow chart

2. Principle of binocular stereo vision ranging

Binocular stereo vision distance measurement is modeled on the distance perception methods of human eyes. In this paper, the stereo vision system consists of two cameras with the same parameters that is parallel mounted and looking at the same objects, as shown in Figure 2. The projection model is created based on the imaging principle of pinhole camera. The object points P are given in a world coordinate system. P_l and P_r are imaging points on left and right images. Assuming that the image plane of the left and right cameras is located on the same plane and the two images are aligned, it can be known from the principle of similar triangle that:

$$\frac{B-(x_l-x_r)}{Z-f}=\frac{B}{Z} \tag{1}$$

Therefore, it can be derived from equation (1):

$$Z=\frac{fB}{x_l-x_r} \tag{2}$$

Where, B is the base distance between the two cameras, and f is the focal length of the camera lens. Again, xl and xr respectively represent the pixel distance of Pl and Pr in the X-axis direction in their respective coordinate system. The difference in the coordinates of the corresponding pixels is known as disparity d. Z is the depth distance of the object point from the camera position.

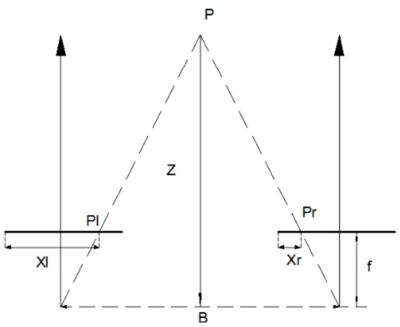


Fig. 2. Binocular vision model

3. Camera calibration

The process of calibration is actually the process of sol-ving the internal and external parameters of the camera. Calibration is the basic and necessary process of binocular stereo vision. The result of calibration will determine the accuracy of distance measurement of the target object. In the process of camera calibration, three coordinate systems are needed: image coordinate system, camera coordinate system and world coordinate system. The camera is approximated to a pinhole model, and the parameter S (scale factor) and the homography matrix H are introduced. The relationship between the points $P=[X \quad Y]^T$ in the imaging plane and $P=[X \quad Y \quad Z]^T$ in the space can be expressed as: $P=sHP$

Where H is represented by two matrices: $H=MW$

$$M=\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

M is the intrinsic parameter matrix of the camera.Where, c_x and c_y represent coordinates in pixel of the image center respectively. Since pixels on a normal camera are not guaranteed to be square, use two different focal lengths f_x and f_y.

$W=[R\ T]$,where, R , T are the extrinsic parameters of the cameras, denoting the rotation matrix and the translation vector from the left camera coordinate system to the right camera coordinate system, respectively.

Considering the distortion of lens, suppose $q_p(x_p, y_p)$ is the corrected point, and $q_p(x_p, y_p)$ is the distorted point, then:

$$\begin{bmatrix} x_p \\ y_p \end{bmatrix}=\left(1+k_1r^2+k_2r^4+k_3r^6\right)\begin{bmatrix} x_d \\ y_d \end{bmatrix}+\begin{bmatrix} 2p_1x_dy_d+p_2\left(r^2+2x_d^2\right) \\ p_1\left(r^2+2y_d^2\right)+2p_2x_dy_d \end{bmatrix} \tag{3}$$

Where (K_1, K_2, P_1, P_2, K_3) constitutes a 5 × 1 matrix, which is the distortion matrix of the camera.

The method of camera calibration used in this paper is based on the checkerboard template calibration method proposed by Zhang, which has the advantages of simple operation and high precision. In this paper, a GUI interface is designed in MATLAB for calibration and display of results. The specific steps are as follows:

(i) Self-made calibration plate; The number of c corners is 7×5 and the size of the calibration target is 20mm×20mm

(ii) Image acquisition; Images from different angles of the calibration target are collected simultaneously with a binocular camera placed in parallel, and the image resolution is 640×480. A total of 13 sets of image pairs were collected, as shown in Figure 3 and Figure 4.

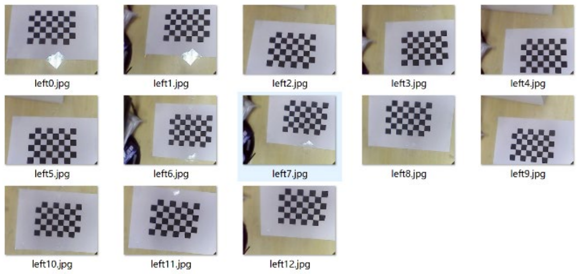


Fig. 3. Left images of calibration target images

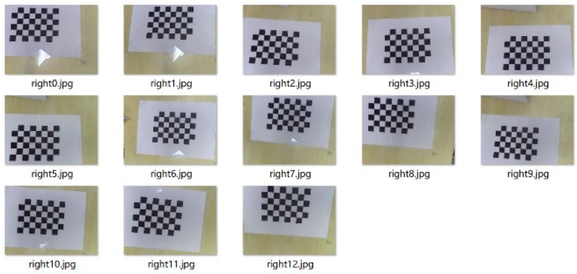


Fig. 4. Right images of calibration target images

(iii) Calibration. Click the calibration button on the GUI interface to import the binocular image pair collected. Then, corner detection and stereo calibration are carried out to obtain the internal and external parameters of the binocular camera and display them on the GUI interface, as shown in Figure 5.

Fig. 5. GUI interface

4. Stereo rectification

The purposes of image rectification are as follows: the distortions of images are removed by the internal parameters in the calibration results, and the two images aligned in non-coplanar lines are rectified to coplanar line alignment such that the polar lines of the two images lie exactly on the same horizontal line. Any point on such an image must have the same line number as its corresponding point on another image. Therefore, in case of searching for corresponding points in two images, it is only necessary to search in the same epipolar line, reducing a 2D search space to 1D to reduce the search time effectively.

The stereo rectification can be realized by the function of cvStereoRectify in OpenCV. Figure 6 shows the left and right image pairs before stereo rectification. Figure 7 shows the image pairs after stereo rectification.

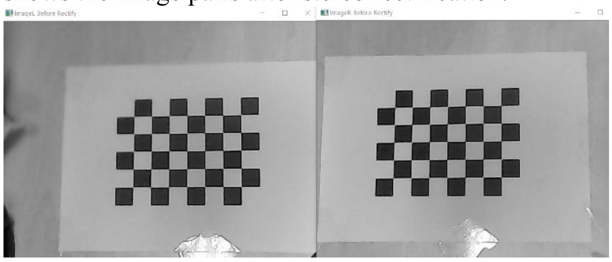


Fig. 6. image pairs before stereo rectification

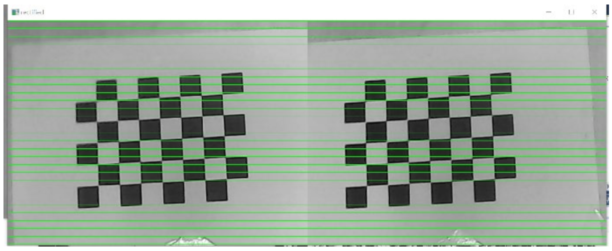


Fig. 7. image pairs after stereo rectification

5. Stereo matching

The stereo matching algorithm is an important step in generating disparity maps. Whether stereo matching is accurate directly affects the accuracy of binocular ranging results. The basic principle of stereo matching is to find the corresponding relationship between the pixels of the same scene in the projected image under different viewpoints, and then to obtain the distance information of object by calculating the parallax between the pixels.

In this paper, CV Find Stereo Correspondence BM, a block-matching Stereo Correspondence algorithm provided by Open CV, is used to find the matching points between the left and right images after stereo rectification through a small window (SAD) called " Sum of Absolute Difference ".

Through matching the left and right rectified image pairs in OpenCV, the disparity map can be obtained. In the disparity map, any two target points can be selected with the mouse to obtain their depth information, and the distance between the two points can be calculated, as shown in Figure 8.

Fig. 5. Disparity map

The actual distance between the corners of the calibration plate is 20mm. From the results of experiments, it can be seen that the measurement method basically meets the requirement of precision, the measuring method is feasible.

6. Conclusion

A procedure for distance measurement using a binocular stereo vision system was presented in this article. The experimental results demonstrated that the system could meet basically the needs of measuring. The measurement method is feasible for applications where accurate distance measurement is not required and could offer better technical support for machine vision applications. However, due to the influence of matching and other factors, the accuracy should be further improved in the future research work.

Acknowledgements

This work was supported by the Key Research and Development Program (No.M18GY300021).

References

1. Gary Bradski, Adrian Kaebler. Learning OpenCV[M]. Translated by Yu Shiqi and Liu Ruizhen. Beijing: Tsinghua University Press, 2009.
2. Hu Hanping. Research on key Technologies of Binocular stereoscopic ranging [D]. Graduate University of The Chinese Academy of Sciences (Changchun Institute of Optics, Fine Mechanics and Physics),2014.
3. Zhao Xunjie, Li Chengjin. Research on key Technologies of binocular Stereo Real-time ranging System [J]. Laser & Infrared,2006(09):874-877.
4. Wang Zhenzhen, Zhao Lianyu, Liu Zhenzhong. Binocular stereo ranging system based on MATLAB and OpenCV [J]. Journal of tianjin university of technology,2013,29(01):45-48.
5. Shen Tong, Liu Wenbo, WANG Jing. Target ranging system based on binocular stereo vision [J]. Electronic measurement technology,2015,38(04):52-54.
6. Wang Hao, Xu Zhiwen, Xie Kun, Li Jie, Song Chenglu. Binocular ranging system based on OpenCV [J]. Journal of jilin university (information science),2014,32(02):188-194.

Authors Introduction

Dr. Jiwu Wang

He is an associate professor, Beijing Jiaotong University. His research interests are Intelligent Robot, Machine Vision, and Image Processing.

Ms. Pei Xin

She is a postgraduate in Beijing Jiaotong University. Her research interest is stereo vision and image processing.

Robustness Verification Against Noise of Self-localization Method Using Omni-directional Camera for Soccer Robot

Yuehang Ma, Kaori Watanabe, Hidekazu Suzuki
Graduate School of Engineering, Tokyo Polytechnic University, 1583 Iiyama
Atsugi, Kanagawa 243-0297, Japan
E-mail: xmaancn@gmail.com, watanabe.kougei.karakuri@gmail.com, hsuzuki@em.t-kougei.ac.jp
www.t-kougei.ac.jp

Abstract

The main focus of the RoboCup competitions is the game of football/soccer, where the research goals concern cooperative multi-robot and multi-agent systems in dynamic adversarial environments[1]. In the field of RoboCup, self-localization technique is important to estimate own position including goal and other robot positions and to decide strategy. This paper presents a self-localization technique using an omni-directional camera for an autonomous soccer robot. We propose the self-localization method with white line information of soccer field, and recognize the robot position by optimizing the fitness function using Genetic Algorithm. Moreover, we also verify the robustness of the proposed method against noise through experiments.

Keywords: Robustness Against Noise, Self-Localization, RoboCup Middle Size League, Soccer Robot, Genetic Algorithm

1. Introduction

In the field of RoboCup, self-localization technique is important to estimate own position including goal and other robot positions and to decide strategy. Basically, we estimate the self-position with the image information, the environment information and the field information. In this paper, we describe a real-time self-localization method that applies a genetic algorithm (GA) for the RoboCup middle size league, and verify the robustness detection of this method.

2. Hardware of vision system

About the omni-directional vision system of our robot is consisted of the camera (FLIR, Flea3[2]), a varifocal lens (Vstone) and a hyperboloidal mirror (Vstone). We developed vision system shown in Fig. 1 for RoboCup MSL robot by combining with above elements[3]. The image captured by this vision system is shown in Fig. 2(a), and the image size and frame rate are 512 × 512 [pixels] and 30 [fps] respectively.

Fig. 1. Hardware of vision system

3. Self-localization

We use a white line of MSL field for self-localization. We have proposed the self-localization method, which generates the searching space based on a model based matching using white line information[4]. And this method recognizes the robot position by optimizing the fitness function, which has the maximum value at correct robot position. Moreover, this proposed self-localization method employs Genetic Algorithm (GA)[5] for optimization of the fitness function.

3.1. *Searching model*

Figure 2 shows the process of making the searching model of the proposed method. At first, we need the detection image of the white line for making the searching model. We obtain the white detection image by employing the converting method of color space from RGB to HSV and to YUV like Fig. 2b. Then we generate the field information by orthogonalizing the white line information like Fig. 2c. Moreover, we determine the searching model by thinning down the field information based on white line like Fig. 3d. Therefore, we use thinned model as searching model for the self-localization.

3.2. *Model-based matching*

The proposed self-localization method generates the searching space by model based matching between geometric information of the white line in the MSL field and above-mentioned searching model. We use this method to calculate the evaluation function $F(\tilde{\phi})$, $\tilde{\phi} = [\tilde{x}, \tilde{y}, \tilde{\theta}]$. The fitness function $F(\tilde{\phi})$ obtains the maximum value when the position of the searching model corresponds to the correct position that robot exist in the MSL field. Then, the problem of detection of robot position is converted to the searching problem of $\tilde{\phi}$ such that $F(\tilde{\phi})$ is maximized. Due to the revolution symmetry of the white line of the MSL field, we may get two maximum value exist in the fitness function. Here, we select only one depending on an electric compass.

3.3. *Genetic Algorithm*

In the proposed self-localization method, we employ Genetic Algorithm (GA) for searching the maximum value of the fitness function $F(\tilde{\phi})$. A GA is an example of an artificial intelligence program and is well known as a parallel search and optimization process that mimics natural selection and evolution. In the proposed method, an elitist model of a GA that preserves the best individual in the population at every generation is utilized and

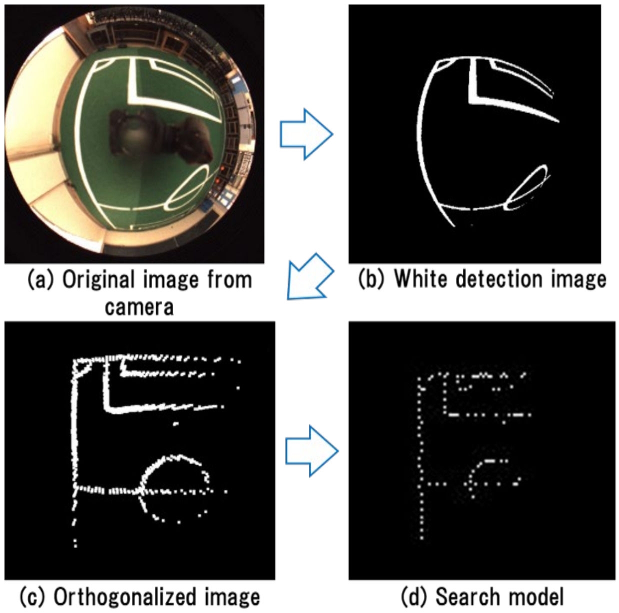
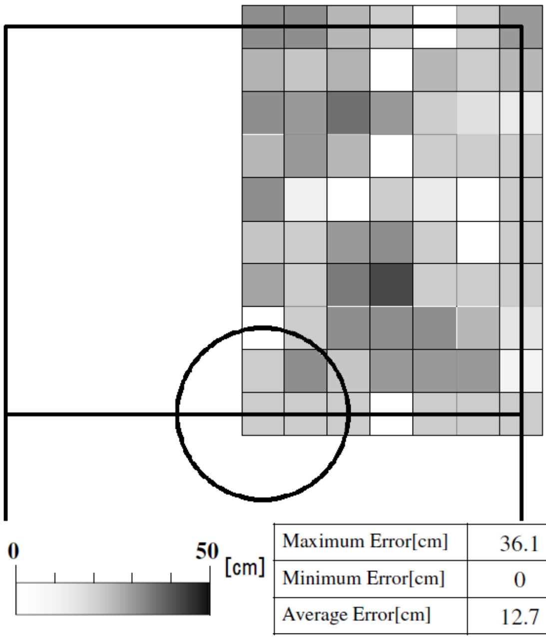


Fig. 2. Process of making search model



Maximum Error[cm]	36.1
Minimum Error[cm]	0
Average Error[cm]	12.7

Fig. 3. Error of the self-localization

genetic coding using gray code, roulette selection and one-point crossover are employed. And, the parameters of the GA process are determined by previous experiments.

3.4. *Verification experiment*

We performed the self-localization experiment to verify the effectiveness of the proposed method. Figure 3 shows the result of the verification experiment that checked the self-localization error between correct position and detected position at the quarter area of the MSL field at interval of one meter. In this figure, each box represents the error as the brightness of gray scale. Average error of this experiment was 12.7[cm], and the accuracy of the self-localization by the proposed method is enough to play soccer.

Fig.5. Experiment location

4. Robustness verification

The accuracy measured so far is the result of an ideal environment when there are no other robots on the soccer field. However, in a real game environment, up to ten robots may exist in the same soccer field. In addition to the robots, there are humans such as the referee and the line referee at the same time on the field. Then the robot may not be able to recognize correct own position, because these become an occlusions. Therefore, to verify the robustness of proposed method against various noise, we conducted experiments.

4.1. *Experiment*

In a real environment, there will be a variety of noises appearing on the image, and we cannot perform qualitative experiments. Here we perform verification experiments by adding artificial noise.

We conducted the experiments at the seven locations in the MSL field indicated in Fig. 5. These seven locations are especially characteristics point in MSL field. Taking the center of the panoramic image as the center of the circle, we set the fan-shaped area with a center angle of 30 degrees as a noise like Fig. 6. By changing the position and number of the noise, we can simulate different noise conditions, and verify the results of self-localization.

Figure 6 shows the change in the number of noise at the position Fig. 5(d), and the change in angle is shown in Fig. 7. The image in the top left of Fig. 6 is original image without any noise. We verify the error of self-localization results measured at the angle and number of various noise at each coordinate location in the MSL field. And the results of each measurement position are

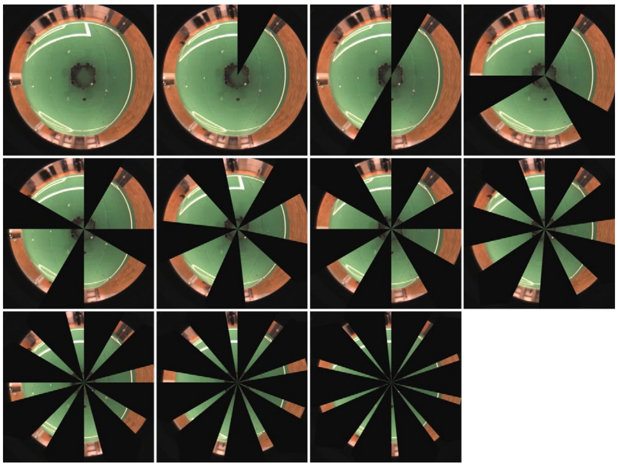


Fig.6. Noise at different amounts

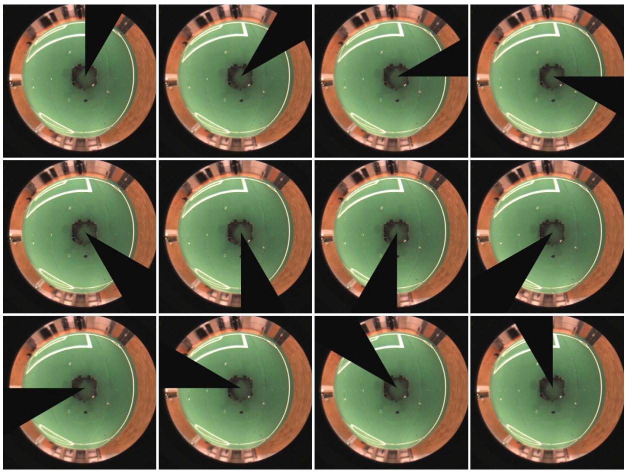


Fig.7. Noise at different angles

Table 1. Error comparison result

	Error[cm]	Noise num									
	Rate	1	2	3	4	5	6	7	8	9	10
Point	A	28.3	28.3	28.3	28.3	150.3	22.4	72.8	1510.0	28.3	131.5
		0.00%	0.00%	0.00%	0.00%	8.33%	0.00%	8.33%	66.67%	0.00%	16.67%
	B	14.1	14.1	14.1	14.1	31.6	14.1	41.2	22.4	648.5	1164.8
		0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	50.00%	100.00%
	C	10.0	14.1	0.0	10.0	14.1	10.0	14.1	10.0	14.1	10.0
		0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	D	20.0	14.1	10.0	20.0	20.0	10.0	1310.3	429.5	676.8	1207.8
		0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	8.33%	66.67%	75.00%	100.00%
	E	22.4	22.4	22.4	58.3	1650.5	22.4	22.4	31.6	31.6	31.6
		0.00%	0.00%	0.00%	33.33%	8.33%	0.00%	0.00%	0.00%	0.00%	0.00%
	F	30.0	30.0	41.2	14.1	20.0	40.0	20.0	40.0	41.2	370.1
		0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	66.67%
	G	10.0	10.0	40.0	120.0	40.0	10.0	820.1	900.1	890.1	900.1
		0.00%	0.00%	0.00%	33.33%	0.00%	0.00%	33.33%	66.67%	75.00%	66.67%

summarized in Table 1. For the position deviation caused by the noise of different angles, we take the average value combined with the deviation value caused by the amount of noise. The red grid in the table indicates that the error value of the self-localization exceeds the width of a robot (50cm). It means that the robot has lost its accurate position.

4.2. *Experiment results*

According to the results in Table 1, when the number of noises increases, the error of self-localization also increase. Overall, when the number of noises is greater than 7, it will cause large errors except for special locations. Here, the special locations is when there are more than two features in the image obtained by the robot (Location B and F) or when the robot is in the center circle (Location C). Location E and G also have errors when the amount of noise is not large. The reason may be that the robot is on the white line crossed vertically, and the 4 noises hide all the white line information. Overall, from the results in Table 1, when the number of noises is less than 3, the error of self- localization is very small. And when the number of noises increases to 4, the maximum error of self- localization is 120cm. In a real game environment, the probability of more than 4 noises existing at the same time is very small. Therefore, this method has robustness against noise enough to play soccer.

5. Conclusion

In this paper, we have proposed the self-localization method that generates the searching space based on a model based matching with white line information of RoboCup MSL soccer field, and which recognizes the robot position by optimizing the fitness function using Genetic Algorithm. Moreover, we have verified the effectiveness and accuracy of the proposed self-localization method using GA. Furthermore, we confirmed that the robustness against noise of the self-localization by the proposed method is enough to play soccer through experiments.

References

1. “RoboCup Federation official website,”, https://www.robocup.org/
2. FLIR, “Flea3(USB3 Vision Camera)”, https://www.ptgrey.com/flea3-usb3-vision-cameras
3. “RV-Infinity Discription paper”, http://www.seit.t-kougei.ac.jp/vision/rv-infinity/
4. K. Watanabe, Y. Ma and H. Suzuki, Real-time Self-localization using Model-based Matching for Autonomous Robot of RoboCup MSL, *Journal of Robotics, Networking and Artificial Life,* Vol. 7(2); September (2020), pp. 98–102
5. D.E. Goldberg, Genetic algorithms in search, optimization and machine learning, Addison-Wesley, Reading, MA, USA, 1989.

Tomato-Harvesting Robot Competition: Aims and Developed Robot of 6th Competitions

Takayuki Matsuo
National Institute Technology, Kitakyushu College
2-20-1, Shii, Kokuraminami-ku, kitakyushu-shi, Fukuoka, Japan

Yasunori Takemura, Takashi Sonoda
Department of Engineering, Nishinippon Institute of Technology, 2-11, Aratsu, Kanda-town,
Miyako-gun, Fukuoka 800-0396, Japan

Yuya Nishida, Shinsuke Yasukawa, Kazuo Ishii
Department of Life Science and System Engineering, Graduate School of Kyushu Institute of Technology,
2-4, Hibikino, Wakamatsu, Kitakyushu-city, Fukuoka 808-0196, Japan

E-mail: matsuo@kct.ac.jp, takemura@nishitech.ac.jp, sonoda@nishitech.ac.jp,
s-yasukawa@brain.kyutech.ac.jp, y-nishida@brain.kyutech.ac.jp, ishii@brain.kyutech.ac.jp
http://www.lsse.kyutech.ac.jp/~sociorobo/ja/tomato-robot2019

Abstract

In general, farm work is often hard work. In recent years, the number of farm laborers has been decreasing due to the aging of the workforce. Therefore, there is a need to automate agricultural work, and we are organizing a tomato harvesting robot competition to automate tomato harvesting. The competition includes a Junior Division for high school students and a Senior Division for college students. In this paper, we report on the results of 6th tomato harvesting robot competition.

Keywords: robots for socio synthesis, robot competition and agriculture robot

1. Introduction

In recent years, the decline and aging of the agricultural workforce has become an issue. According to data[1] released by Japan's Ministry of Agriculture, Forestry and Fisheries, the number of agricultural workers in 2020 was 1,361,000, a decrease of 22.5% or 396,000 compared to the previous survey in 2015. In addition, the percentage of people aged 65 and above rose to 69.8%, up 4.9% from the previous survey. This is due to the fact that a large number of people quit farming due to old age, and securing the number of farmers is an issue. In recent years, in order to cope with the decreasing number of agricultural workers, there has been a lot of research and development of agricultural robots aimed at automating agricultural work. Among them, we have focused on tomatoes and have been holding tomato harvesting robot competitions since 2014. As the research of tomato harvesting robots in Japan, Kawamura et. al. developed a mobile robot with manipulator [2] and proposed the tomato harvesting method using image processing and visual feedback [3]. Kondo et. al. proposed the method to improve the success ratio of tomato harvesting and speed-up technique [4].

In this paper, we will report on the result of the 6th tomato harvesting robot competition.

2. Competition regulations

2.1. *Senior League*

In Senior League, two kind of competition field are designed, the one is rail-style area and the another one is fee-style area as shown in Fig.1 and Fig.2. The rail-style area is designed to have the similar environment with the tomato factory. Free-style area is for the robots of general tomato fields in outdoor environment. In first stage, a tomato is suspended and the team is able to advance to the second stage when an end-effecter of the robot is able to touch the tomato. In second stage, a cluster of tomato is suspended. Teams compete against each other to see how many tomatoes they can harvest. The score is calculated using Eq. 1 below.

$$\boldsymbol{P} = \boldsymbol{\eta C}(2\boldsymbol{\alpha} + \boldsymbol{\beta}) - 2(\boldsymbol{\gamma} + \boldsymbol{\delta}) - \boldsymbol{\varepsilon} \qquad (1)$$

$$\boldsymbol{\eta} = \frac{\boldsymbol{\alpha}}{\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma}+\boldsymbol{\delta}} \qquad (2)$$

, where P is score, C is coefficient of magnification depending on selecting class as shown in Table1, 2, $\alpha$is the number of tomatoes which is no damaged and correct color, β is the number of damaged tomatoes, γ is the number of drop tomatoes and γ is the number of damaged tomatoes which are not harvested. ε is a deducted point when robots damage stalks of tomato plants and the point is deducted in Final Stage. Eq. 2 represents the harvest rate η. If multiple tomatoes are harvested but dropped or damaged, the final score will be less. In final stage, the robots harvest tomato from plant body. In addition, some farms do not actually have rails, in which case the road surface in the vicinity of the tomato plants has a slope. Starting from the 5th competition, we have installed a slope in the freestyle area to try to replicate the harvesting conditions of an actual farm.

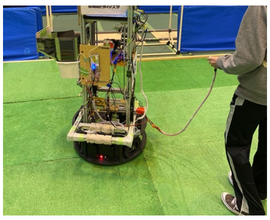


Fig.1 Free-style area

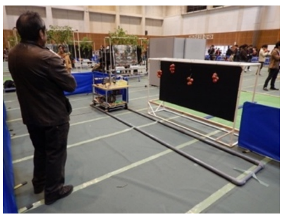


Fig.2 Rail-style area

Fig. 3 A slope set up in the freestyle area.

Table 1 The class number and coefficient at choosing remote control

Method of View	Directly		Indirectly	
Area	Rail	Free	Rail	Free
Number of Category	T1	T2	T3	T4
Coefficient C	1	2	2	4

Table 2 The class number and coefficient at choosing autonomous control

Area	Rail	Free
Number of Category	T5	T6
Coefficient C	8	16

2.2. *Junior League*

In Junior League, the subject is to carry small size tomatoes to assigned positions. Students should develop the robot using LEGO Mindstorm with the functions such as line trace, color recognition, end-effector with mechanism design and motors with control, and their programming [5]. The basic specification for robot is that the size of the robot is within 300mm x 300mm on ground [5].Height is no limited. Students should develop the robot using LEGO Mindstorm. Competition subjects include Line Trace Challenge, Color Identify Challenge, Mechanism Design and Control Challenge and Object Detection Challenge. In Line Trace Challenge, robots should detect black line in the competition area and move along the line using a color sensor. The robot starts from the starting point. In the middle of the course, the tomato harvesting field (harvest field) exists, where tomatoes are arranged. The robot must move to the harvest field in order to get the 6 tomatoes. Until the 4th Tomato Harvesting Robot Competition, the tomatoes were stored in a transparent box called "Tomato Box" beforehand, but from the 5th competition, the tomatoes were placed in a fresh state. Therefore, more careful handling of tomatoes is needed when the robot manipulates and transports them.

In Color Identifying Challenge, robots should explore and recognize color signs in the middle of the course and the same color of tomatoes. As guidance to harvest field, red, yellow and blue lines are drawn in the field as shown Fig.4. It is necessary for the robots to detect guidance line by color sensor. The robot gets the tomatoes of the same color. Along the black line for line trace, each color is signed. The robot can move to the harvest field from the lone detecting the color. In Mechanism Design and Control Challenge, robot should manipulate tomatoes using manipulator and carry to the storage location. To pick up the tomato box by using a manipulator equipment which is made by each team, participants are expected to design and make a device to get tomatoes on their idea. The robot is required to store, transport and relocation depending on tomatoes in each color. After picking up the tomatoes, the robot should return to the course. Then, the robot carries the tomatoes to the specific storage location. In Object Detection Challenge, robots should detect a battery charging station and stop there.

In addition to the above rules, we have added a new rule to handle tomatoes with care. Tomatoes are a food. Tomatoes are a food product, and if they are handled improperly in this competition, they will not be scored for storage points. The procedure for checking tomatoes in the competition and the criteria for judging when tomatoes are not handled properly are as follows:

1. A representative of each team member must check the tomatoes for blemishes before trials.
2. After the trials, all tomatoes must be checked for scratches by the referee and one team member. If there is any objection, one team member and the organizing committee chairman and vice chairman will discuss the matter.

Criteria for improper handling of tomatoes
1. if the liquid inside the tomato has been released to the outside
2. if the tomato is damaged in such a way that the inside of the tomato is visible
3. tomatoes are clearly dented from their pre-race condition
4. tomatoes are being transported in contact with the floor
5. if it is determined that the tomatoes have not been handled in a manner other than the above criteria

By adding the above rules, the goal was to instill in the competitors the awareness of handling tomatoes with care when carrying the actual product.

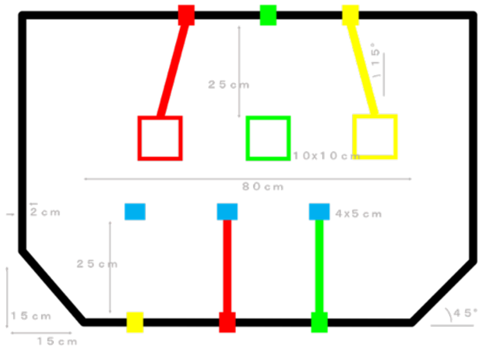


Fig.4 Robot Jr. League competition Field

3. Result of 6th tomato harvesting robot competition

The results of the competition of Senior League are shown in Table 3. The Senior League has revised its rules since the 5th competition, including changes to the scoring equation and the addition of slopes to the freestyle area. The equation is shown in Eq. 1, which introduces the concept of harvest rate, where dropping tomatoes and scratching tomatoes have a significant impact on the score. For practical use, the robot needs to harvest more tomatoes within a certain time frame. Therefore, it is desirable for the robot to work for the entire 10 minutes of the competition, but many teams quit the competition halfway through for fear of losing the harvest rate. As for the slope, many people said that it would be difficult to control the robot because it would slide down the slope without control.

The results of the competition of Junior League are shown in Table 4. The 6th competition was the second competition after the rules were revised to require the teams to harvest and transport fresh tomatoes. In the previous year, some teams damaged the tomatoes, but in this year's competition, there was some ingenuity in handling the tomatoes with care, and the students were able to instill an awareness of the importance of handling raw plants through their crafts.

Overall, the competition scores improved from the previous year, and one team was able to harvest and transport all the tomatoes in one week. Therefore, it would be better to make some rule changes for the next competition based on the competition held this time (since it is a high school competition, we will not change the basic base, but will consider some additional elements).

Finally, some teams commented that it was difficult to make tomatoes of a consistent size. Lastly, some teams commented that it was difficult because the size of the tomatoes was not consistent. This may be a result of the fact that the purpose of the competition, which is to encourage ingenuity in the devices and algorithms due to the inconsistent environment of handling natural objects, was not well understood. I felt that it was necessary to create a mechanism in the regulations and website to appeal the purpose of this competition, that it is not only about winning and losing as a competition, but also about developing the ability to cope with nature

Table 3 Result of Senior League

Ranking	Team
Overall winner and 1st place of rail-style division	Hibikino-Toms (Kyushu Inst. Of Tech.)
Overall runner-up and 2nd place of rail-style division	Syuga-Lab (Nagasaki Inst. Of Applied Science

Table 4 Result of Junior League

Rankin	Team
1st Place	Tangokinasu (Fukuoka Joto High School)
2nd Place	Double Lycopin (Fukuoka Joto High School)
3rd Place	NiASience Ver.2 (Nagasaki Sogo Fuzoku High School)
Special Award	
Best Presentation Award	G-Advance (Fukuoka Joto Highschool)
Perfect Award	Tangokinasu (Fukuoka Joto High School)
Special Judges Award	Kako Tomato Curry (Kashii Technical High school)
Idea Award	Revolutionary (Fukuoka Joto High School)
Challenge Award	SyokugyoMochi (National institute of Technology, Kitakyushu college)

4. Conclutions

In this paper, Tomato robot competition for robot socio-synthesis was introduced. The tomato harvesting robot competition has two leagues which are Senior League and Junior League. In Junior League, number of the participated teams were 24 teams. In Senior League, number of participated team 9 team. In this competition, concept of carefully dealing with tomatoes will be employed.

Acknowledgements

This works was supported by Kyushu Institute of Technology, Center for Socio-Robotics Synthesis, The University of Kitakyushu, Nippon Bunri University, Nagasaki Prefecture University, Kyushu Polytechnic Collage , Nishinippon Institute of Technology(This support was including The Ministry of Education, Culture, Sports, Science and Technology(MEXT), Center of Community), National College of Technology, Kitakyushu College, Hibikinada-Saien, Afrel Co., Ltd., City of Kitakyushu, The Kitakyushu Chamber of Commerce and Industry, Institute of Systems, Information Technologies and Nanotechnologies (ISIT), Japan Chapter of IEEE Society on Systems, Man, and Cybernetics, The Japan Society of Mechanical Engineers(JSME), The Robotics Society of Japan(RSJ), Kitakyushu Foundation for the Advancement of Industry Science and Technology (FAIS), IT and Mechatoronics Chapter and Kyushu Chapter of Japanese Society of Agricultural Machinery and Food Engineers, The MathWorks Co. Ltd., and Ministry of Agriculture, Forestry and Fisheries(MAFF).

References

1. 2020 CENSUS of Agriculture and Forestry in Japan, https://www.maff.go.jp/j/tokei/census/afc/, 2020
2. Kawamura et. al. Study on agricultural robot (1st report), The journal of JASM, vol. 46 No.3,pp.353-358, 1984 in Japanese
3. Kawamura et. al., Fruit harvesting robot, The journal of JASM vol.47 No.2, pp. 237-241, 1985 in Japanese
4. Kondo, et.al., A Machine Vision System for Tomato Cluster Harvesting Robot, Engineering in Agriculture, Environment and Food, Vol. 2, No. 2, pp.60-64, 2009
5. The 6th Tomato robot competition junior rule, "http://www.lsse.kyutech.ac.jp/~sociorobo/wp-content/uploads/2019/07/fb3520b51d93fdb42e034fa83b237857.pdf", 2019, 7

Smart Agriculture IoT Education Course in enPiT-everi (Education Network for Practical Information Technologies - Evolving and Empowering Regional Industries)*

Yasunori Takemura, Keiji Kamei, Atsushi Sanada
Department of Life Science and System Engineering, Kyushu Institute of Technology, 2-4, Hibkinino, Wakamatsu Kitakyushu, Fukuoka 808-0196, Japan
Department of Engineering, Nishinippon Institute of Technology, 1-11 Aratsu, Kanda, Miyakogun, Fukuoka 800-0394, Japan

Kazuo Ishii
Department of Life Science and System Engineering, Kyushu Institute of Technology, 2-4, Hibkinino, Wakamatsu Kitakyushu, Fukuoka 808-0196, Japan
E-mail: takemura@nishitech.ac.jp, kamei@nishitech.ac.jp, sanada@nishitech.ac.jp, ishii@brain.kyutech.ac.jp

Abstract

In Japan, expectations for the utilization of IT such as AI (Artificial Intelligence) and IoT (Internet of Things) are becoming greater and greater to address the shortage of working population due to the declining birthrate and aging population, and the depopulation of rural areas. There is a serious shortage of IT human resources, especially in rural areas, and I believe that IT human resource development is an urgent task to promote the super-smart society, Society 5.0, which is being promoted by the government. Thus, in order to promote business efficiency and productivity improvement, the utilization of IT in companies will become even more important[1].

Then, five universities, mainly in Kyushu and Chugoku regions, have been selected by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) for the enPiT-Pro project, and will collaborate to provide systematic and practical IT education programs to working people from local companies, with the aim of fostering advanced IT human resources. In this paper, we report on a class of the enPIT-everi program, the Agricultural IoT Education course.

Keywords: IoT, EnPIT-everi, Agriculture, Image processing, Deep Learning, PBL education.

1. Introduction

In Japan, expectations for the utilization of IT such as AI (Artificial Intelligence) and IoT (Internet of Things) are becoming greater and greater to address the shortage of working population due to the declining birthrate and aging population, and the depopulation of rural areas. There is a serious shortage of IT human resources, especially in rural areas, and we believe that IT human resource development is an urgent task to promote the super-smart society, Society 5.0, which is being promoted by the government. Thus, in order to promote business efficiency and productivity improvement, the utilization of IT in companies will become even more important [1].

In the midst of the global trend of the fourth industrial revolution, the Kyushu-Chugoku region has the opportunity to realize the development of new key industries that it has been seeking for many years. In order to realize this, it is necessary for the various people already working in society to adapt to the new technological innovations and grow. In order to achieve this aim, we will develop an advanced human resource development program for working people, targeting distinctive industries in the region, complementing each other in a wide range of fields through university collaboration, and incorporating the latest artificial intelligence and robotics technologies into social implementation. Then, five universities, mainly in Kyushu and Chugoku regions (the University of Kitakyushu, Kyushu Institute of Technology, Hiroshima City University, Kumamoto University and Miyazaki University), have been selected by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) for the enPiT-Pro project, and will collaborate to provide systematic and practical IT education programs to working people from local companies, with the aim of fostering advanced IT human resources. In this paper, we report on one of the education courses in enPIT-everi (Education Network for Practical Information Technologies - Evolving and Empowering Regional Industries) program, the Agricultural IoT Education course.

2. enPIT-everi project

The enPIT-evri program aims to provide working adults who are already based in some industrial field with the opportunity to acquire the knowledge, skills, thinking, judgment, and organizational abilities described in the following policies by taking a systematic education program at the graduate school level, despite its short duration of 3 to 6 months. However, in consideration of the various motivations and course-taking methods of the participants, two policies for the completion of the program, "IoT Architect" and "IoT Engineer," will be established [2].

This educational program will support the development of distinctive industrial fields in the Kyushu and Chugoku regions. Companies operating in the northern part of Kyushu are interested in IoT in a wide range of industrial fields, although they are somewhat more interested in manufacturing. Therefore, this program has set up courses that take advantage of the characteristics of the five courses. The five courses are Smart Factory, Intelligent Car, Smart Life Care, Smart Agriculture, Forestry and Livestock, and Hospitality IoT.

In all courses, a case study lecture is provided as a preliminary lecture, and after that, the common lectures can be accessed via VOD (Video on Demand). Finally, there are practical laboratory exercises in each course. In this paper, we report on the "Practical laboratory exercises on agriculture" conducted as an exercise of the Smart Agriculture Course.

3. Practical laboratory exercises on Agriculture

In Practical laboratory exercises on agriculture, we conducted PBL-style problem-solving classes and practical assignments necessary for problem-solving. The classes are held on weekends and consist of 24 sessions (36 hours) in total. In the curriculum of the class, the following lessons were given in each lesson.

The exercises are mainly organized into four major themes: artificial intelligence exercises, image processing exercises, mechatronics (microcontroller programming) and PBL (Project Based Learning) exercises.

1. Introduction
2. Mechatronics practice 1:
3. IoT Exercise using Microcontroller 1
   15. Mechatronics practice 2:
4. IoT Exercise using Microcontroller 2
   Mechatronics Exercise 3:
5. IoT Exercise using Microcontroller 3
6. Artificial Intelligence Exercise 1:
   Introduction to Python
7. Artificial Intelligence Exercise 2:
   Preprocessing of training data (Numpy)
8. Artificial Intelligence Exercise 3:
   Deep Learning Recognizer
9. Artificial Intelligence Exercise 4:
   Time Series Predictor
10. Artificial Intelligence Exercise 5:
    Class determination using self-organizing maps [3]
11. Image Processing 1:
    Development Environment for Image Processing Programs, Basics of Image Processing

12. Image Processing 2:
 Color Detection Processing (Color, Grayscale, Binary)
13. Image Processing 3:
 Color Space (RGB, HSV)
14. Image Processing 4:
 Counting Objects (Labeling, Center of Gravity)
15. Image processing 5:
 Object detection (template matching)
16. Image processing 6:
 Object detection (Hough transform)
17. Problems in agriculture, forestry and livestock industry
18. Brainstorming on agriculture, forestry and livestock industry issues

19. PBL exercise 1
20. PBL exercise 2
21. PBL exercise 3
22. PBL exercise 4
23. PBL exercise 5
24. Presentation of results and evaluation

In order to practice in the real field, we prepared an experimental field (Fig.1, 2). In the experimental field, we prepared a hydroponic field and an open-air field for IoT practice. In addition, the power supply for the exercises is secured by using solar power generation.

3.1. *IoT Data acquisition experiment using Microcontroller*

We practiced how to use a microcontroller to obtain data such as temperature, humidity, and water content in a plastic greenhouse from the cloud via the Internet. The microcontroller is an Arduino microcontroller, and the students practiced how to build a system that can obtain information from the cloud using an MQTT server.

Figure 3 shows the temperature in the greenhouse for one month, obtained by using IoT devices. The data is acquired once every 30 minutes, and can be checked on a web browser.

3.2. *Artificial Intelligent Exercise*

The students have already learned the outline of programming techniques and artificial intelligence in the common courses. Therefore, in this exercise, we used

Fig.1 Experiment green house field

Fig.2 Scene of Experiment (Installation of IoT devices)

Fig.3 Temperature data acquired by IoT devices CNN (Convolutional Neural Network) as a method of application in a farm. CNN is capable of recognizing objects in images. Therefore, in this exercise, we created an artificial intelligence architecture to recognize caterpillars in a farm. In this exercise, we used google collaboratory.

3.3. *Image Processing exercise*

For image processing, we aimed to practice using a miniaturized computer for use in agriculture. In this exercise, we used a small computer, Raspberry Pi, to build a system that can be equipped with CNN inference

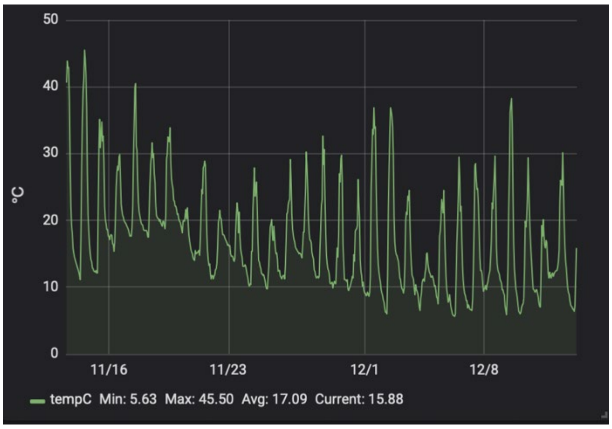


Fig.3 Temperature data in green house field

at the same time. We mainly learned about classical object recognition and filtering methods in image processing, and then performed CNN inference using the Neural Computing Stick.

3.4. *PBL exercise*

In the PBL exercise, the students brought the equipment they had used in the past and discussed how they could use it in the company with graduate students and advisors.

One of the students, who is in charge of town revitalization in the region, proposed a project that uses IoT devices to educate young people. There were many other ideas such as branding with IoT, and we were able to discuss guidelines for applying the technology in the class.

4. Summary

In this paper, we report on an exercise of agricultural IoT in the field of smart agriculture, which is one of the courses of enPIT-everi. In this exercise, we aimed at the development procedure of IoT devices using microcontrollers and how to apply artificial intelligence using image processing and CNN[4] to agriculture. In addition, we conducted a PBL exercise in which the participants, who are working people, think about how to apply the knowledge obtained in this exercise to their own work.

In the future, we would like to establish a system that enables us to carry out the exercise using an actual plastic greenhouse for a long period of time, and to obtain data for a long period of time and to use the data for inference by artificial intelligence.

Acknowledgements

This research was supported in part by enPIT-everi[1]. And, we are thankful to all participants of “Practically laboratory exercises on Agriculture” and helping students in Kyushu Institute of Technology and Nishinippon Institute of Technology

References

1. Enpit-everi homepage : “https://www.enpit-everi.jp”, in Japanese
2. Enpit-everi education course introduction: “https://www.enpit-everi.jp/course/ “in japanese
3. Kohonen, Teuvo (1982). "Self-Organized Formation of Topologically Correct Feature Maps". *Biological Cybernetics*. Vol.43 (1): 59–69.
4. Matusugu, Masakazu; Katsuhiko Mori; Yusuke Mitari; Yuji Kaneda (2003). “Subject independent facial expression recognition with robust face detection using a convolutional neural network”, *Neural Networks* vol.16 (5): 555–559.

Development of a Handy Autonomous Underwater Vehicle "Kyubic"

Toshimune Matsumura
Dept. of Human Intelligent System, Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka 808-0196, Japan
E-mail: matsumura.toshimune399@mail.kyutech.jp

Yuichiro Uemura
Dept. of Human Intelligent System, Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka 808-0196, Japan

Kentaro Yanagise
Dept. of Human Intelligent System, Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka 808-0196, Japan

Yoshiki Tanaka
Graduate School of Life Science Systems Engineering, Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka 808-0196, Japan

Yuya Nishida
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka 808-0196, Japan

Kazuo Ishii
Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka 808-0196, Japan

Abstract

Ocean is one of big challenging and extreme environments, and hard for human to access directly. As the tool for ocean survey, Autonomous Underwater Vehicles: AUVs are expected and developed from '80s. The recent rapid progress of computer and information technologies makes the development of AUVs easier and more practical. We had developed a handy AUV "Kyubic" for the observation of shallow water and artificial structures. In this paper, we describe the system architecture of Kyubic and the experimental results in Underwater Robotics Competition in Okinawa 2020.

Keywords: Autonomous Underwater Vehicle, System design, Robot competition.

1. Introduction

The ocean is one of the most difficult environments for humans to access directly. Autonomous Underwater Vehicle (AUV) are expected to be a tool for marine research [1]. Currently, there is a growing need for underwater robots for surveys and work in shallow water, as well as for surveys and work on artificial structures such as coastal areas, dams, and bridge piers. However, AUV used for marine research require a large support vessel equipped with a crane, which makes it difficult for a few people to operate. In addition, works at shallow water are still often performed by divers or small Remotely Operated Vehicles (ROV). Working at a depth of several tens of meters is a heavy workload for divers. In addition, ROV are limited by their tethers. Therefore, there is a need for a compact and lightweight AUV that can be operated by a few people. Since it is easy to develop a small and lightweight AUV, underwater robot competitions are being held with the aim of activating research and education in the underwater field. In particular, since the Underwater Robotics Competition in Okinawa is a competition held in the actual sea area, it is required to operate an AUV suitable for the actual

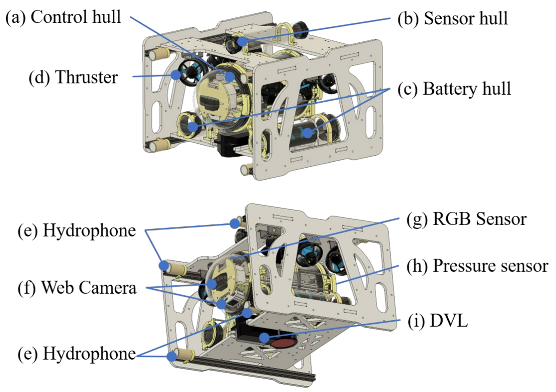


Fig.1 The 3D design of Kyubic

Table 1 Specifications of Kyubic

Structures	Acrylic Pressure Hulls × 4 H : 400[mm] W : 550[mm] L : 570 [mm] Weight : 32 [kg] 15[m] depth pressure resistant
Actuators	Thrusters (Blue Robotics T200) × 6
Batteries	Li-ion 14.8[V], 18 [Ah] ×2
Computer system	PC：NUC 817BEH CPU：Intel Core i7-8559U RAM：32 [GB] SSD：2 [TB]
Communications	Ethernet Wireless LAN Optical LAN
Sensors	USB Camera × 2 IMU(CSM-MG100) DVL(Pathfinder) Hydrophone (SPU0414HR5H-SB) ×4
Software	MATLAB & SIMULINK ROS

environment. The paper explains the configuration of the newly developed AUV "Kyubic" as one of the small underwater robots and reports on its operation in Underwater Robotics Competition in Okinawa 2020.

2. The hovering type AUV Kyubic

2.1. *Kyubic specifications*

In our laboratory have developed "AquaBox" and "DaryaBird" as small and light AUV systems [2][3]. The AUV "Kyubic" aims to be smaller and lighter than conventional robots, and its development concept is as follows.

(i) Small size and lightweight, enabling operation by a few people
(ii) Boxed structure with additional options depending on the mission

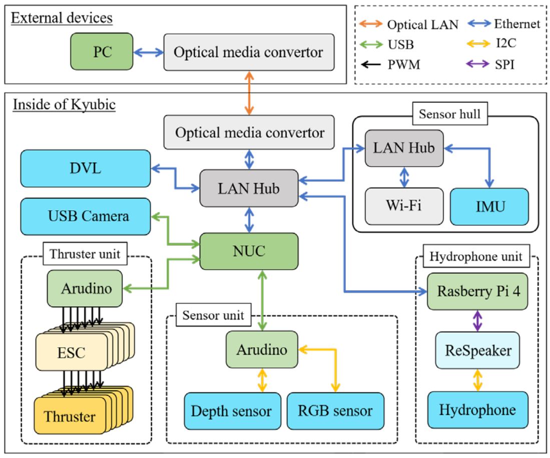


Fig.2 System architecture of Kyubic

(iii) Modular structure for easy maintenance
(iv) AUV and ROV modes can be selected depending on the mission

Table 1 shows the specifications of the AUV "Kyubic". The AUV has a boxed structure and is compact and lightweight.

2.2. *Hardware*

Fig. 1 shows the layout of Kyubic. The AUV is framed of four pressure vessels. The power circuit, motor driver, microcomputer, and control PC are installed in an 8 inch pressure vessel (a). Wi-Fi and IMU are installed in the 2 inch pressure vessel (b). The control and drive batteries are installed in a 3 inch pressure vessel (c). In addition, this AUV is installed with six thrusters (d), four hydrophones (e), two Web cameras (f), an RGB sensor (g), a pressure sensor (h), and a DVL (i). Fig. 2 shows the system architecture of Kyubic. This AUV is intended to be a versatile testbed and software development. For this purpose, a small computer NUC with high processing capacity is installed in a pressure vessel. The AUV uses information from sensors such as the DVL, IMU, and pressure sensors to control the autonomous mode. The depth sensor, RGB sensor, and ESC are controlled by Arduino, and the hydrophone is controlled by Rasberry Pi4. These systems are modularized, and modules can be easily added or removed. In the previous DaryaBird, it was not possible to turn off the power of the sensor by itself. Therefore, it was necessary to turn off all the systems when a problem occurred. To solve this problem, Kyubic has a system configuration that allows the power supply of the sensor to be controlled by itself. In addition, DaryaBird control with control PC directly the motor driver. Therefore, the user could not control the robot if the control PC system terminated improperly. On the other hand, Kyubic has an Arudino between the control PC and the motor driver. Therefore, when the command value from the control PC is not received for a certain period of time, the power supply of the motor is stopped. The thruster control unit is modularized, and the modules can be easily used individually or installed on a robot, resulting in a highly versatile system configuration. The same applies to the sensor unit and hydrophone unit.

2.3. *Software*

Kyubic's control system uses Mathworks' MATLAB / Simulink and ROS control methods [4]. In particular, the Robotics System Toolbox supports ROS, which enables highly functional and fast development using MATLAB and ROS programs. The Parallel Computing Toolbox is a script-based and Simulink model-based program which can build multiprocess and parallel computations. Stateflow is a tool for computing state transition control methods. Fig. 3 shows a architecture of the software. Simulink is divided into 1-5 models (Sim1-Sim5), each of which exchanges data through a ROS network. Sim1A gets the attitude data of the AUV from IMU and the positioning data from GNSS. Sim1B gets the ground speed from the DVL and the depth data from the depth sensor. Sim2 handles the PID control of the thrusters, and Sim3 handles the image processing. Sim4 records data and does action strategies. Stateflow is used for the action strategy, making it easy to visually understand the values and transitions of the process being executed. Kyubic uses positional control to perform waypoint tracking. Waypoint tracking is a method in which a specific position on a defined path is specified and the robot moves to pass through that position. For the path, the CSV data includes the waypoint number, target position (x, y, z), target attitude of the AUV (roll, pitch, yaw), arrival judgment threshold, timeout, and mission number. Hence, load the CSV as external data. The previous DaryaBird needed to write the target value directly to stateflow. However, Kyubic only updates the csv data and does not need to change the stateflow. In addition, this AUV can control the operation values of Sim4 by GUI. Fig.4 shows the

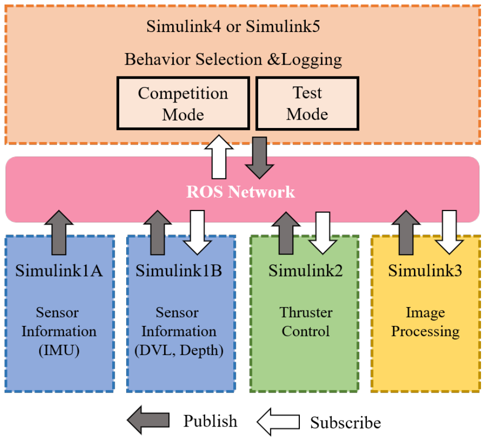


Fig.3 Software architecture of Kyubic

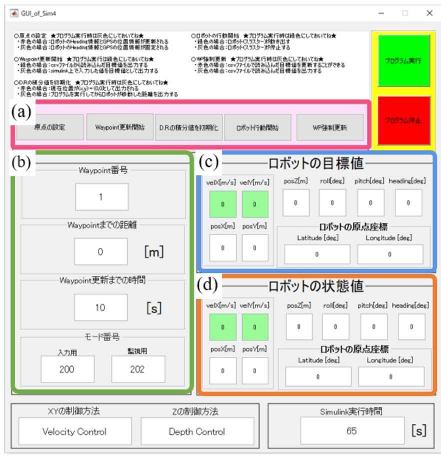


Fig.4 Sim4 operation GUI

GUI of Kyubic. (a) in the GUI is an item that the user inputs before the robot starts its action, such as the origin setting and the command to start the robot's action. (b) shows the set waypoint number, distance to the waypoint, and mode number. (c) shows the target value of the robot, and (d) shows the state value of the robot. This GUI can be used to avoid human errors during robot operation and to monitor the robot status for efficient operation.

3. Sea trial in shallow water

3.1. *Underwater Robotics Competition in Okinawa*

Kyubic was operated at the Underwater Robotics Competition in Okinawa , and the results are described. The Underwater Robotics Competition in Okinawa is the only competition in Japan that is held in actual marine

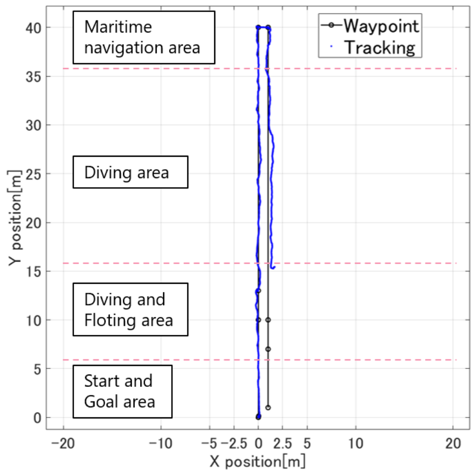


Fig.5 Waypoint and Tracking

environments. The rule of the AUV department is that it navigates there and back in a straight line of about 40 meters in shallow water in five minutes. In the middle section, there is a diving area, and the AUV needs to dive and navigate 20 meters. For details of the rules, please refer to [5].

3.2. *Navigation results of the competition*

Kyubic navigated in the surge direction throughout the entire area. This AUV dived to a target depth of 0.5 m in the diving area and turned for the turnaround in the maritime navigation area. At the turnaround point, the AUV navigated 1[m] in the x direction to prevent the AUV from involving the safety rope. Fig.5 shows the waypoints and the trajectory of Kyubic. In this competition, Kyubic was able to navigate to the turnaround dive area in five minutes. For the trajectory, this AUV achieved path following with a maximum error of 1.5 [m] in the y direction. The cause of the error in the trajectory is considered to effect of tidal currents and cumulative error due to dead reckoning. Fig. 6 shows the control performance of velocity of surge direction, depth, and heading. The graph of the velocity in the surge direction show an estimated value of about 0.2 [m/s] against the target velocity of 0.6 [m/s]. It is assumed that this result was due to inadequate adjustment of PID and inferior to control performance. In the depth direction, the tracking is slow and overshoots the target value, show that convergence is taking time. It is assumed that the PID was not adjusted sufficiently for depth. In the heading, there is a slight overshoot, however, it is generally in correspond with the target value.

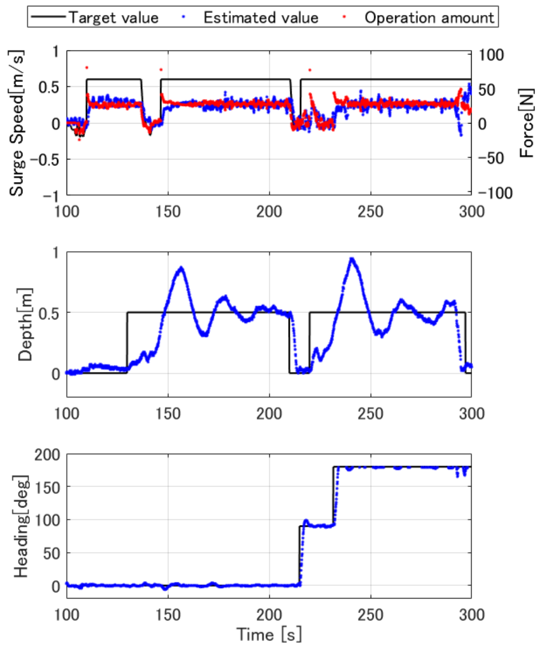


Fig.6 Target value and Estimated value

4. Conclusion

We developed a small underwater robot system "Kyubic" for observation in shallow water. In addition, we were able to confirm its basic performance in a competition in shallow water. In the future, we will improve the control performance and implement acoustic communication and image processing using Kyubic.

References

1. Yuya Nishida, et al., Underwater Platform for Intelligent Robotics and Its Application into Two Visual Tracking Systems, Journal of Robotics and Mechatronics, Vol.30, No.2, pp. 238-247, 2018.
2. Satomi Ohata, et al., Small Underwater Robot Systems "AquaBox" for Shallow Water Observation, The Japan Society of Naval Architects and Ocean Engineers, Vol.4, pp.513-516, 2007.
3. Wataru Hori, et al., An AUV Platform for Robot Behavior Development: Twin-Robot "DaryaBird" and "OCTA", and Their' Mission Strategy, AUV&ONR's 17th Annual RoboSub Competition Journal Paper, 2014.
4. Takashi Sonoda, et al., System Development of AUV's Sampling Device Controller Employing MATLAB/Simulink Toolboxes, ICAROB, 2019.
5. Underwater Robotics Competition in Okinawa, http://www.robo-underwater.jp/2020/rchp/JPN/index.php , (2020.12.25).

A Greenhouse Project toward Smart Agriculture

**Kazuo Ishii[1], Yuya Nishida[1], Shinsuke Yasukawa[1], Kanako Shirahashi[1],
Yasunori Takemura[2], Takayuki Matsuo[3]**

*[1]Center for Socio-Robotic Synthesis, Kyushu Institute of Technology, 2-4, Hibkinino, Wakamatsu
Kitakyushu, Fukuoka 808-0196, Japan*

*[2]Department of Engineering, Nishinippon Institute of Technology, 1-11 Aratsu, Kanda,
Miyakogun, Fukuoka 800-0394, Japan*

*[3]National Institute Technology, Kitakyushu College
2-20-1, Shii, Kokuraminami-ku, kitakyushu-shi, Fukuoka, Japan*

*E-mail: ishii@brain.kyutechg.ac.jp, matsuo@kct.ac.jp, takemura@nishitech.ac.jp, sonoda@nishitech.ac.jp,
y-nishida@brain.kyutech.ac.jp, s-yasukawa@brain.kyutech.ac.j**

Abstract

In the future estimation of 2050, the food demand increases 70% and the production decreases 15% caused by global warming, and farmer population be 1/5 of current workers. Until 2050, we need an agricultural system of twice production with the same farmland area and 5 times effective operations. To realize the sustainable society, smart agriculture including robot technology, AI, IoT is one of the solutions for food issues. We have been working for a greenhouse project under enPiT-everi for implementation of robotic, AI and IoT technologies to agriculture. In this paper, we introduce the project and robotic applications.

Keywords: tomato harvesting, agriculture robot, smart agricultu

re

1. Introduction

Due to the frequent occurrence of intense heat and heavy rain, the harsh global environment caused by global warming, world food production will decrease by 15% to 2.9 billion tons in 2050[1]. Meanwhile, the world population has increased to her 8.6 billion and food demand is expected to be a serious food shortage that will reach 5.8 billion tones [2]. Because of the progress of desertification and global warming, rapid expansion of excellent agricultural land cannot be expected. Therefore, in order to eliminate food shortages, it is necessary to obtain twice the yield of the same area as before. In addition to the intense heat and heavy rain limiting people's outdoor activities, it is estimated that the number of farmers in Japan will decrease sharply to one-fifth due to the declining birthrate and aging population [3], resulting in a serious labor shortage. It is expected that the production will not be able to meet global food demand. To realize the sustainable society, smart agriculture including robot technology, AI, IoT is one of the solutions for food issues.

On the other hand, the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) planned a program to form practical education networks for industry-academia collaboration in order to strengthen the function of developing human resources who can solve specific problems in society by making advanced use of information technology, and practice problem-based learning “enPiT" [4] with the aim of promoting various education and spreading it nationwide. Our university, Kyushu Institute of Technology, proposed an educational program named “enPiT-everi” (Evolving and Empowering Regional Industries) [5] to grow up small and middle size companies in Kyushu region together with Kitakyushu-city university and so on. We have been working for a greenhouse project for evaluation of robotic, AI and IoT technologies as a part of enPiT-everi. In this paper, we introduce the project and robotic applications for agriculture.

2. Smart Agriculture

Smart agriculture is a new style of agriculture that utilizes robot technology and information and communication technology (ICT) to promote labor saving, precision, and high-quality production. In Japanese agriculture fields, there are still many tasks that depend on manpower and can only be done by skilled workers. The labor saving, securing manpower, and reducing the burden in Japanese agriculture are important issues. By utilizing smart agriculture that makes full use of advanced technology, it is possible to overcome the issues in agricultural work, secure new farmers, and improve cultivation technology.

(1) Labor saving and labor reduction of agricultural work
Japanese agriculture is facing a serious labor shortage due to the aging of individual farmers. It is required to support such hardships in Japanese agriculture by utilizing ICT.

2. Succession of agricultural technology
The second issue is the transfer of cultivation technology to new farmers. In the old system, the knowledge is transferred in families or small groups. The shortage of human resources makes difficult to inherit expert knowledge in agriculture in the old fashion.

3. Improvement of food self-sufficiency
Japan's food self-sufficiency rate (calorie basis) was about 40% in FY2018, and imports far exceed domestic production. In order to increase the yield and increase the self-sufficiency rate in the face of the above-mentioned shortage of human resources, automation by IoT, AI and robots is indispensable for reliably growing agricultural products with a small number of workers.

3. enPiT-Pro and enPiT-everi

The Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) planned a program to form practical education networks for industry-academia collaboration in order to strengthen the function of developing human resources who can solve specific problems in society by making advanced use of information technology, and practice problem-based learning “enPiT" with the aim of promoting various education and spreading it nationwide. From 2017, as enPiT-Pro, MEXT supports the universities to develop and implement a systematic, advanced and short-term practical education program for working adults, mainly in the field of information science and technology, in collaboration with industry and multiple universities. By disseminating the program widely throughout Japan, MEXT also supports graduate school to reform efforts aimed at contributing to strengthening the re-learning function of working adults in the entire field in Japan.

Our university, Kyushu Institute of Technology,

proposed an educational program named “enPiT-everi” (Evolving and Empowering Regional Industries) to grow up small and middle size companies in Kyushu region together with Kitakyushu-city univ., Kumamoto univ., Miyazaki univ. and Hiroshima-city univ.. In the enPiT-everi program, we provide practical educational programs for working people in the Kyushu and Chugoku regions to acquire new technologies such as artificial intelligence and robot technology. The educational themes are specialized for manufacturing industry, automobile industry, nursing care industry, agriculture, forestry and livestock industry, tourism industry. The enPiT-everi develops human resources who can promote the introduction to companies such as IoT, AI and robots. One of the educational themes is the implementation of IoT for agriculture, and we have been developing the experimental greenhouse with IoT and robotic technologies.

4. Greenhouse project in enPiT-everi

The greenhouse developed for smart agriculture is shown in Fig.1. The house size is 10m x 20m and the half area is designed for cultivation with soil and the other is for hydroponics.

(a) Overview of the house

(b) Inside the house

Fig.1 The experimental greenhouse for agriculture IoT exercise.

Currently, as agricultural IoT Sensors, temperature, moisture data, pH regular measurement of soil, regular measurement of temperature, humidity and illuminance in the greenhouse, photography of plants, have been implemented. As actuators for control air condition and soil condition, opening and closing the greenhouse walls, spraying mist, water supply control by solenoid valves, water circular system in hydroponics, the system for purification of rainwater and control of nutrients are under implementation.

Using the experimental house, we will encounter to various problems in agriculture. For example, recognition of green worms imitating the leaf vein, disease of plants, harvesting and so on. Transmitting data to cloud data base is also one research topics with limited communication like Sigfox, Lora, etc. The detail of exercise for Agricultural IoT course is presented in the paper [5] in this conference. Also the Tomato-Harvesting-Robot Competition was held in the greenhouse from 7th competition in 2020 [6].

(a) green worms imitating the leaf vein

(b) Disease detection

(c) Insect counting

Fig. 2 Research issues related to the greenhouse.

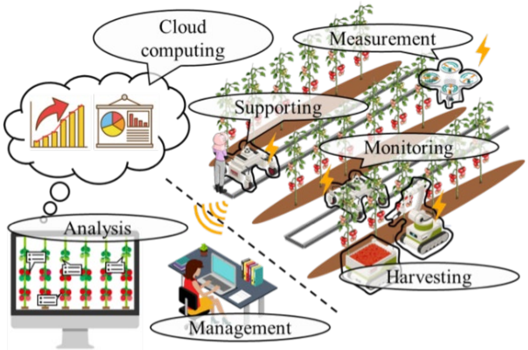


Fig. 3 The future concept of the greenhouse project.

5. Summary

We have been working for experimental greenhouse under enPiT-everi educational program. The agricultural IoT devices are implemented toward smart agriculture. Future vision of the greenhouse is shown in Fig. 3. The house system will include IoT sensors and actuators to control environment, robots for monitoring, harvesting, and transportation. Using the greenhouse, the new research topics with actual demands will come up toward smart agriculture.

References

1. https://www.maff.go.jp/j/tokei/sihyo/index.html
2. https://www.maff.go.jp/primaff/seika/attach/pdf/200403_2019_01.pdf
3. https://www.maff.go.jp/primaff/seika/fsc/faccess/a_map.html
4. http://www.mext.go.jp/a_menu/koutou/kaikaku/enpit/index.htm
5. https://www.enpit-everi.jp
6. Y. Takemura et al., Smart Agriculture IoT Education Course in enPiT-everi (Education Network for Practical Information Technologies - Evolving and Empowering Regional Industries), Proc of ICAROB2021, to be appear, 2021.
7. K. Ishii et al., Tomato-Harvesting-Robot Competition Towards Smart Agriculture, Proc of ICAROB2021, to be appear, 2021.
8. T. Matsuo, et al., Tomato-Harvesting Robot Competition: Aims and Developed Robot of 6th Competition, Proc of ICAROB2021, to be appear, 2021

The Security Challenges with The Widespread Use of IT Infrastructure in ICS

Kuan-Ming Su
Department of Electrical Engineering / Institute of Computer and Communication Engineering,
National Cheng Kung University
No.1, University Rd., East Dist., Tainan City 70101, Taiwan

I-Hsien Liu
Department of Electrical Engineering / Institute of Computer and Communication Engineering,
National Cheng Kung University
No.1, University Rd., East Dist., Tainan City 70101, 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 70101, Taiwan
*E-mail: kmsu@cans.ee.ncku.edu.tw, ihliu@cans.ee.ncku.edu.tw, jsli@mail.ncku.edu.tw**

Abstract

The communication established by Ethernet is becoming more and more common in the industrial control systems (ICS), and it brings not only pros but also cons like vulnerabilities from information technology. We generalized a procedure of attacking an Ethernet-enabled ICS and implemented it to the real industrial system we obtained. The procedure gets the information and access of the devices in the ICS, like identifying the manufacturer of programmable logic controllers (PLCs) and overwriting the configuration of PLCs.

Keywords: Industrial Control Systems (ICS), Programmable Logic Controller (PLC), Information Technology (IT), Network Security, Ethernet

1. Introduction

In the modern industrial control system (ICS), information technology (IT) is getting more and more widespread in the operation technology (OT) field because of Industry 4.0 [1], and serial communication is becoming incompetent to meet the demands. Therefore, in terms of communication of IT in ICS, Ethernet is coming to the most popular one. Ethernet is commonly used in ICS with IT. Comparing to serial connection, Ethernet has much more flexibility and scalability. For example, to access and manage multiple devices, all you need to do is connecting your engineering workstation (EWS) to the network where devices are. Also, Ethernet allows hundreds of devices to communicate with each other. In addition, there have been many industrial protocols supporting Ethernet and Internet Protocol.

Although Ethernet meets the demands of IT and OT, it also brings vulnerabilities to ICS. Considering the cost of building information security, most ICS defense mechanisms only have an external firewall, isolation from the office network, or complete independence from other networks. Moreover, for the purpose of operation stability, many running operation systems are not updated to the latest version including the security patches. Therefore, those devices are very vulnerable to malware. Let us take TSMC for example. TSMC is the most advanced integrated circuit manufacturer in the world, and they have the top of cyber security standards to protect the intellectual property and factory operation. However, an accident occurred in 2018. Because of the operational errors during the software installation on the new equipment, after the new equipment hooked up to the internal network, the ransomware infected other computer systems and fab tools and caused about $170 million US dollars losses [2].

In order to show how fragile information security inside the ICS network is, we generalized a procedure to attack ICS networks, and we carried it out on a real ICS which was used before. Eventually, we compromised the PLC in the ICS network with MODBUS TCP packets. We successfully read and wrote the registers in the PLC and stopped the running PLC.

2. Background

There are devices controlled in ICS, and there must be some methods for the controller to communicate with each other to make those devices work together. We can divide it into the physical connection part and the communication protocol part. For the physical connection in ICS, there are several standards like RS-232, RS-422, RS-485, and Ethernet. Among those standards, Ethernet is the focus of this discussion due to the trend of IT. For the communication protocols in ICS, there are various protocols based on different standards. Many relatively modern protocol versions are based on Ethernet. Some of them are based on Internet Protocol, such as MODBUS TCP and Ethernet/IP, and some of them are not, such as EtherCAT and PROFINET. For the following discussion, we will focus on MODBUS TCP.

2.1. *Industrial Ethernet*

Fig. 1. M12 connector [3]

In addition to Ethernet interfaces on the ICS devices like PLC and HMI, Ethernet switches and cables are also needed to build the network, and there are some differences between common and industrial Ethernet products. Depending on the different environments, the industrial cables may have high-quality foil and braid to protect data transmission from EMI [4], use cable jackets with different materials like FEP and TPE for durability [5], or have M12 and M8 connectors (see Fig. 1) instead of common 8P8C (often miscalled RJ45) in order to be waterproof. Besides the difference of the connectors, the industrial Ethernet switches have other features comparing to common switches. Let us take Cisco industrial Ethernet 4000 series switches (IE-4000) [6] for example. IE-4000 can work in extreme environments and temperature range (-40 to 70 Celsius), has a durable design, support power over Ethernet up to 240W, and so on.

To sum up, the main difference between industrial and common Ethernet products is the durability in different environments and the features in use and management. There is no change to the data transmission standards.

2.2. *MODBUS Messaging on TCP/IP*

MODBUS is a popular communication protocol in industrial environments because it is opened and does not need a license fee. The protocol data unit (PDU) of

TCP/IP Layer	Protocol
Application Layer	Modbus TCP PDU
	Modbus TCP Header
Transport Layer	Transmission Control Protocol (TCP)
Internet Layer	Internet Protocol (IP)
Link Layer	Ethernet (IEEE 802.3)

Fig. 2. MODBUS on TCP/IP

Transaction Identifier	Protocol Identifier	Length	Unit Identifier	Function Code	Data
2 bytes	2 bytes	2 bytes	1 byte	1 byte	n bytes
MODBUS TCP Header				MODBUS TCP PDU	

Fig. 3. MODBUS TCP Data Frame

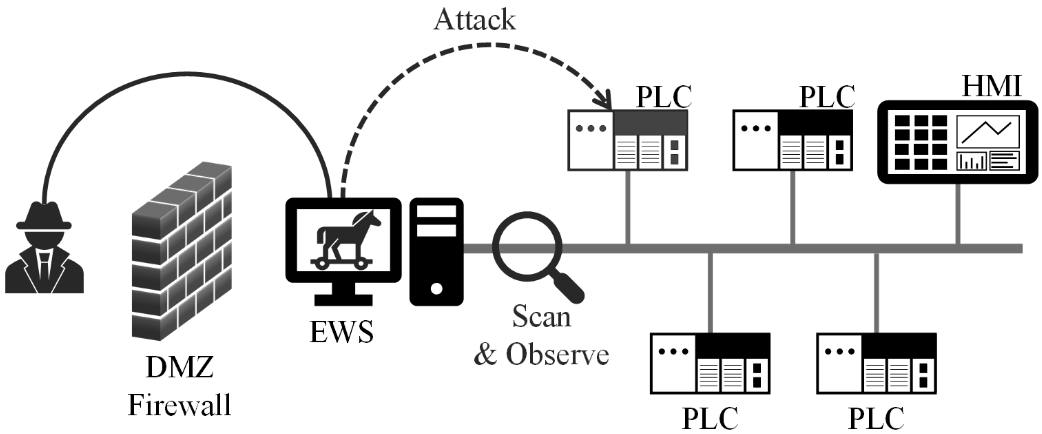


Fig. 4. The Scenario and Procedure of Attack in an Ethernet-enabled ICS Network

MODBUS is simple. It only consists of function code and data. Depending on the function codes and request or response, the following data structure is different. For example, function code 03 is reading multiple holding registers, the request data structure is composed of 2 bytes starting address and 2 bytes quantity of registers, and the response data structure is composed of 1 byte following data size and the value of registers [7].

MODBUS Messaging on TCP/IP, or MODBUS TCP for short, as the name suggests, is MODBUS implemented on the TCP/IP (see Fig. 1), and the port number of it is 502. The data frame of MODBUS TCP is in Fig. 2. Transaction identifier is for pairing the request and response. Protocol identifier is used for internal system multiplexing, and value 0 stands for MODBUS protocol. The length field is the byte count of the rest part including unit identifier, function code, and data field. Unit identifier is used for the internal system routing purpose. Function code and data field are MODBUS PDU [8].

3. Attacking an Ethernet-enabled ICS Network

3.1. *Scenario*

Since ICS with IT brings vulnerabilities, the attacker could use those vulnerabilities to inject a backdoor into the computers in the ICS, bypass the information security protection measures, and perform malicious operations. For example, an attacker could implant a Trojan horse into the EWS in the ICS via social engineering, then bypass the firewall and invade into the ICS network to do whatever he wants. Therefore, assuming that the attacker is able to access the ICS network with some method like backdoor, we generalized a procedure to get the information of the ICS and attack it.

3.2. *Procedure*

The first thing to do is scanning the internal network. In most internal ICS networks, due to the weakness of IT security in ICS, there is no protection method like IDS and IPS. Therefore, by scanning the network, the attacker can observe the information in the ICS network such as enabled services, subnets' range, the number of devices, manufacturer of devices, and so on. After obtaining the information of the ICS network, according to the information, the attacker can formulate detailed attack methods such as man-in-the-middle attack with ARP spoofing to compromise the information security and operation safety in the ICS.

4. Case Study

The ICS that we are going to demonstrate the procedure on is the same as our previous study [9]. The ICS is composed of tens of PLCs to control the field equipment and one computer as the HMI to gather data and control PLCs. All of them are connected to an Ethernet switch.

First, we used a tool called Nmap to scan the ICS network. In the result, we can see the information of online devices, such as IP address, enabled services, and MAC address, and we can identify that the manufacture of PLC is Telemecanique Electrique which is Schneider Electric from the MAC address. The communication

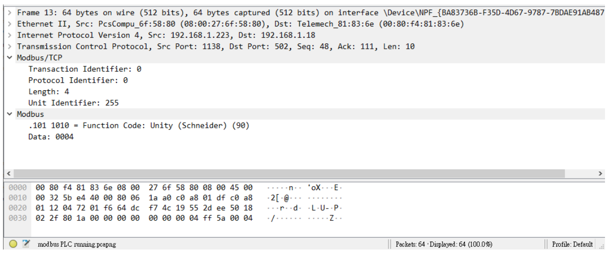


Fig. 5. The MODBUS TCP packets with function code 90

protocol of Schneider PLC is MODBUS, so we tried to read the holding registers with standard MODBUS function code 03 with the MODBUS TCP testing tool, and it works. Furthermore, the writing holding registers' function code 16 also works.

In our previous study, we used the IDE called TwidoSuite to perform some malicious operation on the PLC. This time, we use a tool called Wireshark to record the MODBUS TCP packets. Unlike the standard MODBUS TCP, the function code of those packets is 90 (see Fig. 5.), which is manufacturer defined. It is not able to just do the replay attack because of the authentication mechanism in the data bytes of MODBUS. However, we found some rules of it. We successfully established the connection and commanded the online PLC to stop.

5. Conclusion

In this paper, we discussed the weakness of IT security in the ICS network, generalized an attacking procedure for the ICS network, and implemented it to the real industrial system to support the argument. It turns out that we can easily obtain the information of the ICS with the procedure in the certain scenario without the knowledge of the ICS. As the widespread use of IT infrastructure in ICS, we must pay more attention to the cyber security in ICS.

Acknowledgements

This work was supported by the Ministry of Science and Technology (MOST) in Taiwan under contract numbers MOST 109-2218-E-006-014- and MOST 109-2218-E-006-510-.

References

1. Martin Wollschlaeger, Thilo Sauter, Juergen Jasperneite, The Future of Industrial Communication: Automation Networks in the Era of the Internet of Things and Industry 4.0, *IEEE Industrial Electronics Magazine*, vol. 11, pp. 17-27, 2017.
2. TSMC, *TSMC Details Impact of Computer Virus Incident*, TSMC, Hsinchu, 2018.
3. eicos, *M12 Plug Connector - M12 Male Straight Connector without Cable,* eicos, Online. Available: http://eu.eicos.com/accessories/connectors/m12-male-straight-connector/. Accessed 14 12 2020.
4. Lapp Tannehill, *LAPP ETHERLINE® 2 Pair: CAT5 Flexible - 22 AWG - Green*, Lapp Tannehill, Online. Available: https://www.lapptannehill.com/etherline-2pair-cat5e-flexible-22awg-green. Accessed 15 12 2020.
5. BELDEN, *Category 6 Cable - 7931A*, BELDEN, Online. Available: https://www.belden.com/products/cable/ethernet-cable/industrial-ethernet-cable/7931a. Accessed 15 12 2020.
6. Cisco, *Cisco Industrial Ethernet 4000 Series Switches Data Sheet*, Cisco, 2020.
7. Modbus Organization, *MODBUS APPLICATION PROTOCOL SPECIFICATION V1.1b3*, Modbus Organization, Hopkinton, 2012.
8. Modbus Organization, *MODBUS MESSAGING ON TCP/IP IMPLEMENTATION GUIDE v1.0b*, Modbus Organization, Hopkinton, 2006.
9. Kuan-Ming Su, I-Hsien Liu, Jung-Shian Li, The Risk of Industrial Control System: Programmable Logic Controller Default Configurations, in *International Computer Symposium 2020*, Tainan, 2020.

A Communication System with Equipment's Characteristics

Chia-Chun Lai, I-Hsien Liu, Chi-Che Wu, 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 70101, Taiwan

Chuan-Gang Liu*
Department of Applied Informatics and Multimedia, Chia Nan University of Pharmacy & Science
No.60, Sec. 1, Erren Rd., Rende Dist., Tainan City 71710, Taiwan
*E-mail: {cclai, ihliu, ccwu1988}@cans.ee.ncku.edu.tw, jsli@mail.ncku.edu.tw, chgliu@mail.cnu.edu.tw**
www.cans.ee.ncku.edu.tw, www.cnu.edu.tw

Abstract

Over the past few years, applications on the internet have grown rapidly. In order to identify a specific device, we usually use the information which attains from network packets such as IP address, MAC address and communication port, etc. However, using this kind of information is not enough to identify precisely. Therefore, our research focuses on characteristics of communication devices that can identify precisely and also create a communication system which is able to imitate these characteristics.

Keywords: ICS, Communication, Network, Device Identification, Cyber Security

1. Introduction

In the last few decades, communication technology has grown rapidly and the internet is widely used in our daily lives. The progress result in devices that can connect to the internet have an explosive increase. Compare to the past, some information such as IP address, MAC address, and communication port is commonly used to identify a specific device on the internet. However, using this kind of information is not enough to identify comprehensively because the information can simply be imitated by any other devices. In our research, we focus on the characteristics of communication device. And we also provide a mechanism that can analyze these characteristics and generate configuration references which other devices can apply. With the customized communication module and the references mentioned above, we can make other devices imitate the behavior of the analyzed device and provide a better effect on device emulation.

2. Background

In this section, we are going to discuss some researches and methods which are commonly applied to perform device identification.

2.1. *Packet Header*

When devices communicate with each other, they must use the same protocol. Thus, allow them to understand what the other side transmitted. Nowadays, the most commonly used protocol is the TCP/IP protocol suite. This protocol suite includes a great deal of widely used protocols. Some well-known protocols in the TCP/IP suite are IP, TCP, UDP, etc., which are performed in different layers. These protocols have their own specific packet header when they are used.

Some information in these headers or the combination of this information in the different header can be used to identify which device sends this packet. In the next few paragraphs, we are going to explain the detail of how information in headers are used for device identification.

2.1.1. *IP header*

IP is the principal protocol in network communication. It can simply be divided into two versions: IPv4 and IPv6. Although the packet header of IPv4 and IPv6 is different, they still have some similarities. In the IP header, the major fields are source address and destination address, both of them exist in IPv4 and IPv6 header. The main function of these headers is to identify which device sends out this packet and which device should this packet send to. With these two pieces of information, we could simply identify a specific device, but it's not enough for precise identification. Some networks may use NAT (Network Address Translation) and many devices would share a public IP address. In this situation, using an IP address as the only identifier is not enough for device identification.

2.1.2. *TCP header*

TCP often uses on reliable transmissions such as a webpage, video stream, and file transfer. In the TCP header, there are two fields that can be used for device identification. These two fields are the source port and destination field, but we can't use them independently. We need to combine port information with IP information mentioned in the previous paragraph to carry out a better performance for device identification.

2.1.3. *Ethernet frame*

Ethernet protocol is a protocol that belongs to the data link layer. Recently, it is the principle protocol in our daily internet environment. Each Ethernet packet has an Ethernet frame that contains a header. An Ethernet header has four major fields: destination MAC addresses, source MAC addresses, Ethertype and IEEE 802.1Q tag or IEEE 802.1ad tag.

2.2. *Traffic Patterns*

Various devices will generate different traffic patterns depending on their specific needs. Some researches use traffic patterns to identify devices. Research from Hiroki KAWAI has shown that they can analyze traffic patterns and identify devices in different categories[1]. Also, there are some researches use machine learning to identify IoT devices which can classify all devices that connect to the network into some specific types[2,3,4].

3. Characteristics of Devices

Each device has its special characteristics when they are communicating. In our research, we focus on the behavior of devices when they respond or transmit packets. In our previous research, we obtain that the latency of the response packet between each device is unique. This discrepancy can cause by various factors in different layers.

3.1. *Physical Layer*

The devices produced by the same factory with different models will have their unique characteristics due to the variation of assembly lines. If we focus on the same model devices, they still exist unique characteristics because of the standard error of the machine which is allowed by its supplier. Even more, the manufacturing tolerance which is set by the manufacturer will also increase the diversity of characteristics in devices.

3.2. *Transport Layer*

When devices communicate on the internet they have to determine which protocol to be used in the transport layer. Using different protocols will result in respective behavior therefore generate a distinctive characteristic of devices. For instance, in a general WAN internet environment, if we use UDP protocol for communication, it might provide lower latency than using TCP protocol5. This situation shows that which protocol is used will have a significant impact on the characteristics of devices.

3.3. *Application Layer*

Every device has its own operating systems. Also, the executive application on the device would also affect the behavior. Even the same function programmed in a different language would lead to various characteristics of devices. Furthermore, regardless of having the same model device, if they are running in a different version of firmware, they may also increase the diversity of characteristics. In conclusion, the application layer has many factors that will affect the behavior of the device.

4. Experiment and Result

In our research, we propose a communication system that can analyze the characteristics of devices and then use these data to perform characteristics emulation of devices. We use a Schneider programmable logic controller (PLC) to be the device where we want to emulate and use a Windows 10 PC responsible for analyzing other devices. Owing to the device we used, the experiment will use a specific protocol which is commonly used in industrial control system called Modbus. Also, our communication system is programmed in C++ and run on Linux which can perform a better latency control of communication. Fig. 1. is the overview of our system. We can separate our system into two parts. In Fig. 2 and Fig. 3, we can see it have two major chains. The first chain is responsible for device analyzation and the second chain is responsible for emulation of the device.

4.1. *Analysis Side*

On the analysis side, we use a computer to send the response request to the device being analyzed. When the device response to our computer, we will record its round-trip time. Then we sorted out these data into a characteristic model and arrange a configuration file that can be used for the emulation side.

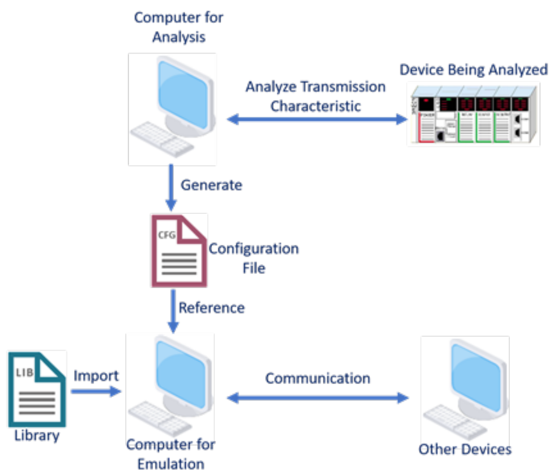


Fig. 1. Overview of our specific communication system

4.2. *Emulation Side*

On the emulation side, a device can either run our specific communication program on it or import our customize program library. Both of them can load the characteristic model generated by the analysis side. Furthermore, we add in a specific parameter called the loading parameter. Using the loading parameter, we can imitate the characteristics of the device. Meanwhile, the device is at working status.

4.3. *Analysis result*

In order to analyze the result of our emulation, our result focuses on the latency of response. In the result analysis, we use the Kolmogorov–Smirnov test for analyzation. Table 1 is the statistical table of our experiment, we let our analysis computer sending 10000 read register Modbus packet to both Schneider PLC and our emulator. In the Kolmogorov-Smirnov test, we assume that two samples come from the same distribution. Table 2 is our Kolmogorov-Smirnov test result. In the Kolmogorov-Smirnov test result, we can obtain a field that is asymptotic significance. In general, if asymptotic significance is greater than 0.05 or 0.1 then we accept the assumption of the Kolmogorov-Smirnov test, which means the two samples come from the same distribution. In table 2 we can see that our asymptotic significance is 0.997 which is greater than 0.01. As a result, we can say our emulation can respond to the same characteristics as the true Schneider PLC.

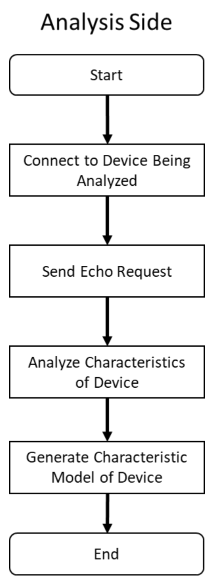


Fig. 2. Flow chart of the analysis side of specific communication system

Table 1. The statistical table of experiment device

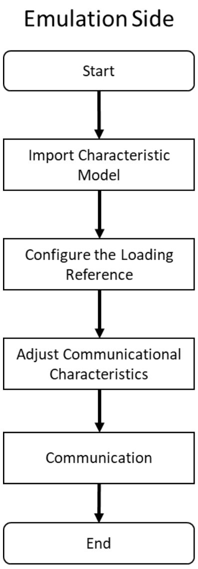


Fig. 3. Flow chart of the emulation side our specific communication system

Device	Sample Size	Average	Standard Deviation
Schneider PLC	10000	5.9258	0.378033
Emulator	10000	5.9304	0.391371

Table 2. The Kolmogorov-Smirnov test result

Most Extreme Differences	Absolute	.006
	Positive	.001
	Negative	-.006
Kolmogorov-Smirnov Z		.403
Asymp. Sig. (2tailed)		.997

5. Conclusion

In this paper, we propose a mechanism to improve the emulation of device as well as focus on the difference of behavior on devices. Thus, we analyze the response latency of device, then create a characteristics model which can be used in our specific commutation system. Our specific communication system can load the model file and provide a load parameter that can imitate the loading of device and imitate the behavior of device which is at work. Also, using the statistical method we can show the result that using our system, we can let normal computer imitate the same behavior as the working PLC.

Acknowledgements

This work was supported by the Ministry of Science and Technology (MOST) in Taiwan under contract numbers MOST 109-2218-E-006-014-, MOST 109-2218-E-006-510- and MOST 109-2221-E-041-001-.

References

1. H. Kawai, S. Ata, N. Nakamura and I. Oka, Identification of communication devices from analysis of traffic patterns, *2017 13th International Conference on Network and Service Management (CNSM),* Tokyo, Japan, Nov. 2017, pp. 1-5
2. Ola Salman, Imad H. Elhajj, Ali Chehab, Ayman Kayssi, A machine learning based framework for IoT device identification and abnormal traffic detection, *Trans Emerging Tel Tech.,* 2019
3. Antônio J. Pinheiro, Jeandro de M. Bezerra, Caio A.P. Burgardt, Divanilson R. Campelo, *Identifying IoT devices and events based on packet length from encrypted traffic,* Computer Communications, vol. 144, 2019, pp. 8-17
4. Yair Meidan, Michael Bohadana, Asaf Shabtai, Juan David Guarnizo, Martín Ochoa, Nils Ole Tippenhauer, and Yuval Elovici, ProfilIoT: a machine learning approach for IoT device identification based on network traffic analysis, *SAC '17: Proceedings of the Symposium on Applied Computing,* New York, United States, Apr. 2017
5. I. Coonjah, P. C. Catherine and K. M. S. Soyjaudah, Experimental performance comparison between TCP vs UDP tunnel using OpenVPN, *2015 International Conference on Computing, Communication and Security (ICCCS)*, Pamplemousses, Mauritius, Dec. 2015, pp. 1-5

Application of the Self-Organizing Map (SOM) to Analyze the Multiple Perspectives on Cross-National Culture

Li-Ming Chuang
The Department of International Business,
Chang Jung Christian University
No.1, Changda Rd., Gueiren District, Tainan City, 711301, Taiwan

Yu-Po Lee
The Ph.D. Program in Business and Operations Management, College of Management,
Chang Jung Christian University
No.1, Changda Rd., Gueiren District, Tainan City, 711301, Taiwan

Shu-Tsung Chao
Graduate School of Business and operation Management,
Chang Jung Christian University
No.1, Changda Rd., Gueiren District, Tainan City, 711301, Taiwan
Email: liming@mail.cjcu.edu.tw, davidlitw@yahoo.com, james@cvig.org,

Abstract

Organizational behavior in different countries and cultures has been the focus of studies in recent years. Following a literature review, we find that there are many different perspectives and features in the related cross-cultural studies. However, whether the analysis methods of these different cultural dimensions can fit into the increasingly complex and diverse topics of cross-cultural studies have not been determined, which is also a topic of great interest to scholars. Therefore, this study integrates the previous cross-cultural literature and aims to construct an analysis model of cross-national culture with multiple dimensions from three important cultural dimension theoretical models commonly used in cross-cultural studies: Hofstede, Global Leadership and Organizational Effectiveness (GLOBE) and World Values Survey (WVS). Traditional statistical analysis seems to be unable to solve the problem of the integration of relevant scales and units in different dimensions of cultural analysis. Therefore, this study uses a self-organizing map (SOM) as an analysis method to integrate 17 cultural variables from this multicultural dimension for cluster analysis and explains the cultural types in 26 countries based on the results. This study explores the differences and similarities of different countries in different cultural dimension analyses and provides a comparative model of multicultural analysis. This study takes samples from three cross-cultural analysis databases as data sources and employs the self-organizing map for analysis based on a neural network algorithm that can be used for type discrimination, map analysis, process monitoring, and error analysis. The results identify the cross-cultural groups of 26 countries and reveal their key cultural similarities and differences. We also elaborate upon the findings of these cultural characteristics and multi-cultural dimensions. The signification of this study is presented as a reference for subsequent studies of transnational and cross-cultural analysis and its applications.

Keywords: Cross-Culture, Self-Organizing Map (SOM), Hofstede GLOBE WVS

1. Introduction

At the end of the 20th century, many scholars adopted large sample empirical methods, which have become the mainstream of current research on cultural differences. Representative scholars include Hofstede (1980, 1991), Trompenaars & Hampden-Turner (1993, 1998), Schwartz (1997), and House, Hanges, & Ruiz-Quintanilla (1997). Although Hofstede's theory is the most representative, there are still some bottlenecks. For example, the samples are from the employees of a single company (IBM), the dimensions of cultural differences are insufficient, the sampling is limited, and the cultural dimensions are not dynamic and developmental. Many studies on national culture have emerged successively, such as the GLOBE (Global Leadership and Organizational

Effectiveness) project conducted by House, Hanges, & Ruiz-Quintanilla (1997). GLOBE expanded Hofstede's five dimensions into nine dimensions, retaining "power distance" and "uncertainty avoidance". Hofstede's "individualism and collectivism" were divided into "group collectivism" and "public collectivism", while " masculine and feminine culture" was divided into "gender equality and decisiveness". The "short- and long-term orientation" was changed to "future orientation". "Humanistic orientation" is consistent with Kluckhohn's dimension of "views on human nature", and the dimension of "performance-orientation" was added.

In addition to the above two analysis models of cross-national culture, the World Values Survey (WVS) has also gained increasing attention in recent years. WVS originated from the European Values Survey (EVS) conducted in 1981 for 10 countries of Western Europe. The findings are instructive in terms of cultural change and can be extended globally. Generally speaking, this transnational survey covers a wide range of topics, including social values, social norms, social issues, social distance, work issues, labor organization, employment issues, political attitudes, national democracy, gender issues, environmental issues, marriage, and family and child rearing issues. The literature of the past decade shows that cross-national culture is an important topic in the field of international enterprise research (Breuer, Ghufran, & Salzmann, 2018; Chand & Ghorbani, 2011). Relevant contextual factors such as cultural distance, cultural value, long-term orientation, individualism and physical distance can all predict different national cultures (Beugelsdijk, Maseland, Onrust, Van Hoorn, & Slangen, 2015; Malik & Zhao, 2013).

After reviewing the past studies, we find that there are many different perspectives and features in the related cross-cultural studies. However, whether the analysis methods of these different cultural dimensions can fit into the increasingly complex and diverse topics of cross-cultural studies have not been determined, which is also a topic of great interest to scholars. Therefore, this study integrates the previous cross-cultural literature and aims to construct an analysis model of cross-national culture with multiple integration dimensions from three important cultural dimension theoretical models commonly used in cross-cultural studies: Hofstede, Global Leadership and Organizational Effectiveness (GLOBE) and World Values Survey (WVS). This study focuses on the application of the self-organizing map to explore the multi-dimensional cross-cultural analysis model. A self-organizing map neural network can gather a large amount of information with similar characteristics through the self-organizing map and then compare and analyze multiple models based on the cluster data. Therefore, this study explores the differences and similarities of various countries under different cultural dimension analyses and provides a comparative model of multicultural analysis. Samples from three cross-cultural analysis databases are used as data sources.

2. Research Design

2.1. Research Method: Self-Organizing Map

A self-organizing map neural network can gather a large amount of information with similar characteristics through the self-organizing map. Since SOM is a neural network for unsupervised learning, the target output value of web-based learning does not have to be defined in advance. Cluster rules can be derived according to data similarity in order to distinguish the differences among data groups. It is an effective analysis tool for Data Mining. Self-Organizing Map (SOM) is an unsupervised artificial neural network model, proposed by Kohonen (1982). SOM is especially suitable for representing the distribution of high-dimensional data vectors in a multidimensional space. The high-dimensional data vectors can be mapped into two-dimensional space, so that a user can understand the relationship between the original data structures, and the number of data groups can be reduced.

2.2. Research Subjects and Data Sources

The data sources for this study are from three important cultural dimension theoretical documents and databases commonly used in cross-cultural studies: Hofstede (Geert Hofstede's Websites), Global Leadership and Organizational Effectiveness (GLOBE), and World Values Survey (WVS). Table 1 lists the data of the cultural dimensions of 26 countries.

Table 1 List Of Country Information

Country Information

Culture Cluster	*Country*	*Ctry. Code*
Anglo Cultures	USA	US
	Canada	CA
	England	UK
	Ireland	IE
	New Zealand	NZ
	South Africa	ZA
	Australia	AU
Latin Europe	France	FR
	Italy	IT
	Portugal	PT
	Spain	ES
	Swiss	CH
Middle East Cultures	Morocco	MA
	Turkey	TR
	China	CN
	Hong Kong	HK
Confucian Asia	Japan	JP
	Singapore	SG
	South Korea	KP
	Taiwan	TW
	Brazil	BR
Latin America	Argentina	AR
	Colombia	CO
	El Salvador	SV
	Mexico	MX
	Venezuela	VE

Date source：World Value Survey

3. Conclusion

This study used a self-organizing map (SOM) as an analysis method to integrate 17 cultural variables from this multicultural dimension for cluster analysis and explains the cultural types in 26 countries based on the results. Moreover, this study explored the differences and similarities of different countries under various cultural dimension analyses, and provided a comparative model of multicultural analysis. It sourced samples from three cross-cultural analysis databases . The self-organizing map is for analysis based on a neural network algorithm that can be employed for type discrimination, map analysis, process monitoring, and error analysis. The results identify the cross-cultural groups of 26 countries, reveal their key cultural similarities and differences, and help elaborate upon these cultural characteristics and multi-cultural dimensions. The significance of this study is presented as a reference for subsequent studies of transnational and cross-cultural analysis and its applications.

According to the results in Figures 2 to 5, Table 2 summarizes the comparisons of cross-cultural analysis patterns in multiple dimensions. Table 2 shows that there are Eastern cultural group and Western cultural group in Hofstede 6 analysis. There are three groups in the analysis of GLOBE 9: high, medium, and low GLOBE cultural groups. WVS 2 analysis shows four groups of country clustering: W1 (High T/R & LOW S/S) culture group, W2 (High T/R & High S/S) culture group, W3 (Low T/R & Low S/S) culture group, and W4 (Low T/R & High S/S) culture group. Among them, most east Asian regions or countries such as Taiwan, Japan, China, Hong Kong, and South Korea are in the W1 (High T/R & LOW S/S) culture group.

The results in Table 2 help us analyze the distribution of 26 countries after the analysis of four cross-national cultural analysis modes by SOM. It is interesting to find that there are two groups in Hofstede 6 analysis: H1 and H2; and the countries of H2 are the same as the countries of G1 and G3 after GLOBE 9 analysis; i.e., Hofstede's Western culture group is equal to the high and low cultural groups of GLOBE, and Taiwan belongs to G2 (i.e., medium GLOBE culture group) in GLOBE 9 analysis. Among the 26 regions or countries, only Taiwan belongs to this group. The cultural attribute and classification of Taiwan are worth discussing, and subsequent research should further analyze its causes.

WVS 2 cultural dimension clustering analysis presents four groups of country clustering. Most countries fall into two of these categories. One part is in the W1 (High T/R and LOW S/S) cultural group, and Taiwan belongs to this group. The other part is in the W4 (Low T/R and High S/S) cultural group, and many advanced countries belong to this group. The W2 (High T/R and High S/S) culture group has both tradition and self-expression ability, represented by two countries: New Zealand and Switzerland. Countries in the W3 (Low T/R & Low S/S) culture group are Australia, Morocco, Turkey, and Singapore. The analysis results of the above three cultural dimensions are close to the clustering results of GLOBE 9 cultural dimension analysis.

Table 2

cultural dimension / Country	Hofstede		GLOBE			WVS			Integration model		
	H1	H2	G1	G2	G3	W1	W2	W3	W4	M1	M2
USA(US)		✓	✓						✓		✓
Canada(CA)		✓	✓						✓		✓
England(UK)		✓	✓						✓		✓
Ireland(IE)		✓	✓						✓		✓
New Zealand(NZ)		✓	✓				✓				✓
South Africa(ZA)		✓	✓						✓		✓
Australia(AU)		✓	✓					✓			✓
France(FR)	✓		✓			✓				✓	
Italy(IT)	✓				✓	✓					✓
Portugal(PT)	✓				✓				✓	✓	
Spain(ES)	✓				✓	✓				✓	
Swiss(CH)		✓	✓				✓				✓
Morocco(MA)	✓				✓			✓		✓	
Turkey(TR)	✓				✓			✓		✓	
China(CN)	✓		✓			✓					✓
Hong Kong(HK)	✓		✓			✓					✓
Japan(JP)	✓		✓			✓				✓	
Singapore(SG)	✓		✓					✓			✓
South Korea(KP)	✓				✓	✓				✓	
Taiwan(TW)	✓			✓		✓				✓	
Brazil(BR)	✓		✓						✓	✓	
Argentina(AR)		✓			✓				✓	✓	
Colombia(CO)		✓			✓				✓	✓	
El Salvador(SV)		✓			✓				✓	✓	
Mexico(MX)		✓			✓				✓	✓	
Venezuela(VE)		✓			✓				✓	✓	

References

1. House, R. J., Hanges, P., Ruiz-Quintanilla, S. A., & Dickson, M. W. The development and validation of scales to measure societal and organizational culture. Under review, 1997.
2. Inglehart, R., & Welzel, C. Modernization, cultural change, and democracy: The human development sequence. Cambridge Univ Pr, 2005.
3. Trompenaars A., Hampden-Turner C.. Riding the waves of culture: Understanding cultural diversity in business[M]. London: Nicholas Brealey, 1997.
4. WORLD VALUES SURVEY 1981-2008 OFFICIAL AGGREGATE v.20090901, World Values Survey Association (www.worldvaluessurvey.org). Aggregate File Producer: ASEP/JDS, Madrid., 2009.

Threats Hidden in Employee Workstation through Office Files

Tung-Lin Lee
Department of Electrical Engineering / Institute of Computer and Communication Engineering,
National Cheng Kung University
No.1, University Rd., East Dist., Tainan City 70101, Taiwan

I-Hsien Liu
Department of Electrical Engineering / Institute of Computer and Communication Engineering,
National Cheng Kung University
No.1, University Rd., East Dist., Tainan City 70101, 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 70101, Taiwan

*E-mail: tllee@cans.ee.ncku.edu.tw, ihliu@cans.ee.ncku.edu.tw, jsli@mail.ncku.edu.tw**

Abstract

With the advent of the Internet of Things era, IoT devices participate in our work environment. From printers, shared files to temperature control, they make our working environment more convenient. Although these IoT devices communicate through different communication protocols, the same identity verification method is often used., hackers can also use this authentication method to steal user identity verification information, This study combines past research results to present our new findings, and further, organize them into a complete attack architecture.

Keywords: Active Directory, IoT, Credentials, Spear-fishing

1. Introduction

In an environment with Active Directory (AD), all resource access authorization authentication will be handled by Active Directory, so that every Server (EX: Exchange server, NAS, printer) does not need to handle every user's authentication credential. Although the server authentication work is shared by Active dictionary, for users, to access a certain function, they still need to enter the password repeatedly once they want to access different resource. To deal with this situation single sign-on implementation was invented.

Due to the single sign-on implementation of windows operating systems, users could save a lot of time without having to enter their account and password again and again. However, this setting also allows hackers to Use some dedicated constructed malicious files to steal the user's identity verification information (NTLM hash), and then launch attacks like passing the hash, and then log in to the user's computer remotely, which becomes a way to gain access to the organization's intranet.

Although there are many ways to cause an NTLM hash leak, fortunately, most of them cannot be exploited by hackers. In this study, we would discuss several NTLM hash leak methods, and focus on some of the most commonly used method of hacking: Office files, and dive deeper to the possibility of extended application and under what circumstances will the existing protection mechanism be bypassed.

In the second section introduced background knowledge related to authentication, as well as traditional attack methods. In the third section, we will introduce our new discoveries and ways to bypass existing protections and combines past research and our findings to try to construct a method that can be used to automate large-scale attacks in real scenarios. Finally, the forth section will be our conclusion.

2. Background

In this section, we will introduce a common authentication method "NTLM". Besides, we will discuss the protocol and application using NTLM, and the past attacks on this authentication.

2.1. *NTLM Overview*

NTLM is a challenge-Response authentication mechanism. First, the client informs the server, the username and informs the server that it wants to log in. Then the server sends out a set of random number as a challenge and requires the client to hash it with its own password. Third, the client sent the hash result to the server, and then Server will verify whether the password is correct or not.

The NTLM protocol suite was first implemented in a Security Support Provider of Microsoft security protocols and been adopted in other protocols. Because NTLM has undergone many changes, We are not going to talk about its cryptography details here. If you are interested in the detailed, Microsoft official document could provide more information.

NTLM is mainly used for SMB, LDAP, MSSQL, HTTP, and other protocols, not only that, although it is not the preferred authentication method, in fact, NTLM is also used for wi-fi connection and remote desktop authentication. Although NTLM has been around for decades and has been updated many times, its security mechanism has often been challenged. MITRE ATT&CK has suggested that enterprises or institutions should not use NTLM for authentication but instead using Kerberos. However, Kerberos is more complicated to set up and there is downward compatibility and the third-party software requirement which does not support Kerberos as an authentication option, using NTLM for authentication is usually a necessary option.

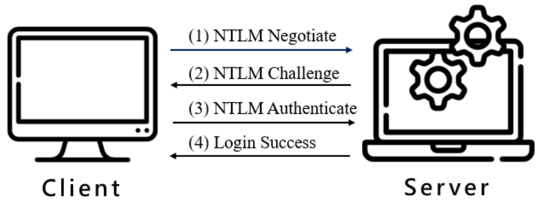


Fig. 1. The NTLM Challenge-Response Mechanism.

2.2. *Typical attack*

The security problems of the NTLM mechanism are reflected in two aspects. First, the NTLM hash can be cracked by rainbow tables or brute force. Common attack methods used Responder as a malicious server which answers to specific queries to receive the credentials sent by the victim to obtain the NTLM hash and then start the cracking process.

However, it is difficult to crack in a limited time, so there is another attack method called NTLM relay or so call pass the hash. The attack method is another variant of a man-in-the-middle attack as shown in Figure 2. Although there are some mitigations such as EPA (Enhanced Protection for Authentication), there are still ways to bypass those mechanisms such as “drop the mic”, and NTLM relay is still an indispensable step while implementing these attack. Through NTLM relay, an attacker can steal the victim’s credentials, and thus obtain confidential data that the victim could access or call the RPC service that is authorized to control the server, thereby affecting the entire operation of the organization or enterprise.

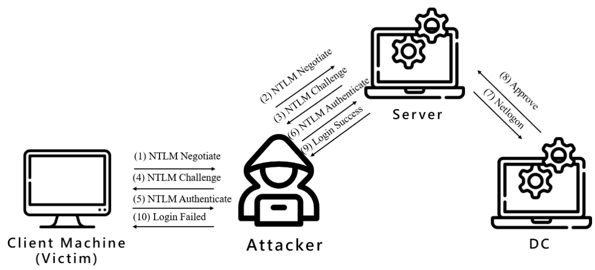


Fig. 2. The NTLM Relay Attack with Active Dictionary

2.3. *NTLM leak*

For the NTLM relay attack, the key relies on how to make the victims actively send a login request to the malicious server. through malicious file is one of the way.

In the article "Living on land: NETNTLM HASHES " introduced those penetrations such as HTML files, PDF, Window media player, Office files, and others to cause NTLM leakage.

The principle is that when we click on these malicious files, the files actively request resources embedded in the UNC path through the SMB protocol. This sounds reasonable, but when the server located on this path asks for a certificate, the client will actively send the NTLM hash, which is the credential of the victim! Because of Windows integrated authentication, Windows integrated authentication is mainly suitable for users not to repeatedly enter account and password when accessing any resources. This is a great function to save users' time and reduce the possibility of Side-channel attack attacks, but when files need to access resources on a certain UNC path, the Windows operating system. It will consider that the security of these UNC paths has been verified when the file is made, so when the manufactured file is opened again, the file would automatically send the resource query and user credentials if needed.

Usually, those malicious files are sent through attachments in phishing emails. Fortunately, most people, rarely click on the extensions called .m3u .url .jnlp and other rare seen extensions. The most vulnerable types are PDF and office documents because these two file types are also the most common types in daily office work.

Table 1. File Type that Could Cause NTLM Leak.

File Type	File extension
Internet shortcut	url
Windows media player	m3u, asx, wax
Java external jar	jnlp
Microsoft Word	docx, xml
Chrome & IE & Edge	htm
Adobe Acrobat Reader	pdf

3. Our findings

3.1. *Office mitigation*

The NTFS file format supports using the named Zone: Identifier in the alternate data stream to mark files from the Internet. In order to avoid attacks similar to NTLM relay and other malicious code being executed while the file is opened, browsers and other internet clients such as email and chat programs utilize IAttachmentExecute interface's methods or write the Alternate Data Stream directly to mark files from the Internet. If this file happens to be an Office document while opening this Office file, we will see it open in Protected View, Similar to a sandbox to prevent malicious cities from being executed, noticed that the Enable all feature in Protected View is because the files is from the internet. It is different from the Enable Content seen by Macro in the file. In our study refers to the former, which is "Enable all feature".

3.2. *Bypassing Office mitigation*

Although there used to be browsers or Mail Desktop failed to check specific file types, and different determination on dragging downloadable files or double clicking to downloadable files. In other words, if the user dragged a file from Mail Desktop to a folder,the protect mode would be bypassed. However, this type of bypass technique has been patched by most software developers, therefore hackers started to focus on other problems: decompression. When compressed files are decompressing, the unzipped file should inherit the characteristics of Zone:identifer. Although most compression software has noticed this, different compression methods in the same decompress software will have different behaviors. Taking the most common 7zip as an example, it will propagate only when the compressed file is open in archive and double click the file. the article "Downloads and the Mark-of-the-Web" discuss it before, we re-experimented and organized, and found a little different result (Table 2)

Table 2. Popular Extractor List

Extractor	Double-click in Archive	Extract all
Windows Explorer	Not vulnerable	Not vulnerable
WinRar	Not vulnerable	Not vulnerable
7zip	vulnerable	Not vulnerable
WinZip	Not vulnerable	vulnerable
IZArc	Not vulnerable	Not vulnerable

3.3. *Is stealing NTLM hash over other protocol possible?*

The malicious files in table I steal the NTLM hash only through the SMB protocol. However, in the real world, experienced network administrators will prohibit SMB from connecting to the internet. Not only there is usually no related requirement, but to avoid NTLM hash through SMB. While hackers install Responder on the infected employee's computer is not possible to open rogue SMB servers, because that TCP port 445 has been bound by the Network Neighborhood or samba. The responder cannot bind the same port with an existed application. Because of those reasons, we try to insert the resource access path of the HTTP URL path in the Office document. Unfortunately, HTTP Request will be sent out successfully, but when the server responds with an access denied and requested, it uses NTLM for authentication
The client-side usually does not respond in the default settings, but when the client-side computer is set to "Automatic logon with current username and password" In an intranet or trusted site, HTTP will take the initiative to issue an NTLM hash and complete the entire authentication process.
This means that once the hacker hijack a host, and the internal network environment is set to Automatic logon with the current username and password in the intranet, Hackers can steal the credential of the entire intranet by sending spearing fishing e-mail attachments with an HTTP URL path pointing to the infected server through the investigation of the group personnel in the intranet.(Figure 3)

4. Conclusion

IoT devices not only appear in our lives but also enter our work environment. They all make the working environment more convenient and improve the work efficiency of employees. These IoT devices often support NTLM as an identity verification method. Furthermore, to prevent employees from constantly entering account passwords, the device automatically sends the user's identity verification information. Although this saves the user's time to enter the account and password, hackers can also use this authentication method to steal user identity verification information, and use the identity verification information to log in to other devices, This article combines past research results to present our new findings, and further, organize them into a complete attack architecture.

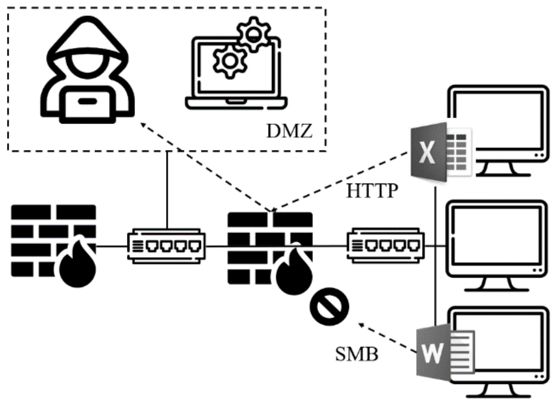


Fig. 3. The Attack Simulation Scenario

Acknowledgements

This work was supported by the Ministry of Science and Technology (MOST) in Taiwan under contract numbers MOST 109-2218-E-006-014- and MOST 109-2218-E-006-510-.

References

1. Jianing Wang and Junyu Zhou, NTLM Relay is dead, Long Live the NTLM Relay, HITBSecConf, 2018.
2. William Mattin, *One-Click to OWA*, DEF CON 26, 2018.
3. Marina Simakov and Yaron Zinar, Relaying Credentials Has Never Been Easier: How To Easily Bypass The Latest NTLM Mitigation, DEF CON 27, 2019.
4. Downloads and the Mark-of-the-Web, 2016
5. I. Ghafir and V. Prenosil and M. Hammoudeh, Dsguised executable files in spear-phishing emails: detecting the point of entry in advanced persistent threat, ICFNDS '18, 2018
6. Mohammed M. Alani, IoT Lotto: Utilizing IoT Devices in Brute-Force Attacks, ICIT 2018: Proceedings of the 6th International Conference on Information Technology: IoT and Smart, pp 140–144, 2018

Microsatellite attitude control approach: combined with Generation Adversarial Networks fault-detection and Cerebellar Model Articulation Controller fault-tolerant control

Ho-Nien Shou
Dep. of Aviation & Communication Electronics, Air Force Institute of Technology
Gangshan, Kaohsiung, Taiwan (R.O.C.)

E-mail: honien.shou@gmail.com

Abstract

In this paper, a new attitude control architecture for microsatellite is proposed. Based on the deep learning fault-detection method, the Cerebellar Model Articulation Controller (CMAC) is used as the fault-tolerant control. With the function of Generation Adversarial Networks (GAN) to recognize images, the microsatellite attitude fault wavelet spectrum is used as a guide for the training of generators and discriminators, and for system real-time fault diagnosis and classification. When the system fault diagnosis determines that the fault occurs, the cerebellar neural network is involved in fault-tolerant control. The GAN learning ability of the generative confrontation network is used to solve the problem of insufficient sample data and sample labeling respectively. The CMAC is used as a local learning network, which has generalization ability, strong convergence speed, and easy software and hardware implementation. The simulation results show that compared with other methods, the GAN method of fault-detection combined with CMAC can achieve higher accuracy and robustness.

Keywords: deep learning, fault-detection, fault-tolerant control, Cerebellar Model Articulation Controller, Generation Adversarial Networks.

1. Introduction

The Generative Adversarial Network (GAN) was proposed in 2014 [1]. The GAN continuously optimizes the network to achieve better results by using the mutual confrontation between generator and discriminator. There are many hidden deep learning network models in GAN, and it has good feature learning ability. At first, it is mainly used in data generation and image generation to solve the problem of insufficient sample data. Based on different research purposes, derivative models for GANs are constantly being proposed. A Semi-Supervised Learning with Context-Conditional Generative Adversarial Networks (SSL) is proposed, in which the generator structure remains unchanged and the discriminator is classified in two categories an output layer is added to achieve the multi classification function [2]. Since the SSL was proposed, it has been applied to speech generation and recognition, image recognition, sample classification and other fields [3]. The SSL is applied to human action recognition, and the images are captured by the video frame as the network input for action recognition [4]. In the above application, the SSL achieves the recognition effect.

The Cerebellar Model Articulation Controller (CMAC) was developed by Dr. J.S. Albus in 1975, based on the biological model of cerebellar cortex proposed by Marr in Medical Research Institute. CMAC are widely used because of characters of fast convergence speed, strong generation ability, it can overcome the local optimization problem of back propagation neural network (BP), simple structure, easy implementation by software and hardware. CMAC has been successfully applied in many application fields, such as fault detection and nonlinear control. In the classification task, Lin et al. [5], a parameter fuzzy CMAC neural network with mixed parameters is proposed. Learning algorithm composed of self-clustering genetic algorithm (GA) and improved GA for facial detection and breast cancer diagnosis.

In recent years, the research on motor bearing fault diagnosis has been widely concerned by scholars and engineers. The vibration analysis method is commonly used for bearing fault diagnosis, which analyzes and classifies the collected bearing vibration signals. However, bearing vibration signal has the characteristics of non-stationary, so it is very difficult to extract the characteristics of the bearing.

As a time-frequency analysis method, continuous wavelet transform can effectively extract the time-frequency characteristics of non-stationary signals. The function of wavelet transform is to take one-dimensional signal as input. The result of time-frequency analysis essentially reflects the two-dimensional spectrum of energy intensity of signal at different time and frequency, and the output is displayed in the form of time-frequency diagrams. At present, continuous wavelet transform has been widely used in the field of mechanical fault diagnosis [6]. In reference, a combination of wavelet transform branch support vector machine (SVM) and convolutional neural networks (CNN) was proposed. In reference [6], wavelet transform and back propagation neural network (BP) were used to realize fault diagnosis of different motors. In the application of the above literature, the feature vector is constructed artificially based on the time-frequency map, and the learning and expression ability of the selected neural network is limited [6].

The content structure of this paper is a brief introduction to Section 1 and a description of SSL in Section 2 in Section 3, the CMAC is explained, in Section 4 wavelet transform of fault model, in Section 5 attitude fault diagnosis and fault-tolerant control simulation of microsatellite in Section 6. Finally, the conclusion is drawn.

2. Semi-Supervised Learning with Context-Conditional Generative Adversarial Networks (SSL)

The SSL routing generator and discriminator are composed of the generator and discriminator. The generator belongs to the generative model. The difference between the generator and other models is that the generator does not need to be modeled in advance. However, due to the lack of constraints, the generated data is too free. This paper proposes a kind of improvement to solve the above problems Based on the semi supervised generative countermeasure network algorithm, conditional variable is added to the model to guide the training process of generator and discriminator by using additional information. Compared with the prototype generated countermeasure network, the main difference is that the discriminator outputs a class signal (the generated sample is No For the discriminator, the loss function consists of two parts, one is supervised learning loss (only need to judge whether the sample is true or false), and the other is unsupervised learning loss (judging sample category). The generator only needs to generate realistic samples as far as possible. After training, the discriminator can be used as a classification model to classify.

The generator of the improved SSL consists of several deconvolution layers, excitation layers and extracted over fitting Fitting) and speeding up the training speed method, for the input n-dimensional noise signal connecting condition variables, the final output is almost the same as the real signal "false" signal, which can achieve the state of "false" with the real signal, which is conducive to expand the sample to achieve better training effect. The discriminator consists of several convolution layers, excitation layers, fully connected layers and classified output layers, as well as the

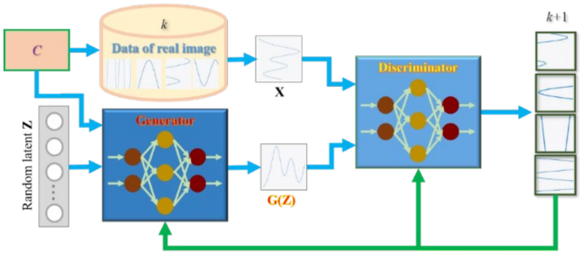


Fig. 1. SSL algorithm process and structure

corresponding methods to prevent over fitting and speed up the training. The discriminator takes the real signal, that is, the signal to be detected and the "false" signal generated by the generator, as the input at the same time.

The flow chart of the improved generative countermeasure network algorithm is shown in Fig. 1.

3. Cerebellar Model Articulation Controller (CMAC)

CMAC is a kind of artificial learning network which imitates the hierarchical storage of information in human cerebellar cortex, and its operation process is basically achieved by a series of mapping methods. As shown in Figure 2, the basic structure of MIMO CMAC with three inputs and three outputs is shown Fig. 2.

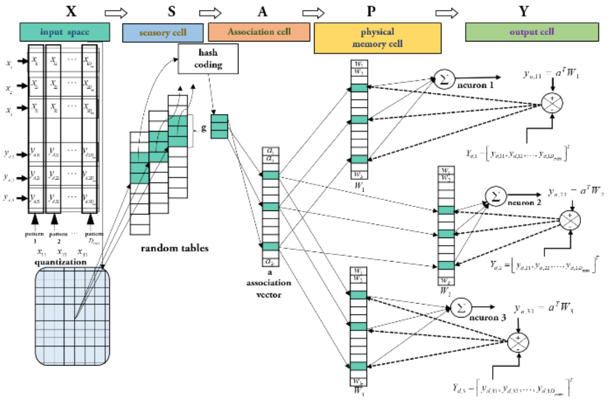


Fig. 2. The MIMO CMAC block diagram

The five units of CMAC neural network consist of (i) input space (**X**), (ii) sensory cell (s), (iii) Association unit (**A**), (iv) physical storage unit (**P**) and (v) output storage cell (**Y**) using a series of images. Fig. 2 shows the network topology vector of three inputs and three output vectors of CMAC neural network; the input is represented by $[x_1, x_2, x_3]^T$, and the actual output is $[y_{d1}, y_{d2}, y_{d3}]^T$. The figure shows an icon for the input mode 1 image, including three input elements $[x_1, x_2, x_3]^T$ and one desired output y_{d1}.

4. Fault Model Wavelet Transform

The attitude fault diagnosis method of microsatellite is based on the fault signal generated by deep convolution and combined with continuous wavelet transform. The ability of wavelet transform for nonstationary signal and the function of SSL processing and image recognition are introduced. Based on the SSL, the model is introduced and the loss function is optimized and optimized. In order to guide the training of generator and discriminator, the generator and semi supervised learning ability are used to solve the problem of sample data shortage and sample marking respectively.

Case1.

$$f_{a_1} = \begin{cases} 0.5, & 40 \le t < 70(\text{sec}) \\ 0 \end{cases}$$
$$f_{a_2} = f_{a_3} = 0 \tag{1}$$

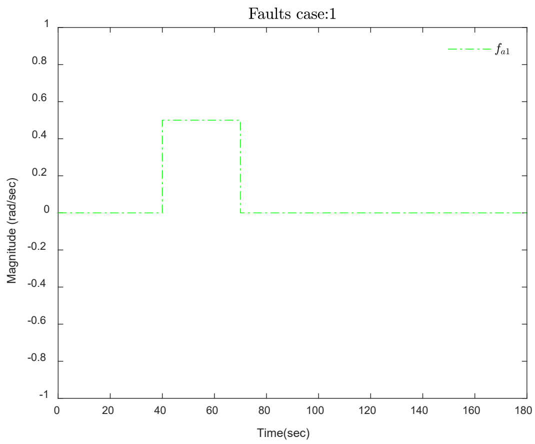


Fig. 3 Fault case 1

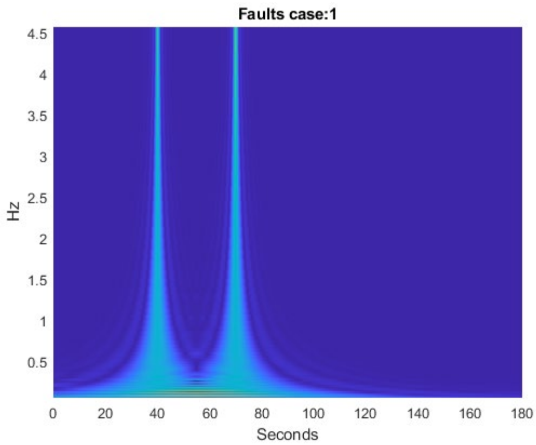


Fig. 4 wavelet transform coefficient diagram of case 1 attitude fault

The fault signal is transformed into frequency time wavelet coefficient graph (time-frequency diagram) to achieve time-frequency analysis of vibration signal and obtain more comprehensive and detailed information.

The feature extraction and image recognition ability of deep learning GAN are used, and the compressed time-frequency diagram is used as input. The wavelet transform coefficients of case 1is shown in Fig. 4.

5. Fault Estimated and Fault Tolerant Control Simulation of Microsatellite Attitude

The microsatellite dynamic equation is expressed as a nonlinear equation

$$J\boldsymbol{\omega} = -\boldsymbol{\omega}^{\times} J\boldsymbol{\omega} + \boldsymbol{\tau}_w + \boldsymbol{\tau}_u \tag{2}$$

where $J = \operatorname{diag}\{J_x, J_y, J_z\}$ are inertia moments of the satellite along principal axes; $\boldsymbol{\omega}$ are the angular velocity of the body-fixed reference frame. $\boldsymbol{\tau}_w$ are space disturbance torques and $\boldsymbol{\tau}_u$ are the control torques along principal axes.

For small attitude angles, the derivation of eq. (2) is as follows

$$\begin{cases} \dot{x}(t) = Ax(t) + g(t,x) + B_u u(t) + B_w w(t) + F_a f_a(t) \\ y(t) = Cx(t) + F_s f_s \end{cases} \tag{3}$$

Where state vector $x = [\psi \quad \theta \quad \phi \quad \dot{\psi} \quad \dot{\theta} \quad \dot{\phi}]^T$, nonlinear term $g(t,x) = \begin{bmatrix} 0 \\ \omega_0 \\ 0 \\ (J_z - J_y)\omega_y\omega_z / J_x \\ (J_x - J_z)\omega_x\omega_z / J_y \\ (J_y - J_x)\omega_x\omega_y / J_z \end{bmatrix}$,

actuator fault matrix $F_a = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T$ and actuator fault vector $f_a = \begin{bmatrix} f_{a_1} & f_{a_2} & f_{a_3} \end{bmatrix}^T$.

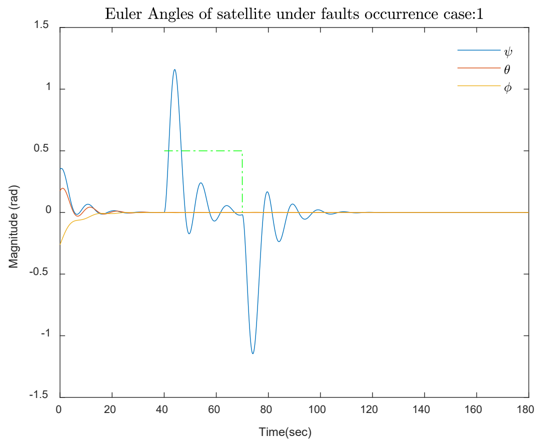


Fig. 5 Time response of case 1 attitude angle failure and recovery

6. Conclusion

According to the numerical simulation results in the previous section, the feasibility of applying SSL to microsatellite fault estimated. Through SSL of GAN samples and training parameters, the CMAC has good generalization ability and stability under the brake fault model.

References

1. Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville and Yoshua Bengio, "Generative adversarial nets," In Advances in Neural Information Processing Systems(NIPS), 2014, pp.2672-2680.
2. Salimans T., Goodfellow I., and Zaremba W., "Improved techniques for training GANs," In NIPS, 2016, pp. 2234–2242.
3. Salimans T, Goodfellow I, Zaremba W, et al., " Improved techniques for training GANs," Advances in Neural Information Processing Systems, 2016, pp. 2234-2242.
4. Lantao Yu, Weinan Zhang, Jun Wang, Yong Yu, "SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient," 31th AAAI Conference on Artificial Intelligence, , 2017, pp.2852-2858.
5. C.J. Lin, J.H. Lee, C.Y. Lee, A novel hybrid learning algorithm for parametric fuzzy CMAC networks and its classification applications, Expert Systems with Applications 35, 2008, pp.1711–1720.
6. Yuan Jian-hu, Han Tao, Tang Jian, et al. "An approach to intelligent fault diagnosis of rolling bearing using wavelet time-frequency representations and CNN," Machine Design and Research **33**(2), 2017, pp.93-97.

Contribution to the Theory of Periodic Reaction of Three Bodies Systems

Yasuhiro Suzuki
Graduate School of Informatics, Nagoya University, Furocho Chikusa
*Nagoya City, Aichi Prefecture 464/0814, Japan**

E-mail: ysuzuki@nagoya-u.jp.com
www.ysuzuki.info

Abstract

The chemical ecosystem is an ecosystem in which chemical substances mediate interactions. When herbivores feed, the plant analyzes herbivore saliva and produces volatile chemicals that attract herbivore natural enemies. Natural enemies are attracted to volatile chemicals and eliminate herbivores. This system is a tripartite system consisting of herbivore, carnivore and volatile chemicals. The basic equation of mathematical ecology is the Lotoka-Volterra equation. This equation is a two-way system of herbivores and carnivores. This paper proposes the basic equations of three chemical ecosystems and shows their mathematical features.

Keywords: Chemical Ecology, Lotoka-Volterra equation, Tritrophic system, mathematicalbiology

1. Introduction

Mathematical ecology has used the Lotka-Volterra equation as the basic equation[1]. This equation is a simultaneous differential equation with the predator as y and the prey as x.

$$\frac{dx}{dt} = k_1 x - k_2 xy, \qquad (1)$$

$$\frac{dx}{dt} = k_1 x - k_2 xy, \qquad (2)$$

The equilibrium points of this simultaneous differential equation are obtained from [3] and [4]

$$0 = \frac{dx}{dt} = x(k_1 - k_2 y), \qquad (3)$$

$$0 = \frac{dy}{dt} = y(k_2 x - k_3), \qquad (4)$$

as

$$(x, y) = \{(0,0), \left(\frac{k_3}{k_2}, \frac{k_1}{k_2}\right)\} .$$

Therefore, if y is larger than k_1 / k_2, x decreases, and if y is smaller, it increases. On the other hand, if x is larger than k_3 / k_2, y increases, and if it is smaller, y decreases. Therefore, x and y oscillate around the equilibrium point. The solution to this equation is periodic[1] (Fig. 1). The Lotoka Volterra equation is a basic equation of mathematical ecology, and various variants have been proposed (such as [2]).

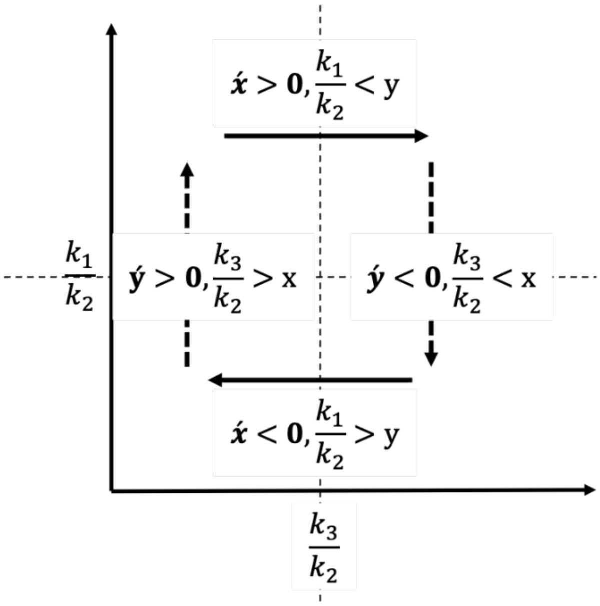


Fig. 1. The isocline analysis of LV equation

1.1. *Chemical Ecology*

Plants are the foundation and resources that support the entire ecosystem, but on the other hand, indirect defense using pests and communication between plants have been revealed. [3]

Since the late 1980s, it has become clear that plants attract predators' natural enemies by releasing volatile substances in response to feeding damage [3]. In this case, there is an indirect interrelationship through odor information between plants and natural enemies that are not in an “eat-eat” relationship that goes beyond the nutritional stage. Plant-predators-predators span three nutritional stages and are linked not only in the food chain but also in exchanging information.

Volatility produced by plants is called Herbivore Induced Plant Volatiles (HIPV). When pests are directly exterminated using pesticides, the side effect is to exterminate the pests to be exterminated and the natural enemies existing in the space. In that case, if pests invade the place where the natural enemy was exterminated again, the natural enemy cannot control it. As a result, plants suffer great feeding damage. To avoid this, it is necessary to spray pesticides over a wider area, and the ecosystem will be seriously damaged.

On the other hand, if a plant that has been damaged by feeding produces HIPV and provides indirect defense by natural enemies, the ecosystem will not be seriously damaged. Plants have their natural enemies exterminate pests in return for the cost of producing scents. Natural enemies can increase the efficiency of finding pests in return for the cost of learning scents. However, pests are natural enemies if they eat plants. It is unilaterally disadvantageous because it is eaten by.

In reality, pests are not unilaterally disadvantaged. Plants do not produce HIPV immediately after being damaged by feeding. Plants do not produce HIPV until they are about to die. Therefore, there is time to spare from the start of feeding to the invasion of natural enemies. Adult females spawn until the invasion of natural enemies and move to other safe strains before invading natural enemies. Therefore, pests get a space where natural enemies do not exist in return for the cost of attracting natural enemies and it can lay eggs safely.

1.2. *Chemical Ecology, rule based expression*

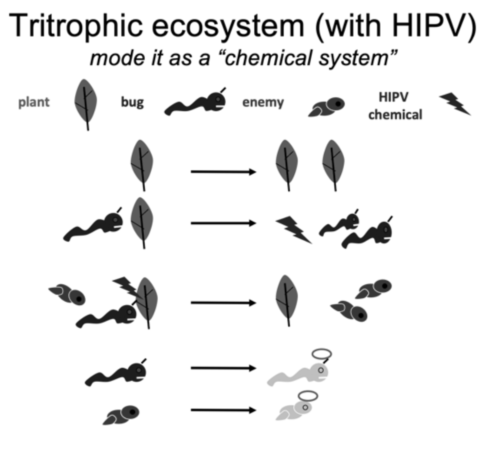


Fig. 2. Schematic rule based model of the tritrophic system composed of Plant-Herbivores and Carnivore

Describe the chemical ecosystem on a rule based (Fig.2). The rule base is an abstract chemical reaction system. An abstract chemical reaction system is a method of describing a phenomenon using the notation of a chemical formula. Reaction kinetics transforms chemical equations into differential equations.

Chemical species used in chemical ecosystems are plants, prey, predators and HIPV.

Plant + Predator-> Prey, Predator, HIPV

This reaction expresses that the prey (herbivore) eats the plant to increase the population and the plant produces HIPV.

HIPV + prey (carnivorous) + predator-> predator, predator

This reaction expresses that HIPV-induced predators eat prey and increase populations. Plants cease to produce HIPV as there are no more herbivores. Therefore, there is no HIPV on the right side of this equation.

Predator->

This reaction represents the natural death of a predator. The reason for not including the natural death of prey is that the number of prey is decreasing too much. Counting the increase and decrease of predators with the above three chemical formulas, it is 0. The increase / decrease in predators is 0. Therefore, the increase / decrease in prey and predator is balanced. If the natural death of the prey is added, the prey will decrease and become extinct even if it is not predated. Its extinction is not essential, but the balance of stoichiometry. On the other hand, without the natural death of predators, the number of predators increases exponentially. This increase is also a problem of stoichiometric balance and is not essential. Therefore, there was no natural death of the predator, and only the predator died of natural causes.

2. Basic equation of chemical ecosystem

Chemical ecosystems are not just related to prey and predators. An information chemical called HIPV intervenes between prey and predators. The indirect interaction between prey and predator by HIPV is the essence of the chemical ecosystem. Therefore, the chemical ecosystem is an interaction of three bodies.

The Lotka-Volterra equation cannot describe the chemical ecosystem in essence. Lotka-Volterra equations with multi-bodies have been proposed [4]. The n species Lotka-Volterra equation has two or more predators or prey but they are essentially two-body interaction, not an n (n> 2) bodies interactions. Therefore, the generalized Lotka-Volterra equation cannot essentially describe the chemical ecosystem.

Let x be the number of prey, y be the predator, and i be the information (HIPV). x, y, i are all positive numbers greater than or equal to 0.

Prey increases by eating plants and decreases by predation by predators. Unlike the LV equation, when a prey eats a plant, the plant produces HIPV, which is represented by i. Predators cannot find any prey if the density of HIPV is low. In addition, the number of prey is reduced by being preyed on by predators. HIPV plays the role of stabilizer, predators cannot have prey as much as possible. Because, if predators get many preys, the concentration of HIPV becomes low and predators cannot find preys. The basic equation of the chemical ecosystem based on the LV equation can be obtained from the above.

$$\frac{dx}{dt} = k_1 x - k_2 xyi, \tag{5}$$

$$\frac{dy}{dt} = k_2 xyi - k_3 y, \tag{6}$$

$$\frac{di}{dt} = k_1 x - k_2 xyi. \tag{7}$$

What makes this equation essentially different from the LV equation is that it includes a term for three-body interactions.

3. Behavior of basic equations of chemical ecosystems

The equilibrium points of this equation are

$$0 = \frac{dx}{dt} = x(k_1 - k_2 yi), \tag{8}$$

$$0 = \frac{dy}{dt} = y(k_2 x i - k_3), \qquad (9)$$

$$0 = \frac{di}{dt} = x(k_1 - y k_2), \qquad (10)$$

and $(x, y, i) = \{(0, 0, -), \left(\frac{k_3}{ik_2}, \frac{k_1}{ik_2}, \frac{k_1}{k_2}\right)$.
The LV equation has two equilibrium points, but in this equation the equilibrium points are all points on c / i and are infinite. And when i = 1, the equilibrium point coincides with the LV equation. When i = 0, x and y have no equilibrium point. When i = 0, dx / dt> 0 and dy / dt <0. If k1, k2, and k3 are positive constants, k3 / k2 is c3, and k1 / k2 is c1, the equilibrium point is (c3 / i, c2 / i, c2).

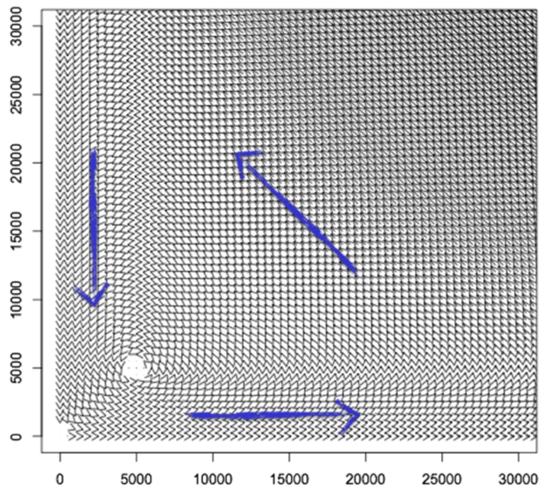


Fig. 3. Vector space of LV equation

3.1. *Vector space of equations*

We compared the behaviour of the trytrophic equations with the LV equation. So we investigated both vector fields. The LV equation's vector field is a periodic solution centred on the origin and another equilibrium point, and this result has been known for a long time (Fig. 3).

The diagonal is the set of equilibrium points. In reality, the equilibrium points on this diagonal are k / i of the equilibrium point curve. Below this diagonal, the concentration of HIPV is low, so Y decreases and X increases. Above this diagonal, the concentration of HIPV is high, so Y increases and X decreases. Since the

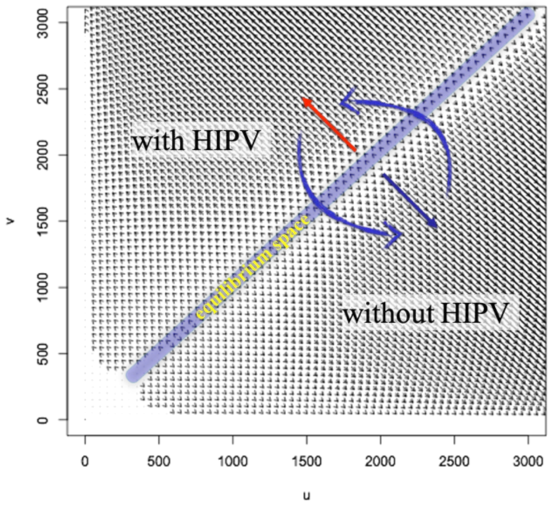


Fig. 4. Vector space of basic equation of tritrophic system., The diagonal is the set of equilibrium points. In reality, the equilibrium points on this diagonal are k / i of the equilibrium point curve.

sum of these two vectors is exactly the opposite, the vector field shows that the tritrophic basic equation has a stable manifold structure. (Fig.4).

From this result, it is foreseen that the basic equation of tritrophic is significantly different from the LV equation and has an equilibrium solution. In the future, I would like to proceed with a more rigorous mathematical analysis.

References

1. Alfred J.Lotoka, Elements of Physical Biology, Williams and Wilkins Company, 1925.
2. Dimas Martins Afonso, Gjini Erida, Modeling Competitive Mixtures With the Lotka-Volterra Framework for More Complex Fitness Assessment Between Strains, Frontiers in Microbiology, 11, 2132, 2020.
3. Takabayashi J., Dicke M. (1997) Herbivore-Induced Plant Volatiles with Multifunctional Effects in Ecosystems: A Complex Pattern of Biotic Interactions. In: Abe T., Levin S.A., Higashi M. (eds) Biodiversity. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-1906-4_9
4. Kerner, Edward H, Comment on Hamiltonian structures for the n-dimensional Lotka–Volterra equations, Journal of Mathematical Physics 38, 1218 (1997); https://doi.org/10.1063/1.531875l, Published Online: 04 June 1998

Deep Micro Vibrotactile, DMV and its Applications

Yasuhiro Suzuki
Graduate School of Informatics, Nagoya University, Furocho Chikusa
*Nagoya City, Aichi Prefecture 464/0814, Japan**

E-mail: ysuzuki@nagoya-u.jp.com
www.ysuzuki.info

Abstract

Deep Micro Vibrotactile, DMV is infrasound. DMVs produce sound effects when mixed with audio. Since the DMV has a transcendental low frequency, mixing it with sound such as music does not affect the frequency spectrum in the audible range. However, through demonstration experiments at concerts, it was confirmed that DMV changes the tactile qualities of the sound. This paper introduces the acoustic features and case studies of DMVs

Keywords: Deep Micro Vibrotactile, Infrasonic, very low frequency sound, sound pressure analysis

1. Introduction

Infrasound is a sound generated by vibrations of 20 times or less per second. In general, the sound loses energy due to vibration as it passes through the medium. Therefore, high frequencies quickly lose energy and are attenuated, while low frequencies are less attenuated. Therefore, it can travel up to thousands of kilometres. Low frequencies are less directional than high frequencies. Proceed in concentric circles from the source. Even if there are obstacles, go around [1)].

In the natural world, low-frequency sounds are generated from wind, air vibrations, self, volcanic activity, avalanches, waterfalls, signals from animals such as elephants, etc. Artificial objects arise from outdoor units of air conditioners, wind turbines for wind power generation, turbines of thermal power plants, machine tools, blasts, pipe organs, etc [1)].

Infrasound is regarded as noise, and its effects on the human body and shielding methods have been investigated [1)]. Kuhn, et al. confirmed that infrasound near the hearing threshold might induce changes in neural activity.

Some of these brain regions are known to be involved in auditory processing, while others are regarded as key players in emotional and autonomic control [2)].

Low-frequency sounds have been used primarily in medicine. One of the most widely used techniques is Whole-body vibration (WBV). WBV is a therapeutic method that exposes the entire body to mechanical oscillations while the patient stands or sits on a vibrating platform. WBV appears to affect balance, muscle function positively, and physical performance, which would benefit the skeletal muscle's treatment, decline seen in sarcopenia [3)].

In music compose, the lowest note on a pipe organ is 16Hz. The d-Moll Toccata and Fugue BWV 565, composed by Johann Sebastian Bach (1685-1750), contains a 16Hz bass.

Fig. 1. St. Thomas Church, Leipzig and the Pipe-organ. J.S. Bach worked as a music director from 1723 until his death in 1750. From Wikipedia and this photograph (by S-Kay) is public domain

Inaudible low frequencies below 20Hz distort the sound in the audible range. Therefore, infrasound is carefully removed in the production of music software. On the other hand, as mentioned earlier, in the Middle Ages, infrasonic was used in music as Bach used. Acoustic studies have shown that pipe organs can generate from 10Hz to 10,000Hz and above [4)].

Pipe organs require 7-meter-long pipes to generate infrasound. Because the low-frequency wavelength is long. Also, because the weight is 150 Kg, the building sinks after the pipe organ is delivered. It will take a long time for the pipe organ to stop subsidence and the building from becoming structurally stable. Why did it generate inaudible low-frequency sound at such a cost?

Infrasound is inaudible. Besides, there are no known musical instruments that generate infrasound other than pipe organs. If there is an instrument that produces infrasound, it cannot produce audible sounds. Therefore, it is natural that music that actively uses infrasound has not been developed.

In contemporary music, some works use infrasound. They use electronically generated infrasound. Franco Schoeman focused on the low frequencies that elephants use for communication, and composed music using low frequencies. He has independently developed a notation for low-frequency composition [5)].

In sound sources such as records, infrasound of 20 Hz or less has been removed. This is because if the sound source contains infrasound, the housing for music playback resonates, causing sound deterioration. The tonearm is a part that picks up sound on a turntable that plays records. The Tonearm has a record needle and makes physical contact with the record, so the record physically vibrates the Tonearm. The resonance between the vinyl record and the tonearm distorts the sound. It causes Tonearm resonance when included in a sound source of 20 Hz or less on a vinyl record [6)]. Therefore, frequencies below 20Hz have been removed.

2. Add a super constant frequency to the sound source

Fourier has shown that a wave with periodicity is the sum of sine waves of any complexity. The Fourier series of the periodic function x (t) with the interval (-T / 2, T / 2 as one cycle can be described as follows because the

$$x(t) = \frac{a_0}{2} + a_1\cos(t) + b_1\sin(t) + a_2\cos(2t) + b_2\cos(2t) + \cdots = f(x) = a_0 + \sum_{n=1}^{\infty}\left(a_n\cos\frac{2n\pi t}{T} + b_n\sin\frac{2n\pi t}{T}\right)$$

waves to be added have orthogonality (Fig. 3). We observe the spectrograph of audible part (higher than 20Hz) does not change compared with the case when

15Hz has added (Fig. 3). This is obvious from the Fourier analysis and pattern of spectrographs do not change.

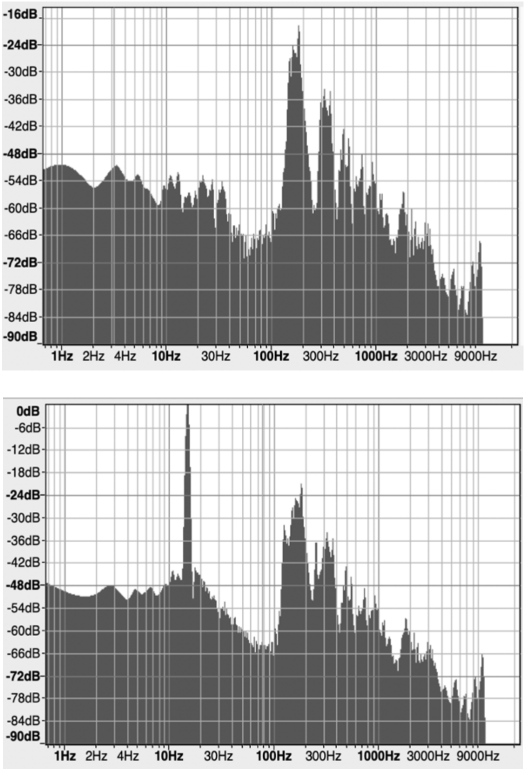


Fig. 2. Example of frequency spectrograph when infrasonic sound added to a sound. Upper: Frequency Spectrograph of No infrasound added, Lower: 15Hz added to the sound. A steep peak of 15Hz shows the added 15Hz. And the frequency spectrographs higher than 20Hz show no differentiations. We used the female sample voice in Mexican, which is given as a sample of the Audition, which is software produced by Adobe Co. Ltd.

As mentioned above, low-frequency sound may induce physical vibration of the playback equipment. Therefore, when a sound source such as a record or Compact Disc, CD is expanded by Fourier series, waves of 20 Hz or less are not included. Therefore, adding infrasound of 20Hz or less to a commercial sound source does not affect each wave of 20Hz or more.

Add sounds below 20Hz to sounds that do not include sounds below 20Hz and analyze the frequency spectrum. The part of the frequency spectrum above 20Hz does not change before and after adding sound below 20Hz.

3. Method and Material

We analyzed Schumann's second-movement -- Symphony No 3 in E-flat major, Op 97, We analyzed the sound intensity by adding a 15 Hz sine wave to this sound source [8]. Besides, frequency spectrum analysis was performed using Audacity 2.4.2.

3.1. *Acoustic sound pressure analysis*

We have proposed analyzing the time change of sound

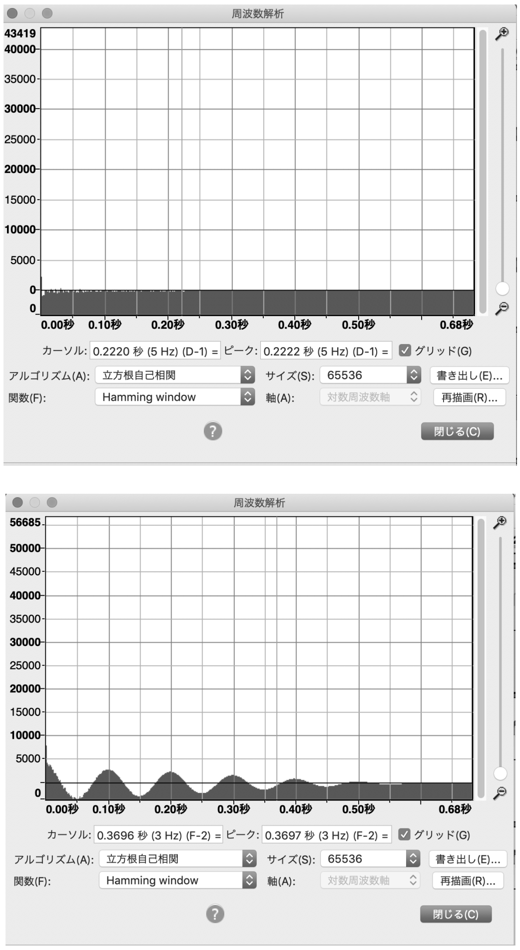


Fig. 3. Upper) Autocorrelation of Schumann's symphony, Bottom) Autocorrelation when 15Hz is put in the same sound source. Autocorrelation is cube root autocorrelation.

pressure of voice data [8]. The method is described below. First, the sound pressure is discretized as follows.

$$F(p) = p_i \,, if\ (b_i \leq p < b_{i+1}),$$

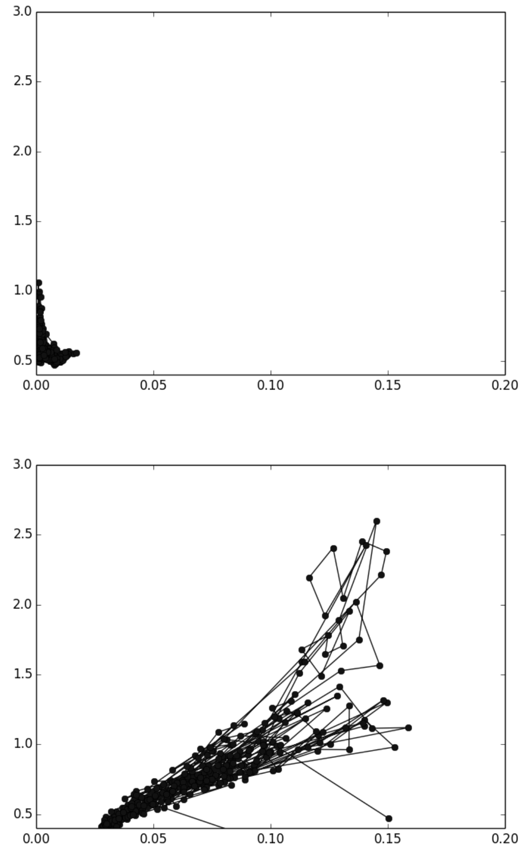


Fig. 4. Upper) Acoustic sound pressure analysis of Schumann's symphony, Bottom) Autocorrelation when 15Hz is put in the same sound source.

where

$$\beta = \frac{p_{MAX} - p_{min}}{n}, \; b_{i+1} = b_i + \beta \, .$$

Next, the sound source is divided into t seconds and averaged. Then, assign bi to pi of each t. From the above, the sound source is converted into a sequence of pi. Calculate the mean and standard deviation of this sequence of points. Then, plot the mean on the vertical axis and the standard deviation on the horizontal axis. Then, the time change of the intensity of the sound source is converted into a two-dimensional pattern.

Even if infrasonic was mixed with ordinary music, there was no difference in the audible part's frequency spectrum. However, when the autocorrelation was measured, it was confirmed that the autocorrelation had a structure when 15 Hz was mixed. Also, acoustic sound pressure analysis showed that mixing 15 Hz made a big difference.

So far, there has been little scientific evidence about the use of infrasonic. However, there are relatively many proven effects. For example, Dave of hard rock band Metallica[5), Dave Pickerel focused on the 16Hz used in Bach's "Toccata and Fugue" mentioned earlier. And he created a subwoofer that was about the size of a guitar case. Then, I selected a song that contained a lot of deep bass from their songs and outputs it from this subwoofer. The super-heavy bass was exposed to whiskey being brewed in a barrel. A whiskey appraiser has tasted the whiskey and confirmed that it is a good quality whiskey. Such examples are used in processes involving fermentation such as winemaking. However, little scientific research has been done.

This study showed that adding super-heavy bass to music causes a large change in the temporal change of acoustic sound pressure. This result opens up possibilities for the use of infrasonic, which was previously unknown.

References

1. Alfred J.Lotoka, Elements of Physical Biology, Williams and Wilkins Company, 1925.
2. Dimas Martins Afonso, Gjini Erida, Modeling Competitive Mixtures With the Lotka-Volterra Framework for More Complex Fitness Assessment Between Strains, Frontiers in Microbiology, 11, 2132, 2020.
3. Takabayashi J., Dicke M. (1997) Herbivore-Induced Plant Volatiles with Multifunctional Effects in Ecosystems: A Complex Pattern of Biotic Interactions. In: Abe T., Levin S.A., Higashi M. (eds) Biodiversity. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-1906-4_9
4. Kerner, Edward H, Comment on Hamiltonian structures for the n-dimensional Lotka–Volterra equations, Journal of Mathematical Physics 38, 1218 (1997); https://doi.org/10.1063/1.531875l, Published Online: 04 June 1998
5. Metallica Talk Top-Secret Distilling Process Behind New 'Blackened' Whiskey, Rolling Stone, SEPTEMBER 15, 2018

Inter-Induce computation and its Philosophical Foundation

Yasuhiro Suzuki
Graduate School of Informatics, Nagoya University, Furocho Chikusa
*Nagoya City, Aichi Prefecture 464/0814, Japan**

E-mail: ysuzuki@nagoya-u.jp.com
www.ysuzuki.info

Abstract

Set theory is based on the distinguishability of elements. How to recognize and identify the world is the essence of set theory. If each element cannot be identified, all the elements are one set. So the set does not make sense. The Heart Sutra is highly rational and can be interpreted mathematically. The mathematical interpretation of the Heart Sutra shows the divergence of how to discriminate. Based on this world view of Heart Sutra, we propose Inter-Induce computation, IIC as a novel calculation paradigm that does not depend on set theory. This paper gives an overview and philosophical foundation of IIC.

Keywords: Inter-Induced computation, Natural computing, Foundation of Computer Science, Heart sutra

1. Introduction

When I started to consider natural computing, when I was studying "the origin of life" using abstract chemistry, the 5th Workshop on Life and Physicality, SOMA 5th~ Theme inter-relationship between "phenomenon" and "things", February 22-23, 2002, I met a priest, Shinsho Kajita, at (at Honen-in temple, Kyoto).

And he suggested to me that "life is unrelenting." I felt something related to the "calligraphy of circle" in every part of the life system and the "circle" of the Zen priest, but at that time, I could not understand the suggestion. Through my research on the origin of life, I finally understood Kajita's "What is life?"

There are several types of the caligraphy of circles (Enso 円相 in Japanese), for example, the calligraphy of the circle of Ikkyu Zen Master is a perfect circle (as shown to the left of the Fig.1). Systematically, this "circle" is in a steady state. Also, since it is a perfect circle, I do not know where it started. That is, there is no time. On the other hand, many calligraphies of circles have "beginning" and "end" from the ink marks, and there is a flow of time (Fig.1 right).

Even if it is a perfect circle, time will flow from one piece to the other if you cut it somewhere. "Dissipation of information" is equivalent to "breaking" a perfect circle by "breaking" a structure that is "steady-state" and creating time.

What if a highly non-linear system with unsteady

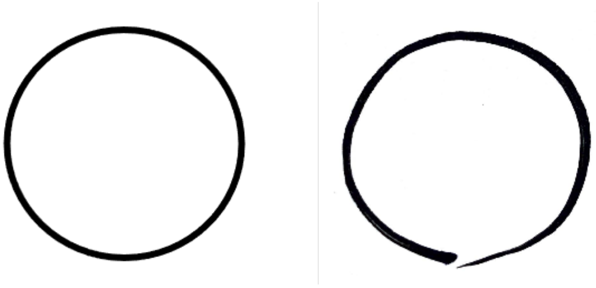


Fig. 1. Example of calligraphy of circles, left) Ikkyu Zen Master (Edo period) created a complete circle calligraphy, right) many calligraphy of circles are not complete circle, which has a “start point” as breaking point; and also density of a stroke illustrates the start-end point

(unsteady) rows (all) does not take a circular structure? If not, all systems have a circular structure in infinite-dimensional state space, it isn't easy to know what the state space looks like unless you can see the entire state space. But if the system has a circular structure, and if it is low-dimensional (two-dimensional or three-dimensional), we can recognize the system.

The calligraphy of a circle that is not closed is topologically a straight line. Then, the straight line disappears at the break of the calligraphy of a circle and regenerates forever. We can recognize the calligraphy of a circle because it is "circular". If it breaks out of the "hanging scroll" as a myriad of straight lines, it is difficult for us to recognize it. The calligraphy of a circle is an example of a case where a system of invariant behaviors can be recognized, such as an attractor of a chaotic dynamical system.

2. Philosophy of Hear Sutra

The Heart sutra (般若心経)[1] is a well-known "sutra". However, its composition is rational and precise. And it is suggested that element reductionism causes bankruptcy. The latter half of the Heart Sutra is a philosophical suggestion. Still, the first half shows that the classification method diverges and becomes meaningless when the theory of element reduction is pursued.

The concept of "colour" is used in the following, but in Buddhism, "colour" corresponds to vision among the five senses. In Chinese, "not unmatch" means "same", but here "not unmatched" is regarded as inclusion. Because, if you want to show "colour, 色 and Kuh, 空 are the same", you can set "colour" = "Kuh". However, in this case, "Kuh" is not specifically shown and becomes undefined.

However, it is possible to define "Kuh" different from "colour" by regarding "Color" ⊆ "Kuh" $ and "Kuh" ⊆ "Color" , "Kuh" is truly included in "Color", and "Color" is truly included in "Kuh". From this equivalent relation, it is mathematically shown that "color" = "Kuh".

"Colorless Kuh, Kuhless colour" are paired. Still, regardless of their religious and philological meanings, mathematically, no matter which one is missing, the identity cannot be proved, so it is inevitable. This is the definition.

"Colorless Kuh" is the definition of the existence of "colour" and "Kuh". Although the existence of "colour" and "Kuh" is defined, this definition claims that "colour" and "Kuh" are the same, and time and dynamics do not occur from this definition. In terms of Enso, it shows the existence of Ikkyu's "completely closed perfect circle." It is the definition of the specific relationship between the two;

Color, that is (即是) Kuh,
Kuh, that is (即是) color.

In Chinese, "即是, immediate correction" means "that is". Therefore, "immediate correction" can be regarded as implies, $\Rightarrow$. $A \Rightarrow B$ by this symbol means "A, that is, if B". Mathematically, $A \Rightarrow B$ and $B \Rightarrow A$ prove that A and B are equivalent. Again, they are opposites, but they must be mathematically paired because they cannot be shown to be equivalent if either one is missing (this $\Rightarrow$ can be regarded as a map, but in category theory. May be regarded as morphism).

In the above, "colour (vision)" of the five senses is discussed, but other senses are omitted because the same discussion is repeated. From this, all "functions of the mind" are equal to "Kuh".

The construction of this concept of "Kuh" is logically clever (aside from religious discussions), and it does not define what "Kuh" is, and what is called "the function of the mind" is individual. It is shown that it cannot be divided into.

If all the "functions of the mind" are equal to "Kuh", the concepts that conflict with the "functions of the mind" such as life-destruction, dirt-purification, and increase-decrease are life = destruction = dirt = purification = increase = decrease. = "Kuh", so everything is the same. In other words, "it cannot be distinguished".

For example, the increase/decrease of "Your juice is more than mine!" Does not hold if you accept the existence of "Kuh". Conversely, if we accept the existence of "Kuh", time-related confrontations such as life-destruction, dirt-purification, and increase-decrease do not make sense (immortality and immortality). Unclean and unclean). Also, as discussed in "colour" and "Kuh", all five senses are the same as "Kuh", so the five senses are all the same from visual = auditory = olfactory = taste = tactile = "Kuh" and cannot be distinguished.

The dynamics are "unsteady", that is, unsteady, but in reality, they are all the same "Kuh", and our perception ("the function of the mind") is "Kuh" and the visual image ("colour") is projected. For example, as Jakob Johann Baron von Uexku[2)] suggests, recognition differs depending on the species even in the same field, but in reality, they are the same field.

Attractor-like mechanical structures can sometimes be seen in (low-dimensional) chaos, but their trajectories are all "solutions" of the same equation. A function does not always have a unique solution if it does not meet the Lipschitz condition. In that case, the solution can be innumerable, and the behaviors of various dynamical systems can be seen depending on the parameters. However, although they are all "unsteady" (unsteady) in terms of dynamics, they are really just "solutions" to the same equation.

On the other hand, the Heart Sutra suggests that recognizing the unchanging natural system leads to the "Kuh." This suggestion is accurate and analytical, and element-reducing science continues to make "lists" of how natural systems work, but this lengthy list, as Yuxcur suggests, can be perceived differently.

Even if a "perfect list" is obtained in the future, it is unlikely that we can understand the natural system that is unrelenting by combining it. An example with a "perfect list" is a cellular automaton. In cellular automata, components and all interactions are perfectly grasped as a finite number of state transition rules. However, that dynamics go beyond our perception and creates a new universe-like world. The universe of cellular automata is vast, and we can only observe it.

3. Framework of Inter-Induced Computation

Based on the Heart Sutra concept, the components of the natural system are subdivided as much as possible. In contrast to searching for the ultimate components, in the following, the components of the natural system are not subdivided but treated as they are. We propose a calculation system.

In formulating the actual nature into a computational system, Heart sutra suggested it is complicated to understand the natural system in element reductionism. It is premised that classification-analysis is impossible. Must be. In that case, it does not become a meaningful set theory. This is because all the elements are the same (one representative source). Therefore, it is necessary to change the basis for constructing a computational system fundamentally.

On the other hand, we do not fully understand the natural system, but we interact and control it to some extent. For example, when communicating with a person, it is impossible to subdivide the other person into elements and fully understand them before communicating.

However, we communicate and compose a society without completely understanding the other person. Then, as I go out with the other person, I gradually feel that I can understand the other person. However, neither knows whether the belief is correct or incorrect (because they cannot know each other's internal state accurately).

Therefore, a framework that guides the target system in the desired direction while interacting with the target, such as communication with humans, is called "inter induce computing".

In the following, the guiding side is referred to as the "inducing subject", and the guided side is referred to as the "inducing target". For the sake of simplicity, in the following, the guiding subject and the guiding target are independent, autonomous systems. It is possible to argue even if they are subordinate to each other. Still, in that case, specifically dependent interactions and independent interactions Since it is necessary to divide into actions and there are innumerable cases, only independent cases are dealt with.

And here is the most important point, "It is impossible to know the internal state of the other party for both mutual subjects and mutual objects by any method." The guiding subject has a "direction/state of being guided" as a solution set. Interactions that bring about change are called "effective actions", and interactions that do not change are called "ineffective actions". In the following, unless otherwise specified, effective action is called an "action".

The inducer "acts" on the induction target. The induction target changes depending on the "action" and leads to a steady-state (including a dynamic steady-state) and a normal set. The guiding subject observes this normal typeset, and if it matches the guided direction and state, it is a "correct answer", otherwise it is a "wrong answer". And from this result, a new "action" may be performed. The inducer decides whether or not to perform a new "action".

4. Remarks

Why is it mutual induction rather than induction?

It is impossible to know the internal state of the other party's system with which any natural / life system interacts, by any method. Therefore, for convenience, we have named the guiding subject and the guiding target. However, the "name" has no essential meaning. Therefore, the mutual induction system is essentially composed of two or more systems.

It is impossible to determine whether any natural or biological system is "inducing or being guided". For example, if I think that someone is "inducing", whether the object to be guided is "induced by me" or "in reverse, I am induced", the "action" is established in the first place. I don't know.

It is impossible to "really know (know the other party's internal state from a transcendental standpoint)".

For example, a virus needs a "host" because it cannot propagate on its own. Emerging infectious diseases such as coronavirus mutate while interacting with the "host". When the "host" is human, humans make vaccines and new drugs and "respond" to the "action" of the virus. In this case, both are the guiding subjects and the guiding targets.

If there is no object of interaction, the concept of induction/delusion does not occur. Self-reflection and delusions are also multiple systems consisting of "self-reflection / delusional subject" and "self-reflection / delusional system" created by that subject. It is a mutual induction. It is not known in principle (unless it is a transcendental existence) which system is mutual subject or mutual object, and even how many of them are unknown.

The essence of induction is that there is an interaction partner. It doesn't matter if the "other party" actually exists. To clarify this, it is called mutual induction.

Isn't mutual induction a feedback control?

The same argument applies to any control system, but when controlling, the control subject and the control target are generally clearly defined, and the internal state of the control target can be known as a state quantity such as temperature or velocity, for example. , In autopoiesis, it is not possible to clearly separate the control subject and the object, but it is part of the definition of autopoiesis, and the discussion, in that case, is similar to mutual guidance. Therefore, in control, the action-response of the controlled object is formulated. It can be handled mathematically.

On the other hand, in mutual induction, it is premised that the guiding subject and the guiding target cannot be separated in principle. Other systems can't know the internal state of the system. It is possible to regard a part of mutual induction as feedback control, etc.. Still, from that point alone, it is an over-generalization and an error to assume that the concept of mutual induction is the same as feedback control.

References

1. Heart Sutra
2. Jakob Johann Baron von Uexküll, Streifzüge durch die Umwelten von Tieren und Menschen: Ein Bilderbuch unsichtbarer Welten, 1934

Digital Testing Device for Active Range of Motion of Finger Joints Utilizing Artificial Neural Network

Huu-Hieu Quang
Department of Electrical and Mechanical Engineering, Graduate School of Engineering
Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Aichi 466-8555, Japan

Yoshifumi Morita
Department of Electrical and Mechanical Engineering, Graduate School of Engineering
Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Aichi 466-8555, Japan

Makoto Takekawa
everfine, LLC, Japan
Email: 31513006@stn.nitech.ac.jp, morita@nitech.ac.jp,
makoto.takekawa@gmail.com
www.nitech.ac.jp

Abstract

This paper proposed a digital goniometer device for measuring the active range of motion (A-ROM) of finger joints by artificial neural network (ANN). A therapist typically measures the A-ROM by a conventional goniometer. However, it consumes a significant amount of time, and imposes a burden on patients. Therefore, we develop a digital goniometer using ANN to overcome these problems. The effectiveness of the proposed device is verified based on the evaluation error on the third joint of the middle finger.

Keywords: Digital goniometer device, Active range of motion, Artificial network, Conventional goniometer.

1. Introduction

The range of motion (ROM) is an essential part in clinical practices when assessing the effectiveness of therapeutics. A therapist typically measures the ROM by conventional, as shown in Fig. 1, or digital goniometers.[1,2]

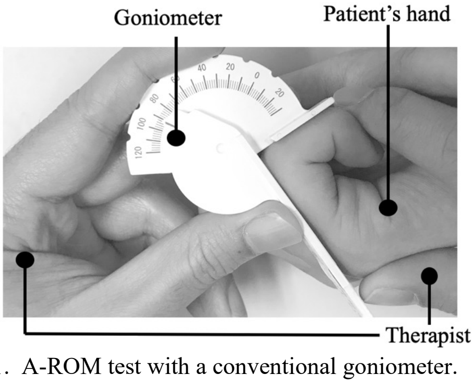


Fig. 1. A-ROM test with a conventional goniometer.

However, he/she cannot carry out the measurement in a short time by these goniometers. Accordingly, several measuring methods were proposed by extant studies.[3,4] A VICON system using markers placed on each joint of the fingers were proposed in order to compute the joint angles. In addition, a non-invasive automatic testing device was developed to boost the measurement time and accuracy. Nevertheless, the developed devices cannot measure the ROM precisely in a short time, and the response of the systems during operation is poor. In addition, a non-invasive automatic testing device was developed to boost the measurement time and accuracy. Nevertheless, the developed devices cannot measure the ROM precisely in a short time, and the response of the systems during operation is poor.

In this study, a digital testing goniometer is developed to overcome these problems. A data collection model is proposed to collect training dataset of the finger joints. Subsequently, an ANN model is utilized to validate the dataset. Hence, it can facilely compute the finger joint angles. From the viewpoint of improving measurement accuracy, the effectiveness of the proposed Digital testing goniometer is evaluated on the third joint of the middle finger in comparison with the conventional goniometer.

2. A-ROM Testing Method and Device

2.1. *Outline of digital testing device*

The active range of motion (A-ROM) is the angle when a patient bends or extends his/her own joints as much as possible. Meanwhile the passive range of motion (P-ROM) is the joint angles when the therapist bends or extends the patient's joints as much as possible. Figure 2 shows an overview of digital goniometer device where the depth camera (Intel RealSense Depth Camera SR300) was connected to the PC via the USB gate. The depth image of depth camera has 640 × 480pixel resolution, and a trigger frequency of 30Hz.

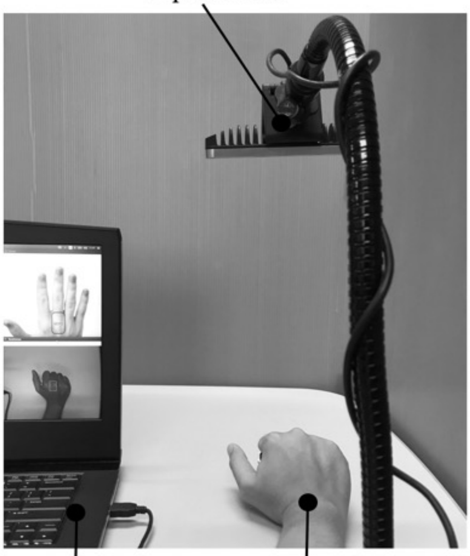
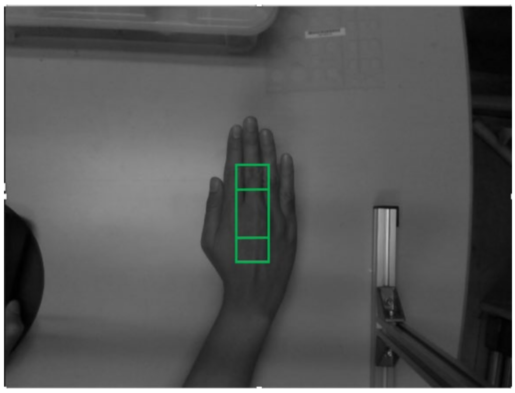


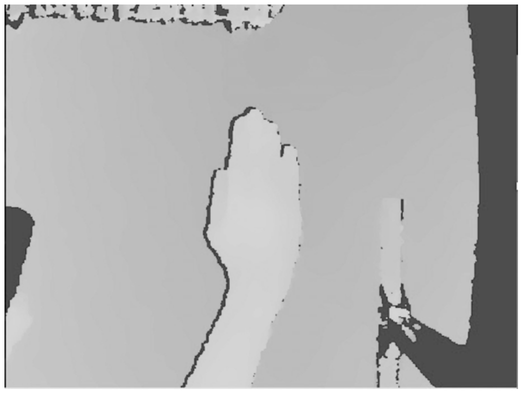
Fig. 2. Digital testing goniometer.

2.2. *A-ROM testing method*

In this study, the metacarpophalangeal joint (MP) of the middle finger is considered. The MP joint is the third joint and the joint between two bones, namely the metacarpal and the wrist bones. The range of motion of MP joints in healthy people is 0 to 90 deg (flexion) and 0 to 45 deg (extension). Let the fully extended position be the neutral position of 0 deg. During A-ROM testing, the therapist moves the depth camera towards to the patient's hand so that the middle finger is into the three-green boxes, as shown in Fig. 3(a). The three green boxes are overlaid on the image of RGB camera. The interface screen shows the taken images on both RGB and depth modes simultaneously, as shown in Fig. 3.



(a) RGB image



(b) Depth image

Fig. 3. Interface of system.

3. A-ROM Measurement by ANN Method

3.1. *Principle*

The points of the proximal interphalangeal, metacarpophalangeal joints of the middle finger and the center point of wrist bones are denoted by PIP3, MP3, and W, respectively, as shown in Fig. 4. The XYZ-axes of the depth camera coordinate system are shown in Fig. 4. The three-dimensional position coordinates of three points are measured with the depth camera. The unit of X- and Y-axes is pixel. The unit of Z-axis is meter. The coordinate of each joint is determined by RGB image and Depth image data. It should be noted that the units are different. Subsequently, the joint angle is predicted by using an ANN model since there is a linear regression relevance between the joint's position coordinate and joint angle.

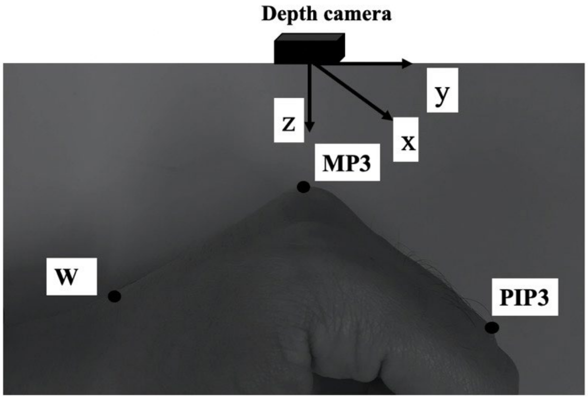


Fig. 4. Positions of the phalangeal joints of the middle finger.

3.2. *Data collection*

We constructed an ANN model from three-dimensional position coordinates of W, MP3 and PIP3 joints to the MP joint angle. We created data collection model for finger joints position detection by using Hand Keypoints, which was a well-known open-source model in the field of image processing for high predictive accuracy of fingers joints recognition.[5] Using the image data acquired by the RGB camera and the position of three points determined by the Hand Keypoints, the two-dimension position coordinates of three joints are obtained in pixel units. Moreover, using the depth data acquired by the depth camera, the depth coordinates corresponding the two-dimensional position coordinates of three points are obtained in meters.

Moreover, three green boxes are set so that the therapist can easily locate the patient's hand, as shown Fig. 5. The overall size of three boxes is 40 x 120 pixel, the size of each green box is 40 x 30 pixel, 40 x 60 pixel, and 40 x 30 pixel from the top to the bottom, respectively. The position of PIP3, MP3, and W joints are marked by red, blue, and green points, and are denoted by I, J, and K points in Fig. 5, respectively. The MP joint angle was determined by the three points. The input dataset consists of the coordinates of these points. In addition, these points are features of ANN model.

$$I_i = (x_{Ii}, y_{Ii}, z_{Ii}), J_i = (x_{Ji}, y_{Ji}, z_{Ji}), K_i = (x_{Ki}, y_{Ki}, z_{Ki})$$

$(i = 1, 2, \dots, n)$, n is the number of data contained in the dataset. When the data is acquired for constructing and evaluating a ANN model, the subject moved the hand until I, J and K points are inside the corresponding green boxes.

The output of the dataset is the MP joint angle, which was measured with a conventional goniometer. When measuring the joints of the fingers, the joints were fixed by sandwiching clay between the joints so that the joint angle of the fingers did not change. Table 1 shows the MP joint angle measured with the conventional goniometer, which are used for constructing a ANN model and evaluating the accuracy.

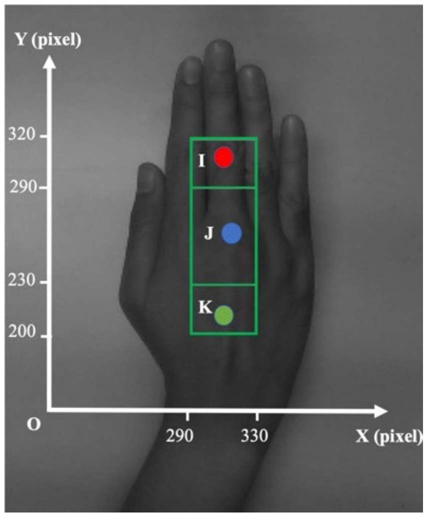


Fig. 5. Processing image to collect data.

Table 1. Measured values of MP joint angle with the conventional goniometer.

Pose No.	1	2	3	4	5	6
Angle[deg]	38	50	64	68	73	80

3.3. *Artificial Neural Network*

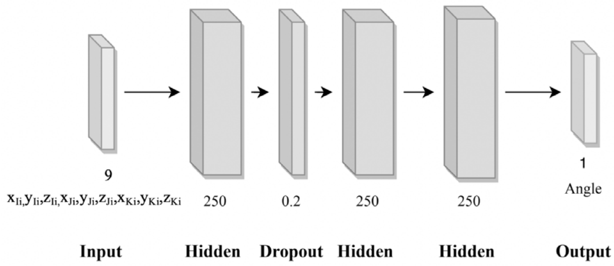


Fig. 6. Artificial neural network architecture.

The principle of the method is creating a model by ANN.[6] The model is determined based on the linear relationship between the joint position and joint angle of sample dataset. Subsequently the model is used to predict the joint angle based on the joint position. We used ANN model for regression to predict angle value. The ANN model used Rectified Linear Unit (ReLu) activation function in both layers. The loss function utilized for regression was Mean Squared Error (MSE). In addition, Adam was chosen as the optimizer function. The architecture of the artificial neural network is shown in Fig. 6.

ANN model was configured of three middle layers of 250 neurons represented by ReLu activation function. In addition, a dropout layer was added as a regularization technique in regression network. The dropout layer randomly drops some of the inputs to a layer is 0.2.

4. Evaluation of Measurement Predicted Accuracy

The evaluation of measurement accuracy in A-ROM testing was conducted to investigate the effectiveness of the proposed digital goniometer device. For this purpose, 75% of the dataset was randomly selected for constructing the ANN model and the remaining 25% was assigned for validation. The effectiveness of the proposed ANN model was evaluated base on the 25% dataset. Figure 7 shows the error between the measured angle with the conventional goniometer and the predicted angle by the proposed testing device. The horizontal axis denotes the trials of 1447. Figure 8 shows the cumulative histogram of the errors. It was found that the errors of 98.69% and 99.52% trials were less than 2 deg and 5 deg, respectively.

5. Conclusion

In order to solve the problems faced when using a conventional goniometer, we proposed a digital goniometer device for measuring the active range of motion of the MP joint of the middle finger. We constructed an estimation model of the finger joint by using RGB and depth data of the finger joint, the Hand Keypoints, and the ANN model. The estimation accuracy of the MP joint of the middle finger was less than 2 deg. The effectiveness of the proposed device was demonstrated.

In our further study, we plan to extend the proposed method so that it can be applied to 12 finger joints. Moreover, we plan to estimate the data collection model to remove outlier for increase high performance of the ANN model as well as reduce error value.

References

1. Goniometer for fingers, SAKAI Medical Co.,Ltd., https://www.sakaimed.co.jp/measurement_analysis/human-body-measurement/hand-goniometer/ (In Japanese).
2. easyangle, Ito Co., Ltd., http://www.medical.itolator.co.jp/product/easyangle/ (In Japanese).
3. J. R. Cook, N. A. Baker, R. Cham, E. Hale and M. S. Redfern, Measurements of Wrist and Finger Postures: A Comparison of Goniometric and Motion Capture Techniques, *Journal of Applied Biomechanics*, 23, 2007, pp.70-78.
4. Reissner L, Fischer G, List R, Taylor WR, Giovanoli P, Calcagni M, Minimal detectable difference of the finger and wrist range of motion: comparison of goniometry and 3D motion analysis, *J. Orthop Surg Res*, 14(1), 2019.
5. Simon, Tomas & Joo, Hanbyul & Matthews, Iain & Sheikh, Yaser. Hand Keypoint Detection in Single Images using Multiview Bootstrapping, *Cornell University, USA,* 2017.
6. Lau, E.T., Sun, L. & Yang, Q. Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Appl. Sci.* 1, 982, 2019.

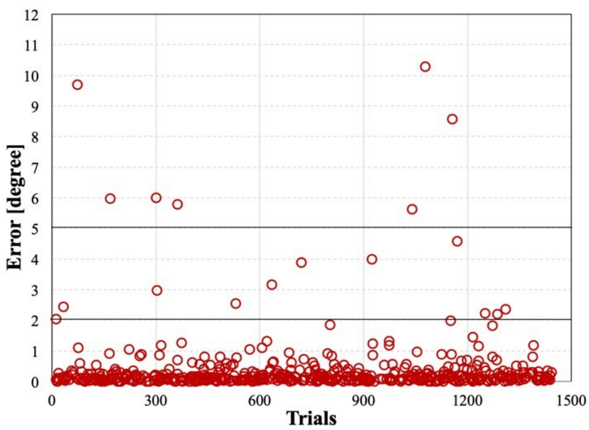


Fig. 7. Distribution of the errors of predicted angles.

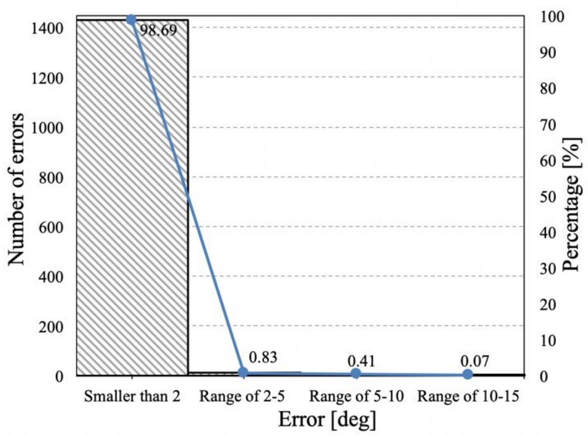


Fig. 8. Histogram of the errors of predicted angles.

Position and Force Teaching Method for 6 DoF Manipulator Using Contact Teaching Tool and Teaching Data Editor

Duy-Do Bui
Department of Electrical and Mechanical Engineering, Graduate School of Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Aichi 466-8555, Japan

Hiroki Tanaka, Quang-Trung Chu, Hideki Inuzuka, Yoshifumi Morita
Department of Electrical and Mechanical Engineering, Graduate School of Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Aichi 466-8555, Japan

Masao Sakai
Industrial Research Center, Aichi Center for Industry and Science Technology
Onda-cho 1-157-1, Kariya-shi Aichi 448-0013, Japan
E-mail: cmc13126@nitech.jp, h.tanaka.683@nitech.jp,
h.inuzuka.776@nitech.jp, morita@nitech.ac.jp, sakai@aichi-inst.jp
www.nitech.ac.jp

Abstract

This paper proposes a direct teaching method for teaching tasks performed by a robot manipulator while the robot's tool makes contact with a workpiece. A contact teaching tool designed for accurately grasping the shape of the workpiece, and a teaching data editor for generating the target trajectory to realize the desired press angle and force, are developed on the assumption that the compliance characteristics of the robot hand are known. The effectiveness of the developed method is confirmed experimentally.

Keywords: Direct teaching method, Robot manipulator, Contact teaching tool, Teaching data editor.

1. Introduction

In the manufacturing industry, efficient robot teaching technologies are essential to bring out the fullest potential of robot manipulators. Currently, for the teaching task of robots, the teaching pendant is widely employed. However, it is difficult to teach complex trajectories by the teaching pendant. Moreover, this method consumes a significant amount of time to teach robots. Therefore, new direct teaching methods have been investigated.[1,2] To solve these problems, a parallel wire-type teaching device (PAWTED) was developed to directly teach robot manipulators using human hands instead of teaching pendants.[3] Nonetheless, it is not possible to determine teaching points precisely by the PAWTED since the operator has to visually checks and adjusts by hand.

Therefore, in this study, a contact teaching tool and a teaching data editor were proposed on the assumption that the compliance characteristics of the robot hand are known. The tool allows us to accurately grasp the shape of the workpiece just by pressing the teaching tool against the edges of the workpiece. Subsequently, the editor will generate the target trajectory to realize the desired pressing angle and force based on the obtained teaching data and the robot hand characteristics. The effectiveness of the proposed method was confirmed experimentally from the viewpoint of improving the teaching accuracy and reducing the teaching time.

2. PAWTED with a contact teaching tool

The PAWTED consists of a moving platform, six wires, and a base. The base, which is attached to an industrial six-degree-of-freedom (6-DoF) manipulator, has six rotary encoders and six flat spiral springs. The moving platform is connected to the base by six metal wires. As a result, the moving platform can move freely in three-dimensional space within the working area.

The operation of the PAWTED comprises teaching and playback modes. In the teaching mode, when the operator moves the moving platform, the system will compute the position of the end-effector based on the drawn length of the six wires. Then, the robot will be controlled to track and maintain a constant distance with the moving platform. In the playback mode, the moving platform is fixed to the base. Subsequently, the teaching trajectory is reproduced by the robot.

In this research, we aimed to develop a contact teaching tool that allows the teaching position to be obtained easily just by touching the contact teaching tool to the edge of the workpiece. Therefore, a contact teaching tool with a design as in Fig. 1 (a) is proposed. As shown in Fig. 1 (b), p_a and p_b are the two teaching points of the teaching tools. The midpoint of the line segment connecting p_a and p_b, which is the rotation center point of these teaching points is denoted as p_0. To improve the flexibility of the teaching tool, it is designed so that it can rotate freely around its central axis as depicted in Fig. 1 (a). Besides, due to the empty v-shaped structure, the contact teaching tool could be used to make contact with both straight edge and arc-shaped edge types workpieces.

In the teaching mode, the contact teaching tool is attached to a moving platform, as illustrated in Fig. 2. When the operator touches the contact teaching tool to the edge of the workpiece, p_a and p_b will make contact with the edge of the workpiece and the position and orientation of the teaching tool are determined. After teaching the trajectory, the editor calculates the trajectory of the teaching points, and desired trajectory on the playback mode. In this study, a robot hand with compliance characteristics and a robot's tool attached to the robot hand were simulated as a joystick in the playback mode. The compliance characteristics of the robot hand are considered to be the compliance characteristics of the joystick.

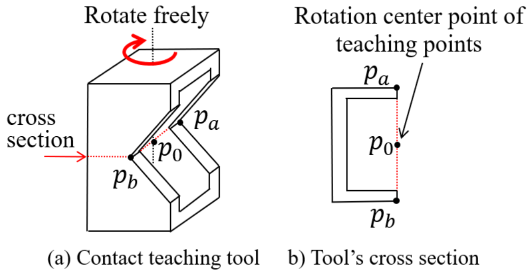


Fig. 1. Contact teaching tool structure

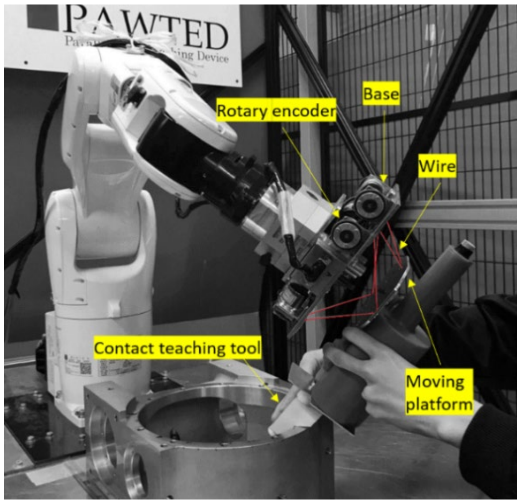


Fig. 2. PAWTED with a contact teaching tool

3. Teaching data editor

A teaching data editor generates the target trajectory to realize the desired pressing angle and force based on the teaching trajectory and the characteristics of the contact teaching tool and playback tool.

A model diagram of the computing process in order to obtain the circular arc-shaped trajectory for the robot is illustrated in Fig. 3. p_{ai} and p_{bi} are the teaching positions obtained when the contact teaching tool touches the edge of the workpiece ($i = 1, 2, \cdots, n$). The rotation center point of p_{ai} and p_{bi} is denoted as p_{0i}. S is the auxiliary circular trajectory created by multiple p_{0i} points which can be calculated based on the PAWTED moving platform. C_0 is the center of S and the hole of the workpiece. The radii of S and the hole are r_1 and r_2, respectively. After at least three rotation center points p_{0i} are obtained, the least squares fitting method is applied to precisely determine C_0 and r_1 from multiple p_{0i} points.

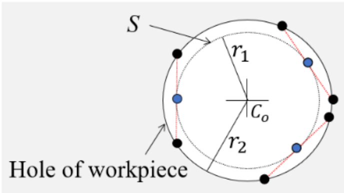


Fig. 3. Getting the hole trajectory

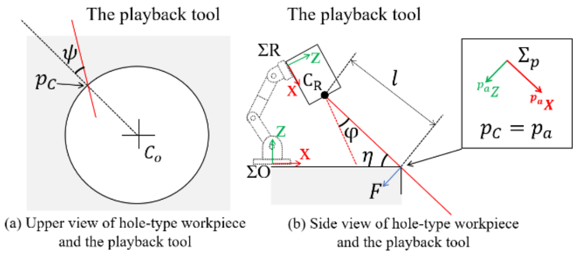


Fig. 4. Hole-type workpiece and playback tool.

Afterwards, the position of an arbitrary number of the teaching points p_{ai} and p_{bi} can be determined from S. Subsequently, the least squares fitting method is adopted once again to acquire the radius r_2 of the circular trajectory that we need to teach the robot. From the newly obtained teaching trajectory, the tool orientation at the time of teaching can be changed to the required orientation.

In Fig. 4, p_C is denoted as the contact position between the playback tool and the edge of the workpiece. C_R is the center pivot point of the playback tool and l is the distance between C_R and p_C. The angle between the line that connect p_C and C_0 and the playback tool is denoted as ψ, as shown in Fig. 4 (a). In Fig. 4 (b), F is the force that the playback tool presses against the workpiece. η is the angle between the workpiece surface and the playback tool when viewed from the side of the workpiece. The inclination angle between the symmetrical axis of the playback tool and its joystick when it is pressed against the workpiece is denoted as φ. The pressing force F is adjusted based on φ. Based on the compliance characteristics between φ and F, φ can be approximated as a function of F as follows:

$$\varphi = f(F) \tag{1}$$

In this study, for simplicity of calculation, we chose p_a as our teaching point. In the coordinate frame Σ_P which is attached to p_a, the position and orientation of the teaching point p_a is calculated to be the desired contact position and orientation of the robot's tool. The position and orientation of C_R as follows:

$$ {}^{O}_{C_R}T = {}^{O}_{p_C}T \begin{bmatrix} {}^{p_C}_{C_R}R & {}^{p_C}_{C_R}X \\ 0 & 1 \end{bmatrix} \tag{2}$$

in which ${}^{O}_{C_R}T$ is the simultaneous transformation matrix from frame Σ_O to Σ_{C_R}. ${}^{O}_{p_C}T$ is the simultaneous transformation matrix from frame Σ_O to Σ_{p_C}. ${}^{p_C}_{C_R}X$ denotes the translation of $-l$ in the X axis of frame Σ_P and ${}^{p_C}_{C_R}R$ denotes the rotation of $-\varphi$ about the Y axis of frame Σ_P.

$$ {}^{p_C}_{C_R}R = R_Y(\varphi) \tag{3}$$

$$ {}^{p_C}_{C_R}X = [-l \quad 0 \quad 0]^T \tag{4}$$

Finally, since the simultaneous transformation matrix from the center of rotation frame Σ_{C_R} to the robot end-effector frame Σ_R is known as ${}^{C_R}_{R}T$, the target position and orientation R of the robot at the teaching points is obtained as:

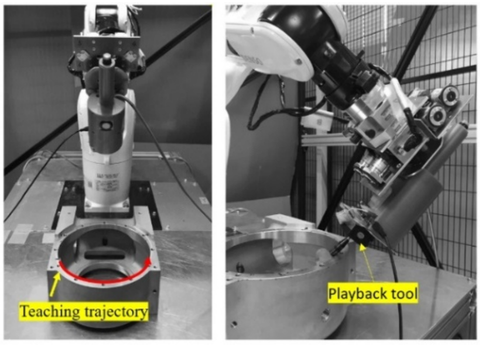
$$ {}^{O}_{R}T = {}^{O}_{C_R}T\,{}^{C_R}_{R}T \tag{5}$$

where ${}^{O}_{R}T$ is the simultaneous transformation matrix from frame Σ_O to Σ_R whose position and orientation are target coordination of the robot end-effector to realize the desired trajectory.

4. Experiment

In order to verify the effectiveness of the new direct teaching method with the PAWTED, a contact teaching tool, and a data editor (PAWTED + ce), experiments were carried out on the 6-DoF industrial robot (DENSO WAVE, VS-060). Specifically, we have performed the experiments with the task of teaching a circular arc-shaped trajectory for the robot, as shown in Fig. 5 (a). The experimental results are compared with the other two methods that used a teaching pendant and PAWTED. For evaluation purposes, each of the three teaching methods is performed three times, and their average value is used.

In the teaching mode, to create the same teaching condition, the operator taught the trajectory that makes the desired values of $l = 45$ mm, $\psi = 0°$, $\eta = 45°$, and $F = 125$ gf for the robot, as illustrated by the red lines in Fig. 6-9, respectively. Afterwards, in the playback mode,



(a) Circular arc trajectory (b) Teaching conditions of the tool

Fig. 5. Teaching trajectory and conditions.

a joystick is used as a playback tool to perform the evaluation experiments, as illustrated in Fig. 5 (b). Based on the compliance characteristics of the joystick, the target position of the manipulator on the playback mode can be determined according to Eq. (5).

In this study, our objective is optimizing the playback trajectory so that the robot can make a contact with the workpiece precisely. The parameters l, ψ, η and F for the desired positions and orientations of the robot's tool are evaluated. The experiment results of l, ψ, η and F, which are analyzed with the unequal variances t-test, are shown in Figs. 6-9, respectively. * shows statistically significant at $p < 0.05$. Regarding the maximum average errors of each parameter, among the three teaching methods, the PAWTED + ce has standout performances.

The teaching-editing time of the three teaching methods for a circular arc trajectory with 11 teaching points is shown in Table 1. Since the position and orientation of the PAWTED + ce method are determined just by touching the teaching tool to the edge of the workpiece, it is possible to quickly teach any point on the edge of the workpiece for the robot. As shown in Table 1, the average teaching-editing time in case of using PAWTED + ce, PAWTED, and the teaching pendant are 11 sec, 29 sec and 104 sec, respectively. PAWTED + ce method reduced the teaching-editing time significantly.

Table 1. Teaching-editing time for a circular arc trajectory.

	Average	Standard deviation
Teaching pendant	104 sec	46.0 sec
PAWTED	29 sec	13.0 sec
PAWTED + ce	11 sec	2.4 sec

5. Conclusion

This study presented a new method for teaching tasks performed by a robot manipulator while the robot's tool makes contacts with the workpiece. We investigated the performance of the proposed teaching method in the task of teaching a circular arc-shaped trajectory for the industrial robot. From the experimental results compared with the conventional methods, it has been shown that the proposed teaching method has the highest teaching position accuracy and the shortest teaching time. A future plan is to verify the usefulness of the proposed method in the actual field. This research was supported by Aichi Prefectural Government through "Knowledge Hub Aichi," Priority Research Project (II) and in the experiments by Kondo Seisakusho Co., Ltd. and FINE TECHNO Co., Ltd.

References

1. S. Masaru, Direct Teaching for Robot, *J. Robotics Society of Japan*. **13**(5), 1995, pp. 627-628.
2. Y. Mochizuki and T. Matsui, Direct Teaching Method for Hybrid Control Applying Impedance Control, *J. Robotics Society of Japan*. **12**(2), 1994, pp. 160-166.
3. S. Sakai, S. Noritaka, M. Yoshifumi, Development of Parallel Wire Mechanism for Robot Teaching, *J. Japan Society of Mechanical Engineers Series C*. **79**(807), 2013, pp. 4321-4329.

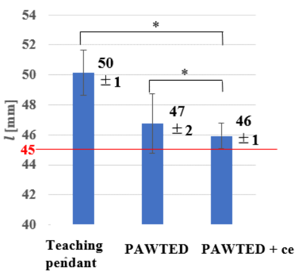


Fig. 6. Teaching distance l on the teaching points.

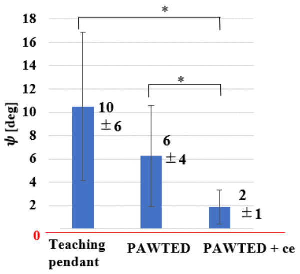


Fig. 7. Teaching angle ψ on the teaching points.

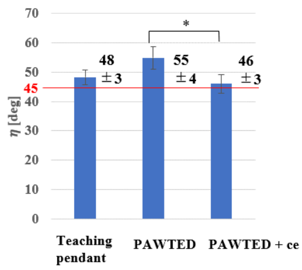


Fig. 8. Teaching angle η on the teaching points.

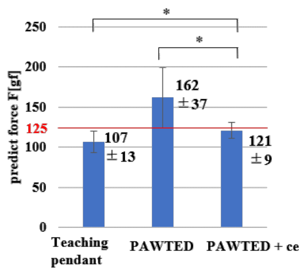


Fig. 9. Teaching force F on the teaching points.

Haptic Device that Presents Sensation Corresponding to Palm on Back of Hand for Teleoperation of Robot Hand
Report2: Consideration on Decided Specification

Kyosuke Ushimaru, Noritaka Sato, Yoshifumi Morita
Department of Electrical and Mechanical Engineering, Graduate School of Engineering,
Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Aichi 466-8555, Japan
E-mail: k.ushimaru.884@nitech.jp , sato.noritaka@nitech.ac.jp

Abstract

Recently, teleoperated rescue robots are required. However, it is known that the teleoperation of a robot hand mounted on a rescue robot is difficult. Therefore, we propose a new haptic device that presents haptic sensation for teleoperation of a robot hand. The device stimulates the back of the hand instead of the palm of the operator. Determination of required specifications by an experiment with subjects is written in this paper. To design the device, the interval of the stimulation points (*i*), the diameter of the stimulation point (*d*), and the force of the stimulation (*f*) should be optimized. As a result of the experiment, we found that the accuracy rate is highest, when (*i*, *d*, *f*) = (30mm, 6mm, 0.9kgf). Moreover, we deeply considered on the decided specification with an additional experiment.

Keywords: Rescue robot, Haptic Device, Teleoperation, Robot Hand, Palm

1. Introduction

Since the Great Hanshin-Awaji Earthquake in 1995, research and development of disaster response robots have been conducted [1]. In recent years, there has been a demand for disaster response robots that can not only gather information but also perform tasks [2]. For this reason, disaster response robots equipped with robotic hands have begun to be developed so that they can work at disaster sites in the same manner as humans [3].

Since disaster response robots were operated in an unknown and extreme environment, they are often controlled by teleoperation [4]. In order to improve the efficiency of teleoperation, it is necessary to provide feedback to the operator on the state of contact between the fingers and palms of the robot hand. However, in the field of robot teleoperation, there are some systems that provide feedback on the contact state of the fingers [5][6], but we cannot find systems that provide feedback on the contact state of the palm.

In this study, we propose a method of substituting the contact state obtained from the palm of the robot hand to the back of the operator's hand. Okano et al. [9] have proposed a method of substituting the tactile presentation to other parts of the hand. However, Okano et al. substituted the sole of the foot for the tactile presentation of the whole hand. In contrast, this study substitutes the back of the hand for the tactile presentation of the hand.

In the previous paper[11], we clarified the characteristic of the tactile sensation of the back of the hand in order to formulate the specifications required for the device that substitutes the back of the hand, as the first step of the research.

We prepared three candidates for each of the three types of stimulus points, i.e., the distance between the stimulus points, the diameter of the stimulus points, and the magnitude of the stimulus force.

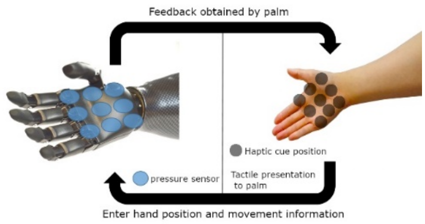


Fig. 1. Ideal method.

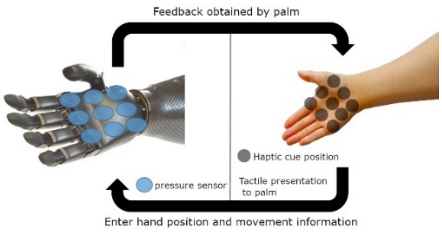


Fig. 2. Proposed method.

Fig. 3. Designed device with parameters which are obtained by experiment.

In this paper, we report the overview of the previous paper and the results of an additional experiment where the number of the stimulus points are one or two and the subjects wear gloves to verify whether the tactile sensation can be presented as a surface instead of a point, without changing the specifications.

2. Proposed Device

2.1. *Overview*

As shown in Fig. 1, it is ideal to present the tactile sensation on the palm of the hand in order to correctly convey the state of contact between the robot hand and the object to the operator. However, this method would interfere with the teleoperation of the robot hand, so we propose a method of presenting tactile sensations on the back of the hand as shown in Fig. 2. The following parameters are necessary to design the device:

- The interval between the stimulation points (interval: *i*)
- The diameter of the stimulus point (diameter: *d*)

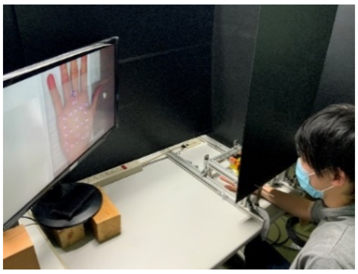


Fig. 4. Experimental scene.

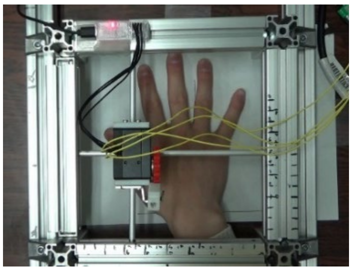


Fig. 5. Device for experiment.(Top view).

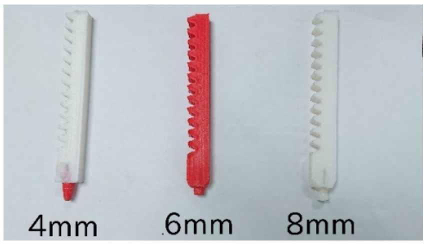


Fig. 6. Rod of experiment device

- The force of the stimulus (force: *f*)

It is known that human skin sensation has a characteristic such as a two-point discrimination threshold.
Therefore, in the previous paper [11], we clarified the characteristics of the tactile sensation on the back of the human hand by subject experiments described in the next section and determined these specifications. Fig. 3 shows a conceptual view of the device designed by using the specifications determined by the experiments.

2.2. *Subject Experiment for Specification*

In this section, we report the optimal combination of the above parameters determined through a subject experiment. The procedure of the experiment is as follows: 10 subjects are prepared, and 10 points on the back of their hands are pressed one by one. The subjects were asked to answer where on the palm the pressed point corresponded to on the back of the hand. The interval of the stimulation points (*i*), the diameter of the stimulation point (*d*), and the force of the stimulation (*f*) were the parameters of the combination. The combination with the highest ratio of correct answers is adopted as the required

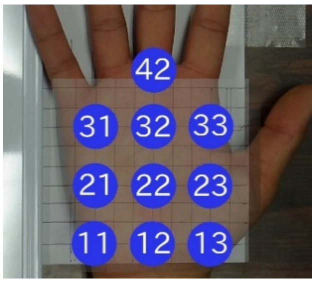


Fig. 7. Reference image for subject

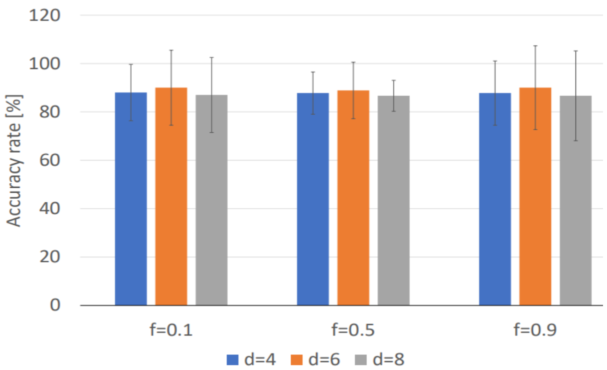


Fig. 8. Experimental result

specification for the device. The objective of the experiment is to find the combination with the highest correct ratio among 27 combinations, where the candidate values of i are 10 mm, 20 mm, and 30 mm, d is 4 mm, 6 mm, and 8 mm, and f is 0.1 kgf, 0.5 kgf, and 0.9 kgf. Fig. 4 shows the experimental scene. Figs. 5 and 6 show the devices for the experiment. Fig. 7 shows the image for the subject to answer the pressed place, which is displayed on the monitor.

2.3. *Experimental Results*

We collected a total of 2,700 data points from 10 subjects, each of whom pressed 10 points. A graph of the results for i = 30 mm is shown in Fig. 8. Note that other results are shown in our previous paper [11]. The vertical axis shows the average ratio of correct responses for each combination, and the error bars show the standard deviation of the ratio of correct responses. In the graph, three bars are grouped together when d is the same. The blue, orange, and gray bars represent the cases where f is 0.1 kgf, 0.5 kgf, and 0.9 kgf, respectively.

The experimental results show that the correct response rate is highest, and the standard deviation is smallest when (i, d, f) is (30 mm, 6 mm, 0.9 kgf). In this case, the ratio of correct answers was 93 % and the standard deviation was 7 %. Therefore, we decided to design the device using this combination since the mean value was highest and the variation was smallest.

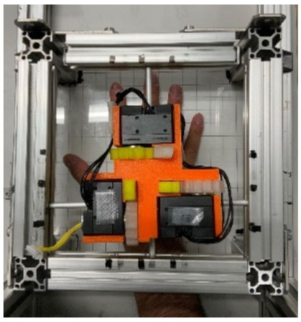


Fig. 9 Device for experiment.(Top view).

(1) Bare hand (2) Thin grove (3) Thick grove

Fig.10 Groves for experiment

3. Experiment with Multiple Stimulation

3.1. *Experimental Method*

In this section, we report a result of an experiment to check whether subjects misidentify the number of points by pressing the back of their hands simultaneously. In the experiment, we prepared three subjects, and asked them to press one or two points on the back of their hands and asked them to answer where the pressed points corresponded to in their palms. We use the combination (i, d, f) = (30 mm, 6 mm, 0.9 kg), which is the required specification for the device.

The subjects were three males (twenties), and in each condition, stimuli were applied to ten points on the back of the hand in a random order. The scene of the subject experiment is like the situation as shown in Fig. 4. We set up the experimental device which can press at two points simultaneously as shown in Fig. 9. The intervals of the points are fixed to 30 mm. In this experiment, the subjects wear gloves to imagine the practical use as shown in Fig. 10. Fig. 11 shows the image for the subject to answer the pressed places.

3.2. *Results and Discussion*

We collected data for a total of 135 points, since 15 points were pressed by three subjects in three conditions.

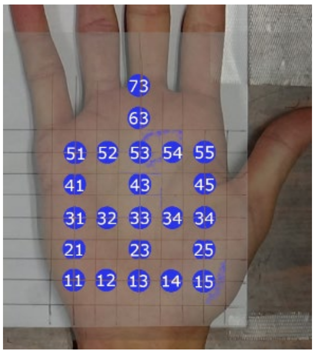


Fig.11 Reference image for subject

We calculated the average ratio of correct answers if the number and position of the points pressed corresponded to the points answered by the subjects. A graph of the results is shown in Fig. 12.

As a result of the experiment, subjects collectively recognized the positions of the pressed points similar to the previous experiment as shown in Fig. 8. Therefore, we think that the device which is developed with the specification obtained in the previous experiment may be effective in the case of multiple stimulations. On the other hand, we think that it is necessary to consider how to stimulate when one large object touches the robot hand. If we want the operator to recognize the tactile sensation as a surface, we may need to set the distance between the stimulus points to less than 30mm, which is the discrimination threshold of the back of the hand.

Moreover, the accuracy rate was decreased when the subjects wore thin gloves. Therefore, it is necessary to pay attention to the selection of the thickness and materials of the gloves. We think that it is interesting that the accuracy rates of some subjects are increased when the subjects wore thick gloves. We will verify this phenomenon with more subjects in the future.

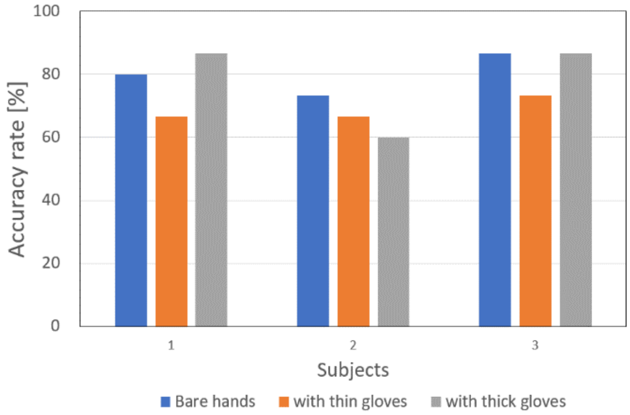


Fig.12 Experiment result

4. Conclusion

In this study, we propose a method to present the contact state of the robot hand to the back of the operator's hand as a substitute for the palm in teleoperation of a disaster response robot. In this paper, we conducted an additional experiment with multiple stimulation.

In the future, we will conduct further verification of the experimental results. After that the device will be developed and the efficiency of teleoperation of the robot hand using the fabricated device will be verified.

References

1. S Tadokoro, et al.: "Rescue Robotics", Journal of the Robotics Society of Japan, Volume 19, Issue 6, Pages 685-688, 2001. (In Japanese)
2. S Tadokoro: "Middle-Long Term Issues of Disaster Robots in Japan", Journal of the Robotics Society of Japan, Volume 32, Issue 2, Pages 154-161, 2014. (In Japanese)
3. S Tadokoro, et al.: "Overview and Outcome of WRS Disaster Robotics Category", Journal of the Robotics Society of Japan, Volume 37, Issue 3, Pages 224-234, 2019. (In Japanese)
4. N Sato, F Matsuno: "User Interface Technology for Rescue Robot Teleoperation", Journal of the Robotics Society of Japan, Volume 28, Issue 2, Pages 156-159, 2010. (In Japanese)
5. K Mima, et al.: "Tele-Manipulation with Humanoid Robot Hand/Arm via Interne", Proceedings of the Japan Joint Automatic Control Conference, Volume 54, 2K303, 2011. (In Japanese)
6. K Negishi, et al.: "i-MarioNET: Remote Operated Humanoid Robot System Controlled by Telexistence FST(Robotics for Hazardous Fields)", The Proceedings of JSME annual Conference on Robotics and Mechatronics (Robomec) 1A1-R14, 2013. (In Japanese)
7. K. Minamizawa, et al. : "A PalmWorn Haptic Display for Bimanual Operations in Virtual Environments", Proceedings of International Conference on Human Haptic Sensing and Touch Enabled Computer Applications, pp 458-463, 2008.
8. T Taniguchi, et al.: "Multi-Point Pressure Sensation Display using Pneumatic Actuators"; Proc. EuroHaptics 2019, LNCS 10894, 58-67, 2018.
9. T Okano, et al.: "Haptic Feedback for Foot Sole Using Pneumatic Pressure Device", Proc.ASIAGRAPH2016, 3-6, 2016.3..
10. Wilska, A : "On the vibrational sensitivity in different regions of the body surface", Acta Physiologica Scandinavica, vol. 31, pp. 285-289, 1954.
11. N Sato, K Ushimaru, Y Yoshifumi : " Haptic Device that Presents Sensation Corresponding to Palm on Back of Hand for Teleoperation of Robot Hand Report 1: Clarification of Required Specifications by Subject Experiment", The 38th Annual Conference of the RSJ, 2I3-03, 2020. (In Japanese)

Robot Assisting Water Serving to Disabilities by Voice Control

Yang Chunxin
MIST, Kyushu Institute of Technology, 680-4 Kawazu
Iizuka-shi, Fukuoka 820-8502, Japan
Sakmongkon Chumkamon
MIST, Kyushu Institute of Technology, 680-4 Kawazu
Iizuka-shi, Fukuoka 820-8502, Japan
Eiji Hayashi
MIST, Hayashi Lab, 680-4 Kawazu
Iizuka-shi, Fukuoka 820-8502, Japan
E-mail: chunxin.yang215@mail.kyutech.jp, m-san@mmcs.mse.kyutech.ac.jp
haya@mse.kyutech.ac.jp
www.kyutech.ac.jp

Abstract

ROS is an open-source robot operating system. In this paper, we use ROS to control Conbe robot arm. By introducing the YOLACT real-time instance segmentation, we trained our own model for Object Detection. Secondly, the Speech-Recognition system is established through Deep speech and Mozilla Text-To-Speech with Tacotron2 DDC model. Deep speech is an end-to-end speech system, where deep learning supersedes these processing stages. Combined with a language model, this approach achieves higher performance than traditional methods on hard speech recognition tasks while also being much simpler. In this way, we create an artificial intelligence, which accomplished a simple conversation with people. And the voice control system is established based on Speech-Recognition system. In the experiment, we successfully control the robot arm move positions and do water serving for disabilities by voice command. With this research, voice control robot arm can be apply in the life support area, it will be more convenient for disabilities in daily life.

Keywords: ROS, Water Serving, Disabilities, YOLACT, Speech-Recognition, voice control, Deepspeec

1 Introduction

1.1 Background

Robot arm has become an important role in industrial production and in people's life using the interplay of robot technology and information technology.

In this research, the robot arm is able to communicate with people accomplished by using Deepspeech and Mozilla Text-To-Speech with Tacotron2 DDC model. Communication with the arm and serving water for those who can't drink water on their own by voice control become possible.

1.2 Purpose of Research

(a) Multipurpose

Voice control robot arms can be applied to home services, commercial services and industrial production.

(b) Can be used in the life support area

The miniaturized robot arm can be used for the human body equipment, and the simple service is possible for the disabled people. By grasping objects using voice control, such as grasping cups to serve water,it will be more convenient for these people in daily life.

2 System Configuration

The system consists of four parts:

(a) Machine Vision

Using YOLACT Real-time Instance Segmentation to train dataset and detect the object that we need.

(b) Speech Recognition

By using Deepspeech and Mozilla Text-To-Speech, we create our speech dataset and trained the speech model, which successfully created an artificial intelligence accomplished a simple conversation with people and control robot arm to move.

(c) ROS And MoveIt

Using ROS and MoveIt to set robot's posture (position and angle). Using Moveit trac-ik to solve inverse-kinematics problem for robot.

(d) Conbe Robot Arm

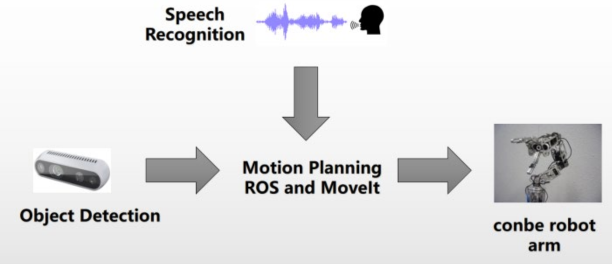


Fig.1 System configuration

3 Machine Vision

3.1 Object Detection

(a) Data collection

In order to detect an object accurately, a large amount of data are necessary. Therefore, we need to collect data by photographing each object and each object needs to take at least 100 pictures from different background, colors, angles and distance.

Then, the object will be labeled with the measurement range and the object name.

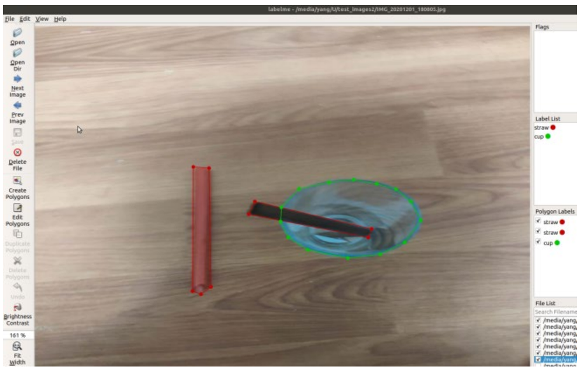


Fig.2 Label The Object

(b) Training with YOLACT

YOLACT is a simple, fully-convolutional model for real-time instance segmentation that achieves 29.8 mAP on MS COCO at 33.5 fps evaluated on a single Titan Xp, which is significantly faster than any previous competitive approach[1].

We trained date using the Resnet101-FPN model, and complete training until the loss is minimized.

(c) Program

The object detection program uses a case program on YOLACT which is eval.py with trained data[2].In order to do water serving, we need to let robot know the position of object, therefore, we add coordinates on the target using opencv centroid, calculate the coordinates of the center point from the coordinates of the bounding box. But in this way we just acquire 2-dimensional coordinates on the target. The real coordinates of the target point are obtained by recognizing the pixel coordinates and depth values of the target.

$$\frac{target\_xy\_pixel[0]-ppx}{fx}=\frac{target\_xy\_true[0]}{target\_depth}$$

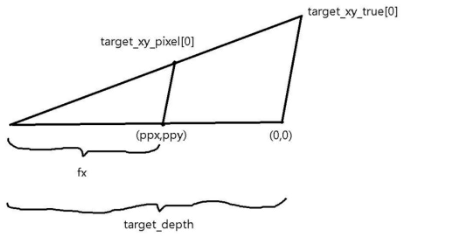


Fig.3 Convert Pixel Coordinates To Depth Value

In the figure3 ,(0,0) is the point position of object target in RGB image, and get the parameter of ppx,ppy,fx,fy from camera internal parameters. In this way, we can acquire 3-dimensional coordinates of the target. Then, we publish markers on this 3-dimensional coordinates using ROS.

3.2 Test

In the test,it can accurately identify the short distance and far distance(2 meter) for cup and straw with 30FPS.The pixel coordinates are displayed in the middle of the object.We can get the real world 3D coordinates based on this pixel coordinates.

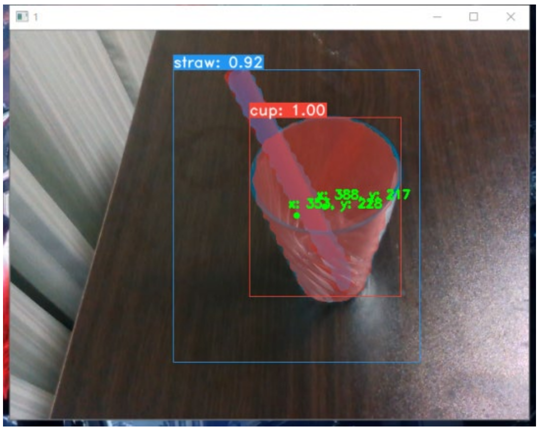


Fig.4 Object Detection Test

4 Speech Recognition

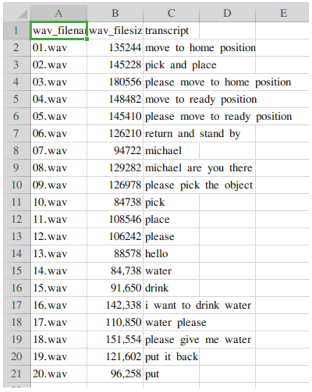
4.1 Deepspeech

Deepspeech is an end-to-end speech system, where deep learning supersedes these processing stages. Combined with a language model, this approach achieves higher performance than traditional methods on hard speech recognition tasks while also being much simpler[3].

We using Deepspeech pre-trained model and add our own database to fine-tune the model, then trained with tensor flow. Currently, we use the Deepspeech 0.9.1 pre-trained model with English alphabet.

(a) Database

For this research, we trained 20 words and sentences that we need. For audio file, we need the audio in Mono 16K with .Wav format[4].



	A	B	C	D	E
1	wav_filena	wav_filesiz	transcript		
2	01.wav	135244	move to home position		
3	02.wav	145228	pick and place		
4	03.wav	180556	please move to home position		
5	04.wav	148482	move to ready position		
6	05.wav	145410	please move to ready position		
7	06.wav	126210	return and stand by		
8	07.wav	94722	michael		
9	08.wav	129282	michael are you there		
10	09.wav	126978	please pick the object		
11	10.wav	84738	pick		
12	11.wav	108546	place		
13	12.wav	106242	please		
14	13.wav	88578	hello		
15	14.wav	84,738	water		
16	15.wav	91,650	drink		
17	16.wav	142,338	i want to drink water		
18	17.wav	110,850	water please		
19	18.wav	151,554	please give me water		
20	19.wav	121,602	put it back		
21	20.wav	96,258	put		

Fig.5 Database of Deepspeech

4.2 Mozilla TTS

Mozilla TTS is a deep learning based Text To Speech project.We use the TTS pre-trained model,Tacotron2 Double Decoder Consistency for our system to get response from our speech.Tacotron2 is a neural network architecture for speech synthesis directly from text and it faster than real-time inference performance[5].

4.3 Speech Recognition Program

To recognize people's language correctly :

(a) The language need to be record and saved as an audio file.

(b) Identify text from the audio files.

(c) Then select the corresponding answer text based on the recognized text.

(d) Convert selected answers to voice.

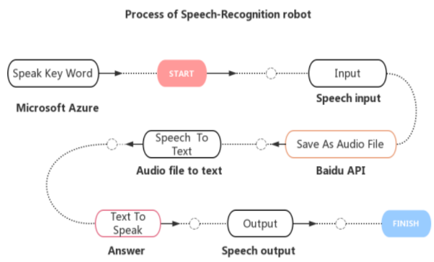


Fig.6 Construction Of Speech Conversation Program

4.4 Speech Recognition Test

In the test, we can have a simple conversation with this AI, it can recognize our words correctly and response based on our words. For example, If we say “pick and place”, it will print the response and answer” right away”.For trained words and sentences, the correct rate is over 90%.

4.5 Voice control

Voice control for robot use this program: “os.system("gnome-terminal -e 'rosrun conbe_moveit_config moveit_fk_demo.py'")”

This program will run the python scripts, in this python file the pose and movement of robot is defined. When we speak the command , it will run the corresponding python scripts.

5 Experiment

5.1 Simulation With MoveIt

Moveit is a Motion Planning Framework for manipulators. The main use is calculate the path to the specified position and angle of the gripper. We use voice control such as saying“please give me water”to call Python scripts and then control robot to serve water.

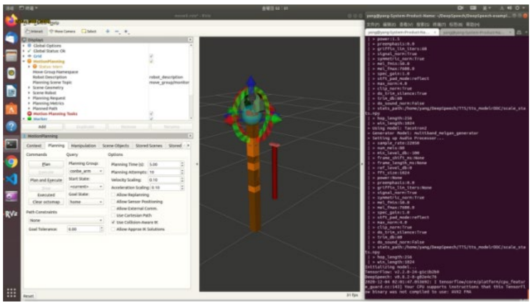


Fig.8 Setting Robot Position Using Voice Control

5.2 Water Serving With Real Robot

The Water Serving program uses MoveIt Command Interface. By setting the joint position and get target position from YOLACT, Moveit will calculate the path and inverse-kinematics for robot planning group. Robot will first pick straw into the cup and then bring cup to people.

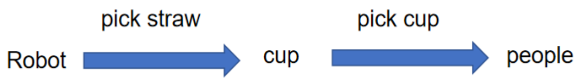


Fig.9 Water Serving

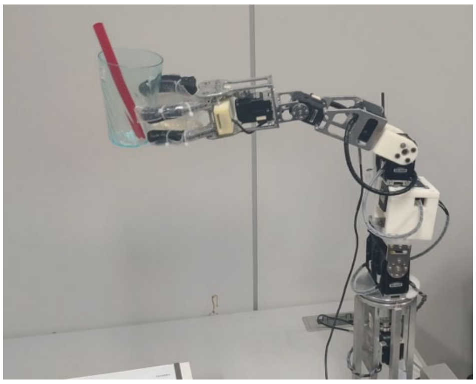


Fig.10 Pick Cup With Real Robot

6 Future Work

In the future, we would like to create more task for the robot, such as making coffee and tea. And we want to use Soft-Actor-Critic Reinforcement Learning method for our robot in order to make robot do pick and place object faster and more accurate.

For speech recognition part, we will create a interface with webserver, and a virtual character, people can communicate with robot through this interface and using voice control easier.

References

1.https://arxiv.org/abs/1904.02689

2.https://github.com/dbolya/yolact

3.https://arxiv.org/abs/1412.5567

4.https://medium.com/visionwizard/train-your-own-speech-recognition-model-in-5-simple-steps-512d5ac348a5

5.https://arxiv.org/abs/1712.05884

6.https://github.com/Eruvae/yolact_ros

A Method of Role Differentiation Using a State Space Filter with a Waveform Changing Parameter in Multi-agent Reinforcement Learning

Masato Nagayoshi, Simon Elderton
Niigata College of Nursing, 240 shinnan-cho
Joetsu, Niigata 943-0147, Japan
E-mail: nagayosi@niigata-cn.ac.jp, elderton@niigata-cn.ac.jp

Hisashi Tamaki
Kobe University, 1-1 Rokkodai-cho, Nada-ku,
Kobe, Hyogo 657-8501, Japan
E-mail: tamaki@al.cs.kobe-u.ac.jp

Abstract

Recently, there have been many studies on the multi-agent reinforcement learning (MARL) in which each autonomous agent obtains its own control rule by RL. Here, it is considered that different agents having individuality is more effective than uniform agents in terms of role differentiation in MARL. Then, we have proposed a promoting method of role differentiation using a waveform changing parameter in MARL. In this paper, we confirm the effectiveness of role differentiation by introducing the waveform changing parameter into a state space filter through computational experiments using "Pursuit Game" as one of multi-agent tasks.

Keywords: reinforcement learning, role differentiation, meta-parameter, waveform changing, state space filter

1. Introduction

Engineers and researchers are paying more attention to reinforcement learning (RL) [1] as a key technique for realizing computational intelligence such as adaptive and autonomous decentralized systems. Recently, there have been many studies on multi-agent reinforcement learning (MARL) in which each autonomous agent obtains its own control rule by RL. Then, we hypothesize that different agents having individuality is more effective than uniform agents in terms of role differentiation in MARL. Here, we define "individuality" in this paper as being able to be externally observed, but not a difference that we are incapable observing, such as a difference of internal construction.

We consider that differences in interpretations of experiences in the early stages of learning have a great effect on the creation of individuality of autonomous agents. In order to produce differences in interpretations of the agents' experiences, we utilized Beck's "Cognitive distortions" [2], which is a cognitive therapy.

Then we have proposed a "fluctuation parameter" which is a wave-form changing meta-parameter in order to realize "Disqualifying the positive" * which is one of the "Cognitive distortions", and a promoting method of role differentiation using the fluctuation parameter in MARL [3].

In this paper, we introduce the "fluctuation parameter" into a state space filter [4] in order to realize "Overgeneralizing" † which is one of the "Cognitive distortions", and confirm the effectiveness of role differentiation by introducing the fluctuation parameter into the state space filter through computational experiments using "Pursuit Game" as one of multi-agent tasks.

2. Q-learning

In this section, we introduce Q-learning (QL) [5] which is one of the most popular RL methods. QL works by calculating the quality of a state-action combination, namely the Q-value, that gives the expected utility of performing a given action in a given state. By performing an action $a \in \boldsymbol{A}_{\mathrm{Q}}$, where $\boldsymbol{A}_{\mathrm{Q}} \subset \boldsymbol{A}$ is the set of available actions in QL and $\boldsymbol{A}$ is the action space of the RL agent, the agent can move from state to state. Each state provides the agent with a reward r. The goal of the agent is to maximize its total reward.

The Q-value is updated according to the following formula, when the agent is provided with the reward:

$$Q\big(s(t-1), a(t-1)\big) \leftarrow Q\big(s(t-1), a(t-1)\big) + \alpha_{\mathrm{Q}}\{r(t-1) + \gamma \max_{b \in \boldsymbol{A}_{\mathrm{Q}}} Q(s(t), b) - Q(s(t-1), a(t-1))\} \quad (2)$$

where $Q\big(s(t-1), a(t-1)\big)$ is the Q-value for the state and the action at the time step $t-1$, $\alpha_{\mathrm{Q}} \in [0,1]$ is the learning rate of QL, $\gamma \in [0,1]$ is the discount factor.

The agent selects an action according to the stochastic policy $\pi(a|s)$, which is based on the Q-value. $\pi(a|s)$ specifies the probabilities of taking each action *a* in each state *s*. Boltzmann selection, which is one of the typical action selection methods, is used in this research. Therefore, the policy $\pi(a|s)$ is calculated as

$$\pi(a|s) = \frac{\exp\,(Q(s,a)/\tau)}{\sum_{b \in A_{\mathrm{Q}}} \exp\,(Q(s,b)/\tau)} \quad (3)$$

where τ is a positive parameter labeled temperature.

3. Reinforcement Learning with a State Space Filter

We have proposed a state space filter based on the entropy which is defined by action selection probability distributions in a state[5].

The entropy of action selection probability distributions using Boltzmann selection in a state *H(s)* is defined by

$$H(s) = -(1/\log|\boldsymbol{A}|)\ \textstyle\sum_{a \in \boldsymbol{A}} \pi(a|s)\log\pi(a|s) \qquad (2)$$

where $\pi(a|s)$ specifies probabilities of taking each action *a* in each state *s*, $\boldsymbol{A}$ is the action space and $|\boldsymbol{A}|$ is the number of available actions.

The state space filter is adjusted by treating this entropy *H(s)* as an index of a correctness of state aggregation in the state *s*. In particular, in case of mapping from the inner state space roughly digitized to the inner state space, a perceptual aliasing problem is happened. That is, the action which an agent should select cannot be identified clearly. Thus, the entropy may not be small in the state space should be divided. In this paper, sufficiency of the number of learning opportunities is judged using a threshold value θ_{L}.

Therefore, if the entropy does not get smaller than a threshold value θ_{H} despite the number of learning opportunities is sufficient, the state space filter is adjusted by dividing the state due to that the perceptual aliasing problem is happened.

Similarly, if the entropy is smaller than θ_{H} in a state *s* and a different state mapping from a transited input state *s'*, and representative actions in each other's states are same, the state space filter is adjusted by integrating the states due to that the states is too divided.

4. Fluctuation Parameter

RL has meta-parameters κ to determine how RL agents learn control rules. The meta-parameters κ include the learning rate α, the discount factor β, ε of ε-greedy which is one of the action selection methods, and the temperature τ of Boltzmann action selection method.

In this paper, the following fluctuation parameter using damped vibration function is introduced into this κ.

$$\kappa(t_{\mathrm{p}}) = \begin{cases} \kappa + A cos(2\pi(t_{\mathrm{p}}/\lambda) + \phi) & (t_{\mathrm{pa}} < t_{\mathrm{ps}}) \\ \kappa + A cos(2\pi(t_{\mathrm{p}}/\lambda) + \phi) \times t_{\mathrm{ps}}/t_{\mathrm{pa}} & (otherwise) \end{cases} \quad (4)$$

where $A, t_{\mathrm{p}}, t_{\mathrm{pa}}, t_{\mathrm{ps}}, \lambda$ and ϕ is the amplitude, the phase, the damped phase, the initial phase of damping, the wavelength, and the initial phase parameter of the fluctuation, respectively. The phase t_{p}, the damped phase t_{pa}, the initial phase of damping t_{ps}, and the wavelength λ are needed to set proper units.

5. Computational Experiments

5.1. *Pursuit Game*

The effectiveness of the proposed approach is investigated in this section. It is applied to the so-called

Table 1. Parameters for Q-learning with a state space filter

Parameter	Value
α_{Q}	0.1
γ	0.9
τ	0.1
θ_{H}	0.3
θ_{L}	1,000

"Pursuit Game" where three RL agents move to capture a randomly moving target object in a discrete 10×10 globular grid space. Two or more agents or an agent and the target object cannot be located at the same cell. At each step, all agents simultaneously take one of the 5 possible actions: moving north, south, east, west or standing still. A target object is captured when all agents are located in cells adjacent to the target object and surrounding the target object in three directions.

The agent has a field of view, and the depth of view set at 3. Therefore, the agent can observe the surrounding $(3 \times 2 + 1)^2 - 1$cells. The agent determines the state by information within the field of view.

The positive reinforcement signal $r_t = 10$ (reward) is given to all agents only when the target object is captured, and the positive reinforcement signal $r_t = 1$ (sub reward) is given to the agent only when the agent is located in the cell adjacent to the target object and the reinforcement signal $r_t = 0$ at any other steps. The period from when all agents and the target object are randomly located at the start point to when the target object is captured and all agents are given a reward, or when 100, 000 steps have passed is labeled 1 episode. The period is then repeated.

5.2. *RL Agents*

All agents observe the only target object in order to confirm the effectiveness of role differentiation, e.g. moving east of the target object. Therefore, the state space is constructed with a 1 dimensional space.

Computational experiments have been done with parameters as shown in Table 1. In addition, all initial Q-values are set at 5.0 as the optimistic initial values, and θ_{H} was set referring to about 0.288: the maximal value of the entropy when the highest selection probability for one action is 0.9,

5.3. *Experiment (A): Learning Rate*

The effectiveness of role differentiation by introducing 3 the fluctuation parameters, in which the initial phase $\phi = 0$, the amplitude $A = 0.09$, and the wavelength $\lambda = \{50, 100, 500\}$ [episode] , into the learning rate of QL with the state space filter (hereafter called "50", "100", and "500", respectively) are investigated in comparison with an ordinary QL with the state space filter without fluctuation parameter (hereafter called "constant"). Here, the fluctuation parameters of all agents take the same value. The unit of the phase $t_{\mathrm{p}}$is set [episode] which is the same as the wavelength λ, the unit of the damped phase $t_{\mathrm{pa}}$is set [episode], and the initial phase of damping is set at $t_{\mathrm{ps}} = 250$ [episode]. The range of values which the fluctuation parameter for $\alpha_{\mathrm{Q}} = 0.1$ can take e.g. [0.01, 0.19] on the condition of $A = 0.09$.

The average numbers of steps and the average size of the state space required to capture the target object were observed during learning over 20 simulations with various wavelength parameters in the learning rate, as described in Figs. 3 and 4, respectively.

It can be seen from Figs. 3, 4 that, (1) "50", "100" and "500" show a better performance than "constant" with regard to the obtained control rule, (2) "50", "100" and "500" are smaller than "constant" with regard to the size of the state space.

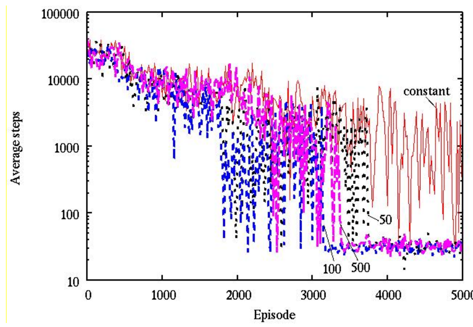


Fig. 3. Required steps of various wavelength parameters in the learning rate (ϕ = 0 [rad]).

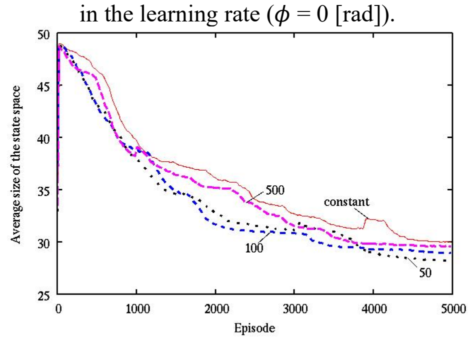


Fig. 4. Required size of the state space filter of various wavelength parameters in the learning rate (ϕ = 0 [rad]).

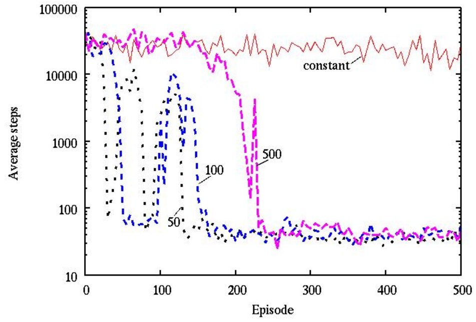


Fig. 5. Required steps of various wavelength parameters in the temperature (ϕ = 0 [rad]).

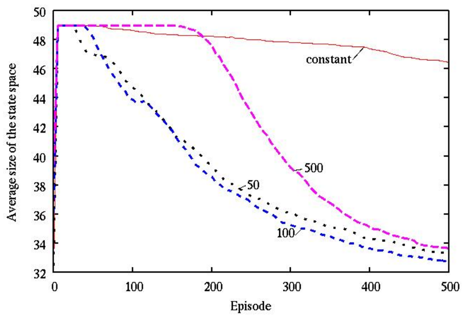


Fig. 6. Required size of the state space filter of various wavelength parameters in the temperature (ϕ = 0 [rad]).

5.4. *Example (B): Temperature*

The effectiveness of role differentiation by introducing 3 the fluctuation parameters, in which the initial phase $\phi = 0$, the amplitude $A = 0.09$, and the wavelength $\lambda = \{50, 100, 500\}$ [episode] , into the temperature of QL with the state space filter (hereafter called "50", "100", and "500", respectively) are investigated in comparison with an ordinary QL with the state space fliter without fluctuation parameter (hereafter called "constant"). Here, the fluctuation parameters of all agents take the same value. The unit of the phase $t_{\mathrm{p}}$is set [episode] which is the same as the wavelength λ, the unit of the damped phase $t_{\mathrm{pa}}$is set [episode], and the initial phase of damping is set at $t_{\mathrm{ps}} = 250$ [episode]. The range of values which the fluctuation parameter for $\tau = 0.1$ can take e.g. [0.01, 0.19] on the condition of $A = 0.09$. If the temperature is zero, then action selection of the agent is greedy and the situations where agents cannot capture the target object occur. Therefore, A is set at 0.09

The average numbers of steps and the average size of the state space required to capture the target object were observed during learning over 20 simulations with various wavelength parameters in the temperature, as described in Figs. 5 and 6, respectively.

It can be seen from Figs. 5, 6 that, (1) "50", "100" and "500" show a better performance than "constant" with regard to the obtained control rule, (2) "50", "100" and "500" are smaller than "constant" with regard to the size of the state space.

Thus, it could be confirmed that the effectiveness of role differentiation by introducing the fluctuation parameter into the state space filter. It could be considered that this is the result of overgeneralizing.

6. Conclusion

In this paper, we introduced a "fluctuation parameter" into a state space filter in order to realize "Overgeneralizing" which is one of the "Cognitive distortions". Through computational experiments, we confirmed the effectiveness of role differentiation by introducing the fluctuation parameter into the state space filter. It could be considered that this is the result of overgeneralizing.

Our future projects include to apply real world problems, e.g. improving a schedule for each nurse.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP19K04906.

References

1. R. S. Sutton and A. G. Barto, Reinforcement Learning (A Bradford Book, MIT Press, Cambridge, 1998).
2. A. T. Beck, Cognitive Therapy and the Emotional Disorders (International University Press, New York, 1976).
3. M. Nagayoshi, S. J. H. Elderton and H. Tamaki, A Promoting Method of Role Differentiation using a Learning Rate that has a Periodically Negative Value in Multi-agent Reinforcement Learning, *Journal of Robotics, Networking and Artificial Life*, 6(4), 2020, pp.221–224.
4. M. Nagayoshi, H. Murao and H. Tamaki, A State Space Filter for Reinforcement Learning, *Proc. of AROB 11th'06,* 2006, pp.615-618.
5. C. J. C. H. Watkins and P. Dayan, Technical note: Q-Learning, *Machine Learning* **8** (1992), 279-292.

The research about editing system of performance information for player piano
- Make inferences about both handed musical composition by using DP matching system -

Ryo Kinoshita
Department of Intelligent and Control Systems, Kyushu Institute of Technology,
680-4, kawazu, Iizuka-City, Fukuoka, 820-8502, Japan[*]

Eiji Hayashi
Department of Intelligent and Control Systems, Kyushu Institute of Technology, Hayashi Lab,
680-4, kawazu, Iizuka-City, Fukuoka, 820-8502, Japan
E-mail: kinoshita@mmcs.mse.kyutech.jp, haya@mse.kyutech.ac.jp
http://www.kyutech.ac.jp/

Abstract

We have developed a system that allows a piano to perform automatically. In order to play music in the manner of a live pianist, we must add expression to the piano's performance. Therefore, we have developed an interactive musical editing system that utilizes a database to edit music more efficiently.

Keywords: Automatic Piano, Knowledge Database, Computer Music, Music Interface

1. Introduction

We have developed a system that allows a piano to perform automatically. In this system, 90 actuators have been installed on the keys and pedals of a grand piano. These actuators execute key strokes and pedal movements to govern the piano's performance, e.g., "Fig. 1.(*Continued*)" [1,2].

Fig. 1. The automatic piano

Our research focuses on giving player piano performance with human-like skill and expression. Playing the piano expressive by player piano, it is necessary to adjust the volume, length, and timing of music. We aimed to develop a system that, like a skilled pianist, can perform even the first musical score based on information related to previous skills and experience. So, we developed a system that automatically estimates the performance expression of unedited music using edited performance data and score data. In this paper, we described a phrase search using Dynamic Programming (DP) matching and a method for selecting an optimal phrase, how to infer parameters of notes, evaluation of an inferred entire song played by both-handed.

2. Editing Support System

2.1. *Performance Information*

The automatic piano that we have developed uses a music data structure that is similar to Musical Instrument Digital Interface (MIDI). This system edits four parameters involved in producing a tone: "Velo" (velocity), "Gate", "Step" and "Time". "Velo" is the dynamics, given by the value of 1-127. "Gate" is the duration of the note in milliseconds. "Step" is the interval of time between notes, and it also exhibits tempo. "Time" is the time since the sound started.

2.2. *How to Make the Data for Player Piano*

We show the structure of the edit system how to make the data for player piano in "fig. 2. (*Continued*)". The system extract features of a pianist from the music data that the pianist has played. Next, the system infers performance data from the features and score information on a music that the pianist has not played.

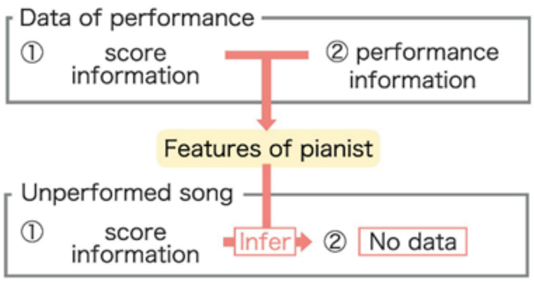


Fig. 2. The structure of the edit system

2.3. *Search System*

As a result of the analysis, it was found that phrases of the same pattern existing in the same tune are performed in a similar expression (Fig. 3.). This time, we use DP matching to search for similar phrases for arbitrarily determined search phrase.

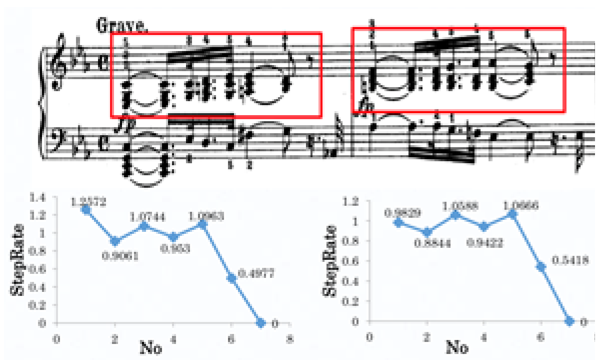


Fig. 3. The discovery about the relation of same tunes and similar expression.

DP matching has a feature that can calculate the similarity between two words that are different in a number of characters from each other.

2.4. *Select System*

After done the Search system, there are a lot of similar phrases which has the same DP matching points. Similar phrases are phrases we found by doing DP matching from music data. Objective phrase is phrase we want to inference in this time. Then, we need to select the optimal phrases for objective phrases by doing a select system. In this system, five indicators were used from the viewpoint of music theory: Dynamic symbols, Beats and Steps, Similarity in change of interval, Dynamic symbols and Velo, and Musical forms.

2.4.1. Dynamic symbols

If the Dynamic symbols are different between similar phrase and objective phrase, the performance is affected even if the phrase is the same. Then, the search phrase and the similar phrase that matches the dynamics on the score are preferentially selected.

2.4.2. Beat and step

Based on the musical grammar, it is known that strong beats are closely related to Step [3]. In places considered to be strong beats, similar phrases are selected using the property that the Step value is larger than in other places. The position of the strong beat depends on the rhythms.

2.4.3. Similarity in change of interval

If the phrases have similar change of interval, it is assumed that the performance expression is also similar. Therefore, we select a phrase with more similarity in change of interval.

2.4.4. Dynamic symbols and Velo

There is likely to be a trend in Velo value depending on dynamic symbols. Therefore, we examine the range of Velo for each dynamic symbol from performance data, and select a phrase which Velo value is correspond to search phrase's dynamic symbol.

2.4.5. Musical forms

If the Select system in 2.4.1.-2.4.4. does not narrow down the number of similar phrases, the selection is performed according to the Musical Forms. Musical Forms is a music format such as the Rond Forms or the Sonata Forms. For example, if different songs were in the sonata format, it was predicted that the exposition part had a presentation part feature, the development part had a development part feature, and each part had a unique feature. We decide same part has similar performance.

2.4.5. Case of Left Hand

Left hand phrases can divide four categories: Main theme, Broken chord, Single, Chord. When select left hand's phrase, select from same category. If similar phrase does not exist in same category, the system selects optimal phrase from other categories.

2.5. *Infer System*

If we use the optimal phrase's expression to search phrase, it will be unnatural expression. Hence, this system infers appropriate Step, Velo, Gate value for the search phrase from the optimal phrase.

2.5.1. Investigation Required for Inference

For appropriate inference, it need investigation from performance data. This year, we use four songs, and its title is shown in "Table 1". All of them are made by W.A.Mozart and Pianist is Maria Joao Pires.

Table 1. The studied title of musical compositions.

	Title of musical compositions.
①	Piano Sonata No.11 in A major, K.311 3rd movement "Turkish March"
②	Piano Sonata No.15 in C major, K.545 1st movement "Allegro"
③	Piano Sonata No.15 in C major, K.545 2nd movement "Andante"
④	Piano Sonata No.15 in C major, K.545 3rd movement "Rondo"

2.5.2. Inference of Step

The investigation of tempo and normalization factor are shown in "Table 2" by using music data in "Table 1". Normalization Factor is s number that "All Tempo Average" divide "Tempo Average".

Table 2. Investigation about tempo and Normalization Factor

Music Number	①	②	③	④
Tempo Avg.	0.80	0.82	1.00	0.92
All Tempo Avg.	0.89			
Normalization Factor	1.11	1.07	0.89	0.96

The inference equation for Step is "Eq. (1)". "PStep" represent provisional Step value which calculate by optimal phrase and "NF" represent normalization factor.

$$\mathrm{Step} = \mathrm{PStep} \times \mathrm{NF} \times \mathrm{All\ Tempo\ Avg.} \qquad (1)$$

In the case of left-hand, adjust the timing using the results of right-had inference when the note is same timing in musical score.

2.5.2. Inference of Velo

The investigation of the ranges and average value about Velo are shown in "Table 3" by using music data in "Table 1".

Table 3. Velo's average value for each Dynamic symbol

Dynamic symbol	Avg. (right-hand)	Avg. (left-hand)
p	64	44
mf	62	36
f	64	45

When the system infers Velo's value, it uses similar interval phrase. Similar interval phrase is the phrase with the highest similarity of interval changes among similar phrases.

The inference equations for Velo are "Eq. (2)" and "Eq. (3)", "Eq. (4)". "n" represents the number of notes in objective phrase and "SVelo" represents similar interval phrase's value of Velo. Eq. (2) is used when the first note in the phrase and the dynamic symbol is different from the previous note. Depending on the results of DP matching, the search phrase may have to be split. In this case, the first note of split phrase is calculated by Eq. (3) except case of Eq (2). Other notes are calculated by Eq (4).

$$\mathrm{Velo(n)} = \mathrm{average\ in\ dynamics\ symbol} \qquad (2)$$
$$\mathrm{Velo(n)} = \mathrm{Velo(n-1)} \qquad (3)$$
$$\mathrm{Velo(n)} = \mathrm{Velo(n-1)} - \mathrm{SVelo's\ value} \qquad (4)$$

2.5.2. Inference of Gate

The investigation of factors for each musical symbol that gives a change in the length of a note are shown in "Table 4" by using music data in "Table 1".

Table 4. Factor for each musical symbol

Musical symbols	staccato	No symbol	slur
Factor	0.4	0.9	1

The inference equation for Gate is "Eq. (5)".

$$\text{Gate} = \text{inferred Step} \times \text{Factor} \quad (5)$$

3. Inference Experiment

An experiment was conducted to compare the music reproduced using the editing support system with the performance of the pianist. The target song is Mozart Piano Sonata No.11 in A major, K.311 1st movement "Theme". Some of the inference result are shown as graphs. "Fig. 4" is about right-handed and "Fig. 5"is left-handed Step. "Fig. 6" is about right-handed Gate. "Fig. 7" is about right-handed Velo.

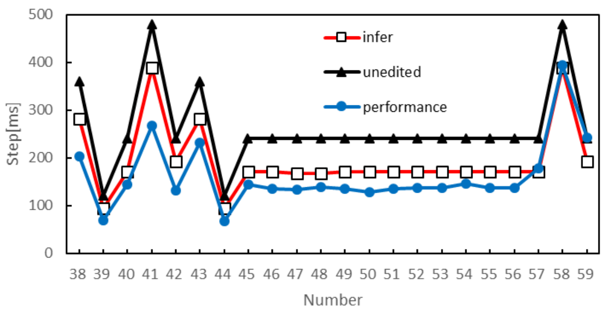


Fig. 4 Inference result about right-handed Step

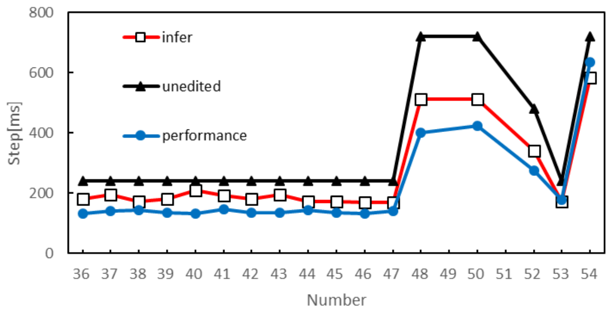


Fig. 5. Inference result about left-handed Step

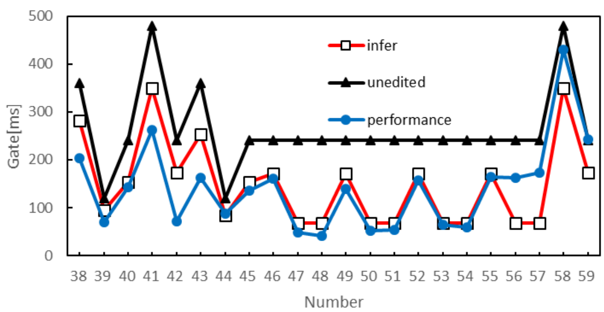


Fig. 6. Inference result about right-handed Gate

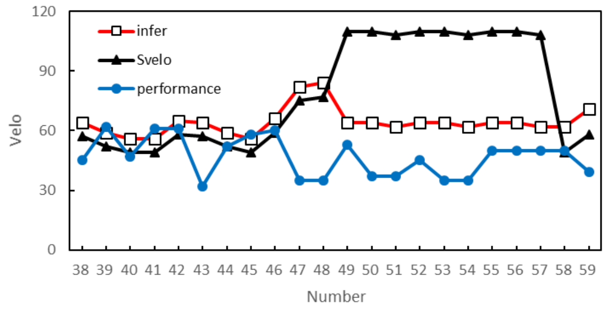


Fig. 7. Inference result about right-handed Velo

4. Consideration

We can confirm that inferred Step value is closer to the pianist's performance than unedited data. It is common to both hands and the same could be seen in Gate. Thus, the results suggest that infer system about Step and Gate are effective. However, inferred Velo's value is not similar to performance data. Therefore, infer system about Velo is need to improve.

5. Conclusion

This year, we introduced a left-handed inference system and Gate's inference system in the editing system of performance information. The inference experiment showed that we were able to infer a performance expression in an unperformed song from data of performance.

Editing system of performance information is infer a phrase from similar phrase in other songs. And, when the same phrase is repeated, the infer result is exactly same. However, pianists should subtly change in inflection and timing for the same phrase. Furthermore, the current system can't infer performance information when the score doesn't have dynamic symbols. In order to solve these problems, we need to develop a new versatile system that adds other methods to current system.

References

1. Y. Isomichi and N. Ogawa, Pattern Matching by Using Dynamic Proganing, 1975, Vol 16, No.1, pp.15-22.
2. 2. E.Hayashi, M.Yamane and H.Mori, Development of Moving Coil Actuator for an Automatic Piano, Inrwenational Journal of Japan Society for Precision Engineering, Vol.28 No.2, 1994, pp.164-169.
3. Shimaoka Jo, Analize about Harmony and Musical style, ONGAKU NO TOMO SHA CORP, 1964.

Development of LiDAR Based Navigation System for Automation of Tree Harvesting Process

Ayumu Tominaga*
Graduate School of Computer Science and System Engineering, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, 820-0044, Fukuoka, Japan†

Akihiro Koubara
Graduate School of Computer Science and System Engineering, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, 820-0044, Fukuoka, Japan

Ryusuke Fujisawa
Graduate School of Computer Science and System Engineering, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, 820-0044, Fukuoka, Japan

Eiji Hayashi
Graduate School of Computer Science and System Engineering, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, 820-0044, Fukuoka, Japan

Abbe Mowshowitz
Iizuka, 820-0044, Fukuoka, Japan

E-mail: tominaga@mmcs.mse.kyutech.ac.jp, koubara.akihiro880@mail.kyutech.ac.jp, fujisawa@ces.kyutech.ac.jp, haya@mse.kyutech.ac.jp

Abstract

This study focuses on an autonomous moving system for the automation of the harvesting process by high-performance machines in the forestry. Many fatal accidents occur due to the harvesting process. In this research, a navigation system has been developed to enable autonomous travel between accumulation sites and trees to be harvested to improve productivity and safety. A 3D map is generated by LiDAR observation, and harvester moves autonomously towards the tree as specified by the operator. A test of the harvesting process was performed in an experimental environment. The evaluation focused on the required time of the autonomous movement in the process. The effectiveness of the system was confirmed in operations such as row thinning by the results.

Keywords: Field Robot, Forest Industry, Harvesting, Autonomous Moving

1. Introduction

Forestry is an industry that nurtures and economically utilizes forest. The safety of forestry work has become an issue, and the development and proliferation of high-performance forestry machinery has been increasing. In the future, it is necessary to introduce robotics technology to further improve safety and reduce labor burdens by utilizing the mobility applied to the forest environment that has been developed so far.

In this study, we developed a navigation system in the forest environment that is aimed to be operational for the harvesting process. The management style of Japanese forestry is wide-area, small-scale afforestation.

.

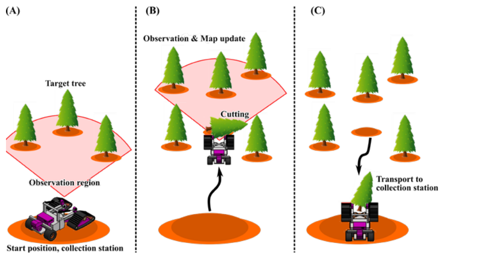


Fig. 1. The scenario of automatic harvesting

Conventional manned machines require long-term, and continuously tree felling and transporting of felled tree, which places a large burden on the operator and leads to serious accidents. The proposed system would navigate that to approach to the target tree selected by the operator autonomously. This navigation was constructed using LiDAR and LiDAR based SLAM which could generate the 3D map and self-localization. We demonstrated in the experimental field and verification of the navigation system.

2. Tree harvesting scenario

Fig. 1 shows the scenario of harvesting process assumed in this study. The details of the scenario are as follows.

1) Selection of a target trees to be felled by the operator (Fig. 1(A)).
2) Approaching the tree by autonomous movement (Fig. 1(B)).
3) Felling of the tree (Fig. 1(B)).
4) Skidding of the tree by autonomous movement (Fig. 1(C)).

In this scenario, the approach and skidding movements are automated by the autonomous moving system, the preparation for the felling and the execution of the sawing are subject to the operator’s operation.

Considering a situation where *N* trees are lined up in a straight line, and the machine fells and skids them from the vicinity of the roadside in order. T_{An} [sec] is a time it takes to approach the n-th tree, T_{Sn} [sec] is a time it takes to skid the felled tree to the roadside, define the time T_n [sec] for autonomous movement as $T_n = T_{An} + T_{Sn}$. Let L [m] be the distance of neighborhood tree form the roadside, l [m] be the interval of trees, and v [m/s] is the average movement velocity of the machine. T_{An} and T_{Sn} is approximately equal and can be estimated by Eq. (1).

$$T_{An} = T_{Sn} = \left\{\frac{L + (n-1)l}{v}\right\} \tag{1}$$

The sum of the time to prepare for felling and to execute of felling is defined as T_F [sec], the sum of the time to select the target tree, it takes for the felled tree to be lowered to an unobstructed position and for the operator to carried-out the operation in τ [sec]. The total time T_N [sec] required to complete the harvesting of the whole trees can be estimated by Eq. (2). Here, T_F and τ are assumed to be the equal for all trees.

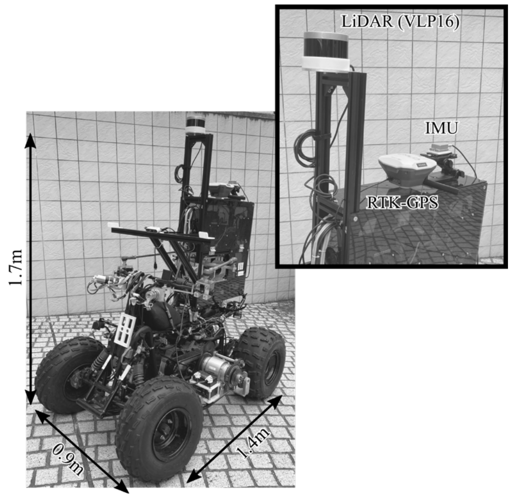


Fig. 2. Forestry robot "SOMA"

$$\begin{aligned} T_N &= \sum_{n=1}^{N} Tn + NT_F + N\tau \\ &= 2\sum_{n=1}^{N}\left\{\frac{L + (n-1)l}{v}\right\} + N(T_F + \tau) \end{aligned} \tag{2}$$

The first term in Eq. (2) is the total time involved in moving, with the progression of the work, the ratio to the total work time $\sum_{n=1}^{N} T_n / T_N$ will increase.

3. System structure

In this study, we developed autonomous ground vehicle “SOMA” shown in Fig. 2. SOMA is based on an All-Terrain Vehicle (ATV) that is robust designed to movement on rough terrain. Fig. 3 shows the system structure that also includes the operation station. The surrounding trees are observed using LiDAR (Velodyne, VLP-16). The robot pose is measured by IMU (Xsense, MTi-30). In this system, the robot was navigated by LiDAR based SLAM which combined IMU’s output.

3.1. *HDL Graph SLAM*

The HDL Graph SLAM was developed for the generation of 3D map and self-localization using

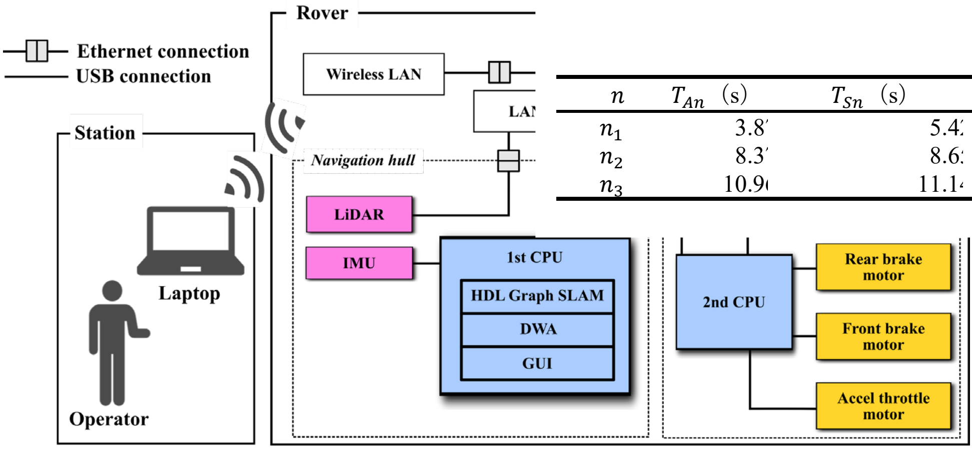


Fig. 4. System structure

Table 1. Experimental

n	T_{An} (s)	T_{Sn} (s)	$\overline{T_F}$ (s)	τ_n (s)
n_1	3.8	5.4		70
n_2	8.3	8.6	20.2	43
n_3	10.9	11.1		28

LiDAR[1]. It is the SLAM that integrates NDT scan-matching[2] using point clouds obtained from LiDAR and the graph optimization.

3.2. *DWA*

The motion plans to approach the target tree is performed by Dynamic Window Approach (DWA) algorithms[3]. The operator could select the target tree of harvesting, the position (x, y) of it's the point will be global target position. The control input (liner velocity, and angular velocity) would be determined by DWA.

4. Verification experiment

We demonstrated the automatic harvesting process to verify the effectiveness of the proposed system in the experimental field. Fig. 4 shows the setting of the initial situation of experiment. The 9 imitation trees were placement equally interval. Approaching and skidding performed for the second row. In this experiment, the robot approached the target autonomously and transitioned to the waiting state for the felling command. After an appropriate amount of time, the robot returned to its initial position by issuing a discharge command. Table 1. Experimental Results the time spent on each element work. 250.8 seconds to complete the process of felling in three targets. The ratio of autonomous moving time to total work time was 9.3, 14.6 and 19.2% as the work progressed. This suggests that this system is effective in operations such as row thinning, where the trees are felled and skidded in sequence. It was confirmed that the round-trip movement was achieved as expected, and the ratio of autonomous moving time in working time increased.

5. Conclusions

In this study, the navigation system based on LiDAR and SLAM was developed to reduce the burden of workers in the harvesting process. The future work includes evaluation of the approaching accuracy and motion planning to achieve a stable and accurate approach and return to the roadside.

References

1. K. Koide, J. Miura, and E. Menegatti, A portable three-dimensional LIDAR-based system for long-term and wide-area people behavior measurement, Int. J. Adv. Robot. Syst., 16(2), 2019.
2. P. Biber and W. Strasser, The normal distributions transform: a new approach to laser scan matching, Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003), 2013, 3, pp. 2743–2748.
3. D. Fox, W. Burgard, and S. Thrun, The dynamic window approach to collision avoidance, IEEE Robot. Autom. Mag., 4(1), 1997, pp. 23–33.

Dynamic Characteristics Analysis of a Multi-scroll Conservative Chaotic System with Sinusoidal Nonlinearity

Zhonggao Chen, Minghan Song

Department of Automation, Tianjin University of Science and Technology,

1038 Dagunanlu Road, Hexi District, Tianjin 300222, PR China

*E-mail: * 313263747@qq.com*

Abstract

In this paper, a multi-scroll conservative chaotic system with sinusoidal nonlinearity is studied. Based on theoretical analysis and numerical analysis, such as equilibrium point and its stability analysis, Lyapunov exponent spectrum and bifurcation diagram, the system is found to show some complex dynamics. In addition, NIST test also verify that the pseudo randomness of the proposed system is satisfactory.

Keywords: multi-scroll, conservative, NIST test

1. Introduction

In 1993, Suykens[1] first proposed a method to generate n-double scroll attractors by adding breakpoints. Compared with single scroll chaotic system or double scroll chaotic system, multi-scroll chaotic system has more control parameters and corresponding key parameters. In addition, it can present complex multi-directional grid like vortices in phase space. The number and shape of vortices can also be controlled and adjusted by the parameters of the system. It has more complex dynamic behavior and is more conducive to image encryption. Therefore, multi-scroll chaotic systems have attracted many attentions, and different types of multi-scroll chaotic systems have been reported, such as unidirectional (1D) multi-scroll chaotic systems[2,3], bidirectional (2D) multi-scroll chaotic systems[4-8], three-dimensional (3D) multi scroll chaotic systems[9], and multi-directional multi-scroll chaotic systems[10-12]. In addition, the realization of chaotic system based on FPGA is also studied[13,14]. In three-dimensional differential equations, the general jerk system has fewer terms, so chaotic systems based on general jerk system have been widely studied[15-19]. In practical application, the increase of vortex number in chaotic system is realized by increasing the number of equilibrium points. Due to the periodicity of sinusoidal functions, many researchers design multi-scroll chaotic systems by modifying the expression of sinusoidal functions or adding control functions to chaotic systems with sinusoidal functions[15,16,18,20-22].

However, there are few studies on the multi-scroll attractors of conservative chaotic systems. In this paper, a new conservative chaotic system is constructed by using sinusoidal function. Through Lyapunov exponent diagram, bifurcation diagram and phase diagram, it is found that there are different attractors with different scroll numbers in the case of different initial values.

2. Construction of Four-dimension Conservative System

Firstly, a new four-dimension conservative system is constructed. It can be expressed as:

$$\dot{\mathbf{x}} = J(\mathbf{x})\nabla H(\mathbf{x}) \begin{bmatrix} 0 & a & 0 & by \\ -a & 0 & 0 & 0 \\ 0 & 0 & 0 & c \\ -by & 0 & -c & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \tag{1}$$

Where, $J(\mathbf{x}) = \begin{bmatrix} 0 & a & 0 & by \\ -a & 0 & 0 & 0 \\ 0 & 0 & 0 & c \\ -by & 0 & -c & 0 \end{bmatrix}$, $\nabla H(\mathbf{x}) = \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$.

Secondly, set $(a, b, c) = (4, 4, 4)$, equation (1) can be described as:

$$\begin{cases} \dot{x} = 4y + 4yw \\ \dot{y} = -4x \\ \dot{z} = 4w \\ \dot{w} = -4xy - 4z \end{cases} \tag{2}$$

Where x, y, z and w is the state variable, and a, b, c is system parameter, respectively.

2.1. *Conservative characteristics of system (2)*

The divergence of the system (2) is

$$\nabla f = \frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{y}}{\partial y} + \frac{\partial \dot{z}}{\partial z} + \frac{\partial \dot{w}}{\partial w} = 0 \tag{3}$$

Moreover, the time domain derivative of the energy function is

$$\dot{H} = \nabla H(\mathbf{x})^T J(\mathbf{x}) \nabla \mathrm{H}(\mathbf{x}) = 0 \tag{4}$$

According to the above analysis, the system satisfies both Hamiltonian energy conservation and volume conservation.

2.2. *Equilibrium point of the system (2)*

The equilibrium point of the system (2) is obtained as $(0, 0, 0, 0)$, and the corresponding eigenvalue is $(6i, -6i, 6i, -6i)$. Set $\nabla H = (\sin(x), y, z, w)$, an infinite number of equilibrium points are obtained. And equation (2) can be express as:

$$\begin{cases} \dot{x} = 4y + 4yw \\ \dot{y} = -4\sin(x) \\ \dot{z} = 4w \\ \dot{w} = -4sin(x)y - 4z \end{cases} \tag{5}$$

It can be found that the equilibrium point of the system (5) is $(k\pi, 0, 0, 0), k \in Z^*$. When the equilibrium point meets $(2k\pi, 0, 0, 0)$, the corresponding eigenvalue is $(4i, -4i, 4i, -4i)$, and the equilibrium point type is a center point; when the equilibrium point is $((2k-1)\pi, 0, 0, 0)$, the corresponding eigenvalue is $(4, -4, 4i, -4i)$, and the type of the equilibrium point is an unstable saddle point.

3. The Dynamic Characters of the System (5)

In this part, the dynamic characteristics of the system (5) are analyzed through Lyapunov exponent spectrum and phase diagram .

3.1. *Lyapunov exponent spectrum of the system (5)*

Set the initial value $(x, y, z, w) = (x(0), 1, 1, 1)$, Lyapunov exponent diagram of the system (5) with the change of the initial $x(0)$ is shown in Fig. 1.

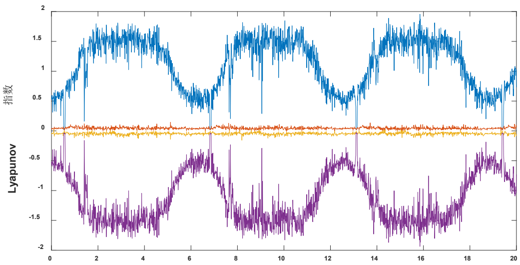


Fig. 1. Lyapunov exponent diagram

It can be seen from Fig. 1 that the Lyapunov exponent of the system (5) is symmetric about the x-axis, and the maximum exponent is always greater than 0, so the system is always in a chaotic state. Secondly, the Lyapunov exponent of the system shows a periodic distribution of 2π.

3.2. *Multi-stability of the system (5)*

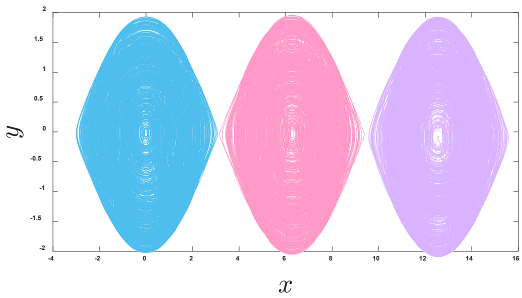
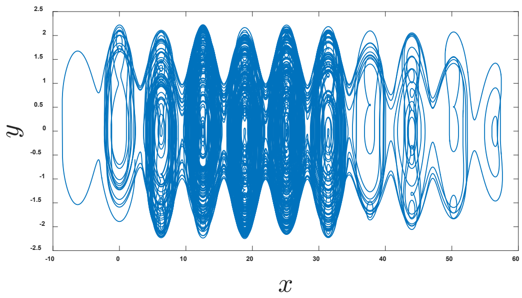
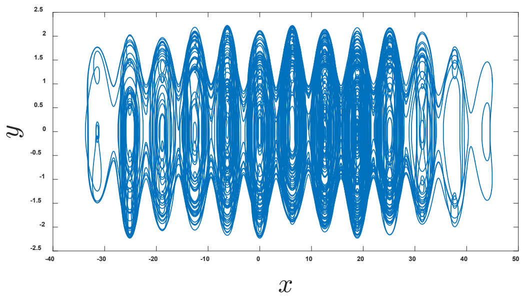


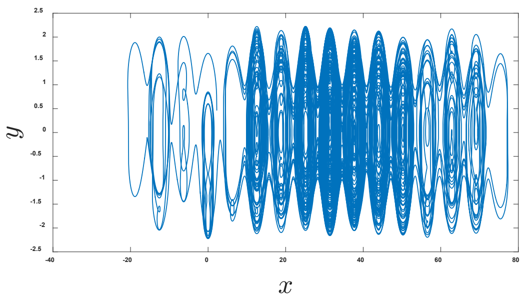
Fig. 2. Coexistence of three independent scrolls



(a) 11-scrolls



(b) 13-scrolls



(c) 16-scrolls

Fig. 2. Different numbers of multi-scroll flows with the change of initial value

It can be found that different flows of the system (5) will appear with the change of initial value, which are shown in Fig. 2 and Fig. 3. When the initial value is $(\frac{\pi}{3}+2k\pi,1,1,1)(k=0,1,2)$, there are three independent scroll flows in the system (5), which are represented by blue, pink and purple lines in Fig. 2. When the initial value is $(\frac{\pi}{2}+2k\pi,1,1,1)(k=0,1,2)$, there are 11-scrolls flow, 13-scrolls flow and 16-scrolls flow in the system (5), which are represented by blue, pink and purple lines in Fig. 2. According to the above analysis, it can be seen that the coexistence of different multi-scrolls flows with the change of different initial values illustrates the multi-stability of the system (5).

3.3. Grid type multi-scroll flows

In the above, multi-scroll flows are obtained in the system (5) in the x-axis direction by expanding the equilibrium point in the x-axis direction. Similarly, the construction of the grid type multi-scroll flows needs to expand the equilibrium point in the x-axis and Y-axis directions. Set $\nabla H=(f(x),\ f(y),z,w)$, and equation (2) can be express as:

$$\begin{cases}\dot{x}=4f(y)+4f(y)w\\ \dot{y}=-4f(x)\\ \dot{z}=4w\\ \dot{w}=-4f(x)f(y)-4z\end{cases} \tag{6}$$

Where $f(x)$ and $f(y)$ satisfy mapping:

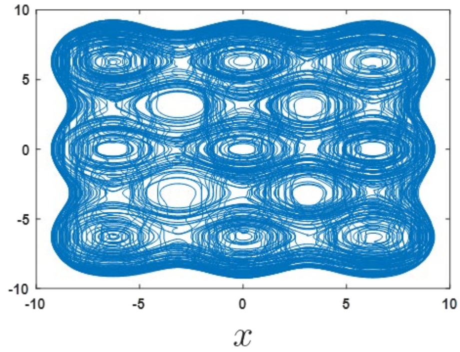
$$f(u)=\begin{cases}u+N, & u<-N\\ \sin(u), -N\le u\le N\ , N=2n\pi, n\in Z^{*}\\ u-N, & u>N\end{cases} \tag{7}$$

It can be found that the equilibrium point of the system is $(k\pi,k\pi,0,0), k\in Z^{*}$. When the equilibrium point is $(2k\pi,2k\pi,0,0)$ and $((2k-1)\pi,(2k-1)\pi,0,0)$, the corresponding eigenvalue is $(4i,-4i,4i,-4i)$ and the equilibrium point type is a center point. When the equilibrium point is $((2k-1)\pi,2k\pi,0,0)$ and $(2k\pi,((2k-1)\pi,0,0)$, the corresponding eigenvalue is $(4,-4,4i,-4i)$ and the type of the equilibrium point is an unstable saddle point.

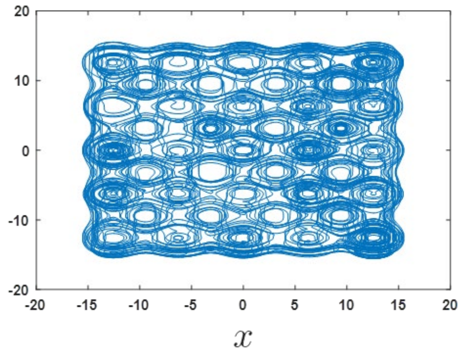
In addition, the scroll of the system (6) will be generated at the center point, and the bond band will be generated at the unstable saddle point. When $N=2n\pi$, the equilibrium point of the system (6) has $(2n)^2+(2n+1)^2$ central points and $4n\times(2n+1)$ unstable saddle points. Then, a multi-scroll flow with $(4n+1)\times(4n+$

1) grid type is formed, it is composed of $(2n)^2 + (2n+1)^2$ vortices and $4n \times (2n+1)$ bond bands.

Set $n = 1, 2$, 5×5 and 9×9 grid type multi-scroll flows are shown in Fig. 3.



(a) 5×5 grid type multi-scroll flow



(b) 9×9 grid type multi-scroll flow

Fig. 3. grid type multi-scroll flows

3.4. NIST test of system (6)

At present, NIST SP800-22 standard is the most widely used and authoritative standard used to detect the pseudo-random sequence generated by the system (6). It has 15 test indexes. The ideal random sequence is taken as a reference to test the deviation degree of pseudo-random sequence from different angles in statistical characteristics. It is generally believed that the sequence that can pass the detection has a good pseudo-random performance. Each test of SP800-22 standard will provide the pass rate and the uniformity of P value distribution. If the significance level $\alpha = 0.01$ and the test sequence is group β, the confidence interval of the passing rate can be defined as:

$$\left(1 - \alpha - 3\sqrt{\frac{\alpha(1-\alpha)}{\beta}}, 1 + \alpha - 3\sqrt{\frac{\alpha(1-\alpha)}{\beta}}\right) \qquad (8)$$

The test conditions used in this paper are: significance level a = 0.01, test sequence B = 100, length of each group is 10bit, and the confidence interval is 1. The test results are shown in Fig. 4.

No.	Statistical Test	P-value	Test Times	proportion	Result
1	Frequency	0.867692	1	0.99	Pass
2	Block Frequency	0.319084	1.00	1.00	Pass
3	Cumulative Sums	0.236810	2	0.99	Pass
4	Runs	0.678686	1	1.00	Pass
5	Longest Run	0.897763	1	0.97	Pass
6	Rank	0.924076	1	0.99	Pass
7	FFT	0.236810	1	0.99	Pass
8	NonOverlapping Template	0.017912	148	0.97	Pass
9	Overlapping Template	0.474986	1	0.98	Pass
10	Universal	0.137282	1	1.00	Pass
11	Approximate Entropy	0.657933	1	1.00	Pass
12	Random Excursions	0.051391	8	1.00	Pass
13	Random Excursions Variant	0.010606	18	0.97	Pass
14	Serial	0.350485	2	1.00	Pass
15	Linear Complexity	0.062821	0.99	1.00	Pass

Fig. 4. The results of NIST test

The performance of pseudo-random sequence is comprehensively analyzed in the experiment. The experiment can pass only when the following conditions are met:

- All P-values must be greater than significance level α.
- The pass rate must be in the confidence interval.
- P-values must be evenly distributed.

It can be seen from Fig. 4 that the system (6) has passed the test and proved that the system has good pseudo randomness.

In this paper, taking Non-Overlapping Template as an example, the system (6) can be found to satisfy the uniform distribution of P-value. P-value distribution of Non-Overlapping Template is shown in Fig.5.

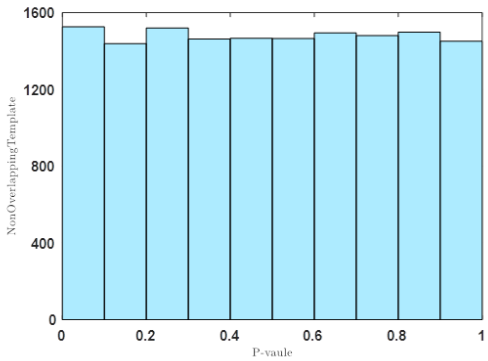


Fig. 5. P-value distribution of Non-Overlapping Template

4. Conclusion

In this paper, a four-dimensional conservative system with a unique equilibrium point is constructed. Firstly, the sine function is introduced to extend the equilibrium point in the x-axis direction so that it has a linear equilibrium point with infinite extension in the x-axis direction. Secondly, it is found that the system (5) has different numbers of multi-scroll flows under different initial values, which proves that the system has multi-stability. Thirdly, by introducing a sinusoidal function to extend the equilibrium points in the x-axis and Y-axis directions, a controllable grid type multi-scroll flow is found. Finally, the system is tested by NIST, and the test results show that the system has good pseudo randomness.

Acknowledgements

References

1. 1. Suykens, J.A.K., Vandewalle, J., Generation of n-double scrolls (n= 1, 2, 3, 4, . . .), IEEE T rans. Circuits Syst. IFundam. Theory Appl, 1993, 40, 861–867.
2. Zhang, C.X.; Yu, S.M. Design and implementation of a novel multi-scroll chaotic system. Chin. Phys. B 2009,18, 119–129.
3. Wang, F.Q.; Liu, C.X. A new multi-scroll chaotic generator. Chin. Phys. 2007, 16, 942–945.
4. Liu, X.Z.; Shen, X.M.; Zhang, H.T. Multi-scroll chaotic and hyperchaotic attractors generated from Chen system. Int. J. Bifurc. Chaos 2012, 22, 1250033.
5. Chen, D.Y; Sun, Z.T.; Ma, X.Y .; Chen, L. Circuit implementation and model of a new multi-scroll chaotic system. Int. J. Circuit Theory Appl. 2014, 42, 407–424.
6. Luo, X.H.; Tu, Z.W.; Liu, X.R.; Cai, C.; Liang, Y .L.; Gong, P . Implementation of a novel two-attractor grid multi-scroll chaotic system. Chin. Phys. B 2010, 19, 070510.
7. Chen, Z.; Wen, G.L.; Zhou, H.A.; Chen, J.Y . A new M x N-grid double-scroll chaotic attractors from Rucklidge chaotic system. Optik 2017, 136, 27–35.
8. Chen, Z.; Wen, G.L.; Zhou, H.A.; Chen, J.Y . A new M x N-grid double-scroll chaotic attractors from Rucklidge chaotic system. Optik 2017, 136, 27–35.
9. Lü, J.H.; Han, F.L.; Yu, X.H.; Chen, G.R. Generating 3-D multi-scroll chaotic attractors: A hysteresis series switching method. Automatica 2004, 40, 1677–1687.
10. Peng, Z.P .; Wang, C.H.; Luo, X.W. A novel multi-directional multi-scroll chaotic system and its CCII+ circuit implementation. Optik 2014, 125, 6665–6671.
11. Wang, C.H.; Xu, H.; Yu, F. A novel approach for constructing high-order Chua's circuit with multi-directional multi-scroll chaotic attractors. Int. J. Bifurc. Chaos 2013, 23, 13500223.
12. Deng, W.H.; Lü, J.H. Generating multi-directional multi-scroll chaotic attractors via a fractional differential hysteresis system. Phys. Lett. A 2007, 369, 438–443.
13. Zhang, Y .; Yu, S.M.; Liu, M.H. Generating multi-scroll hyperchaotic attractors based on FPGA technology. Circuits Syst. 2007, 12, 39–43.
14. Tlelo-Cuautle, E.; Rangel-Magdaleno, J.J.; Pano-Azucena, A.D.; Obeso-Rodelo, P .J.; Nunez-Perez, J.C. FPGA realization of multi-scroll chaotic oscillators. Commun. Nonlinear Sci. Numer. Simul. 2015, 27, 66–80.
15. Yu, S.M.; Lü, J.H.; Leung, H.; Chen, G.R. Design and implementation of n-scroll chaotic attractors from a general jerk circuit. IEEE T rans. Circuits Syst. I Regul. Pap. 2005, 52, 1459–1476.
16. Yalcin, M.E. Multi-scroll and hypercube attractors from a general jerk circuit using Josephson junctions. Chaos Solitons Fractals 2007, 34, 1659–1666.
17. He, S.B.; Sun, K.H.; Wang, H.H.; Ai, X.X.; Xu, Y .X. Design of n-dimensional multi-scroll Jerk chaotic system and its performance. J. Appl. Anal. Comput. 2016, 6, 1180–1194.
18. Ma, J.; Wu, X.Y .; Chu, R.T.; Zhang, L.P . Selection of multi-scroll attractors in Jerk circuits and their verification using Pspice. Nonlinear Dyn. 2014, 76, 1951–1962.
19. T chitnga, R.; Nguazon, T.; Fotso, P .H.L.; Gallas, J.A.C. Chaos in a Single Op-Amp-Based Jerk Circuit: Experiments and Simulations. IEEE T rans. Circuits Syst. II Express Briefs 2016, 63, 239–243.
20. Tang, W.K.S.; Zhong, G.Q.; Chen, G.; Man, K.F. Generation of n-scroll attractors via sine function. IEEE T rans. Circuits Syst. I Fundam. Theory Appl. 2001, 48, 1369–1372.
21. Hu, X.; Liu, C.; Liu, L.; Ni, J.K.; Li, S.L. Multi-scroll hidden attractors in improved Sprott A system. Nonlinear Dyn. 2016, 86, 1725–1734.
22. Luo, X.H. Circuitry implementation of a novel nonautonomous hyperchaotic Liu system based on sine input. Chin. Phys. B 2009, 18, 3304–3308.

Development of in-Home Wireless Continuous Temperature Data Logging and Alarming System for Fever Monitoring in Pediatrics

Ali S. Al-Mejrad
Department of Clinical Laboratory Sciences, Faculty of Applied Medical Sciences University of Hail, Hail, Kingdom of Saudi Arabia.
E-Mail: a.almejrad@uoh.edu.sa

Abstract

Fever is one of the most common concern to be taken care by parents that sometimes force them to bring their child to the hospital emergency for continuous follow up especially at night when sleeping. Fever is defined when a rectal temperature reaches over 38.0 C (100.4 F). Temperatures measured from other sites of body are usually less. There is no defined threshold for fever due to variation of body temperatures between different individuals as much as 1 F. Fever itself is not life threatening unless the rectal temperature is persistent over than 41.6 C (107 F). Fever may be due to a serious illness and usually is caused by common infections that are not serious. However, a fever can be caused by other causes other than infections. Since continuous temperature monitoring is inapplicable at the rectal site, the paper aims at developing a prototype system using Wireless Sensor Network (WSN) for wireless continuous fever monitoring system from the head skin or wrist that are the most applicable area even so one must add 2 degrees for the read temperature correction. The system can record the temperature at the patient terminal and transmit the fever status and critical situations to an alarming system in the side of taking-care people that will help in in-home care for pediatrics suffering from fever.

Keywords: Wireless; Temperature monitoring; data logging; alarming system.

1. Introduction

Fever is one of the most important symptom of many diseases. Fever may be due to a serious illness and usually is caused by common infections that are not serious. However, a fever can be caused by other causes other than infections. The body's temperature increase is controlled by the part of the brain called the hypothalamus to fight the infection that may reach the fever level. Fever is considered one of the most common concern to be taken care by parents that sometimes force them to bring their child to the hospital emergency for continuous follow up especially at night when sleeping. Fever is defined when a rectal temperature reaches over 38.0 C (100.4 F). Temperatures measured from other sites of body are usually less. There is no defined threshold for fever due to variation of body temperatures between different individuals as much as 1 F. Fever itself is not life threatening unless the rectal temperature is persistent over than 41.6 C (107 F) [1][2].

In the past, body temperature monitoring was using a thermometer constructed of glass filled with mercury. "However, the American Academy of Pediatrics (AAP) recommends against using mercury thermometers because the glass can break, and mercury is poisonous [3]". In addition, this kind of thermometer can work with adult and cannot be used properly with infant even at the rectal site. The most commonly used and recommended are Electronic thermometers with a digital temperature readout display. Another type of thermometers is the plastic strip that change color to indicate the temperature. This type is the least accurate. The strip is used by placing it on the forehead and after one minute, the temperature can be read while the strip is in its place. Another type of plastic strip thermometer is used for the mouth. Hence, there are many kinds of temperature measurement methods but most of it is inconvenient to infants and child monitoring, because monitoring means continuous reading of the temperature [4].

Therefore, a method for continuous temperature monitoring has to be developed to be used in home and still be safe and convenient [5]. This method will enable a continuous read out of the temperature data utilizing the wireless technology from multiple transmitter node that enable the person on charge to monitor the temperature of many patients (infant, and child). This paper aims at developing a prototype system using embedded technology and Wireless Sensor Network (WSN) technology for wireless continuous fever monitoring system from the head skin or wrist that are the most applicable area taking in consideration adding 2 degree for the read temperature correction [6]. The wireless temperature monitoring system (WTMS) provides 24/7 continuous temperature monitoring of critical patient (Infant, and Child). Continuous monitoring provides the means to detect abnormalities (fever) at an early stage. Utilizing wireless technology eliminates the need for cabling. WTMS allows for easy field data collection. Temperature data is transmitted from the sensor at the transmitter node to a nearby receiver node via wireless connection. The receiver node can be connected to a computer via USB. In addition, the receiver node can work standalone without the PC. A universal receiver node can be utilized as a data logger that can collect data from any transmitter node. This universal receiver and the attached PC can work to trigger visual and/or audible alarm for preset temperature values for fever detection even more it can send SMS to the smart phone belonging to the person in charge such as parents. In addition to that, the data can be analyzed later for diagnosis and reported by a dedicated software package.

2. System Development

2.1 Hardware and System Block Diagram

The temperature wireless system designed consists of:

A. Transmitter Node
B. Receiver Node
C. Data Access Module

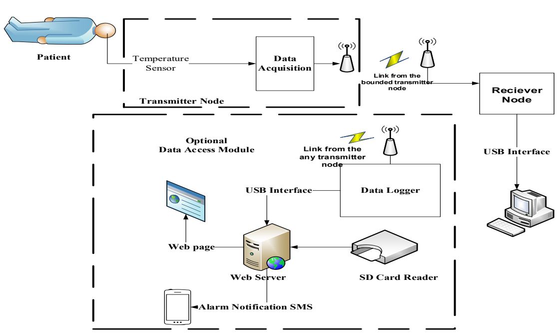


Figure 1. System Block Diagram

Fig. 1 shows a general block diagram of a wireless temperature monitoring system (WTMS). The system consists of three major parts: transmitter node, receiver node, and optional data access module.

A. Transmitter Node

The system is comprised of Wireless transmitter node that consists of a power subsystem (battery), temperature sensor, and data acquisition. This module is based on wireless Wi-Fi technology. This module is designed to operate in moisture environment and is protected by adding moisture resist material that will protect and isolate the patient from being connected to the circuit via moisture. The transmitter node has a lead off detector to trigger an alarm on the receiver node when there is no or bad contact on the patient skin. The transmitter is attached to the patient skin all the time by the mean of head bandage or wrist strap that will give the patient a convenient way of every day practice. Figure 2 shows the transmitter node, which consists of temperature sensor; lead off detector, microcontroller, battery, and the transmitting antenna.

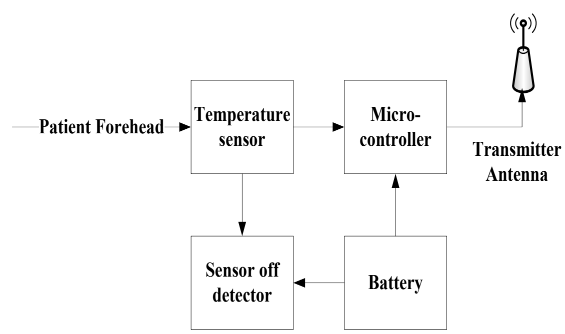


Figure 2. Transmitter Node

B. Receiver Node

It consists of a power subsystem (battery and battery management; recharge circuit operate from USB port), and readout device (LCD). This module is based on wireless Wi-Fi technology that can be directly connected to the PC via USB for data collection and further analysis. The receiver node contains audible and visual alarms, and data storage unit (SD card). Any receiver node can be programmed to collect data from many transmitters, as the case when there are many children or infants to be monitored for temperature rise. The receiver contains SD card to store the data for further analysis by physicians. The collected data can be sent to physician over the internet under user control when the receiver node is connected to a PC. Figure 3 shows the receiver node, which consists of receiving antenna, microcontroller which control: LCD as a reading out device for the temperature and alarm values, buzzer for audible alarm, mass storage (I2C EEPROM, and SD card). The controller has a USB interfacing capability to the PC in order to collect the data for further analysis. In addition, the receiver node battery could be charged from the PC USB port.

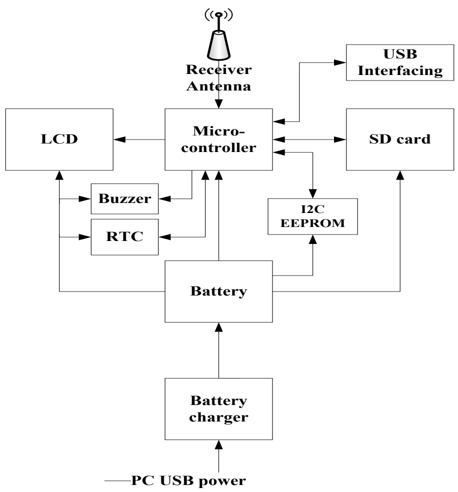


Figure 3. Receiver Node

C. *Data access module*

It consists of data logger module which can be connected to the PC or a web server for data distribution over the web, sending alarm notification SMS to smart phones. SD card
reader can be connected to the server for reading SD card from any receiver. The data logger can connect to any transmitter node(s) for data collection and analysis. The data logger is mainly powered from the PC USB port.

2.2 Software and User Interface

A user interface is developed using Laboratory Virtual Instrument Engineering Workbench to control the wireless module and collect the temperature data of patient events of high temperature threshold of fever as shown in Figure 4 [7].

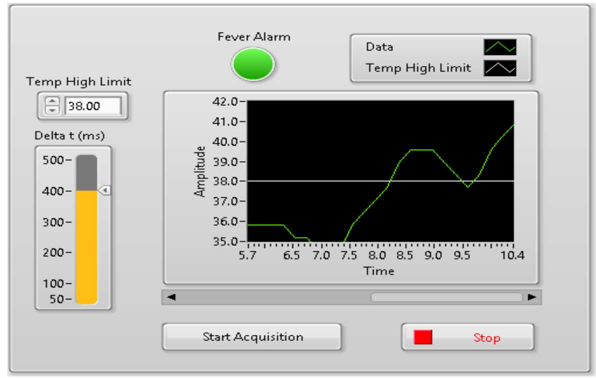


Figure 4. Front Panel of the User Interface of Temperature Data Collection and Fever Alarms Events

3. Results and Discussion

Analysis of Body Temperature Measurements Comparison Test

Testing the sensor DS18B20 as body temperature sensor is done with comparison to the measurement results with mercury thermometer. The selection of this type of thermometer is due to its accuracy compared to other types of thermometers. Table 1 below shows the comparison results, the measurement difference and the average measurement error.

No.	Temperature Sensor (^{0}C)	Mercury Thermometer (^{0}C)	Measurement Difference (^{0}C)
1	36.75	37.60	0.85
2	36.63	37.25	0.62
3	36.25	36.97	0.72
4	36.35	36.94	0.59
5	36.56	37.25	0.69
6	36.06	36.70	0.64
7	35.95	36.49	0.54
8	36.85	37.28	0.43
9	36.75	37.57	0.82
10	36.70	37.53	0.83
11	36.50	37.30	0.80
12	36.53	37.16	0.63
13	36.54	36.96	0.42
14	36.26	36.71	0.45
15	36.23	36.75	0.52
Average - Error			0.64 %

Table 1. Body temperature test results

From the test results, obtained results showed the temperature has an average error around 0.64% using the Mercury Thermometer compared to the developed system. Therefore, it can be stated that the sensor reading s are very good.

Analysis of Events of High Temperature Threshold of Fever Test

From the test results of some of collected temperature data shown in Table 2 which gives continuous monitoring for critical events of fever (as shown in RED colors) , and alarming system to prevent deterioration of the patient condition

No.	Time Data	Temp Data	Time Temp High Limit	Temp High Limit	Events of Alarms (in RED)
1	5.7	35.8	5.7	38	
2	5.9	35.8	5.9	38	
3	6.1	35.8	6.1	38	
4	6.3	35.8	6.3	38	
5	6.5	35.2	6.5	38	
6	6.7	35.2	6.7	38	
7	6.9	34.6	6.9	38	
8	7.1	34.6	7.1	38	
9	8.4	38.9	8.4	38	FEVER
10	8.6	39.6	8.6	38	FEVER
11	8.8	39.6	8.8	38	FEVER
12	9	39.6	9	38	FEVER
13	9.2	38.9	9.2	38	FEVER
14	9.4	38.3	9.4	38	FEVER
20	9.6	37.7	9.6	38	

Table 2. Events of High Temperature Threshold of Fever Test Results

Analysis of Data Transmission Performance Test

From the test results of this communication between the master module and the slave, the slave module used can accept data from master module according to data delivered.

The performance test of Nodemcu ESP8266 works to find out if the sensor is accessible by other devices with a certain distance. Testing is done by sending data on Nodemcu ESP8266 to display the results on the website and then test access with a certain range. Distance parameter of the maximal use range is about 50 meters. From the tests that have been done, it can be concluded that the website is inaccessible in the range above 50 meters. One of the most important factors that affect the process of access and delivery of the data is indoor so possible transmitted signals will be hindered and cannot be well reached.

4. Conclusion and Future Work

The designed system goal is to provide in home, self-dependent, continuous monitoring for critical events, and alarming system to prevent deterioration of the patient condition. The system has many features with wireless (Wi-Fi) connectivity in range of up to 50 meter (the distance between the transmitter and receiver nodes), typical signal reception of once every 1-10 minutes under user control, frequency of data transmission increases with fast temperature rise, temperature range 25°C to 50 °C with 0.1 °C resolution, preset alarm values, moisture resistant for humid environments.

Acknowledgement

I like to thank the Deanship of Scientific Research of University of Hail, for their support of my research.

References

1- Herzog LW and Coyne LJ., "What is fever? Normal temperature in infants less than 3 months old", Clin Pediatr, 32, pp.142-6, 1993.

2- Teach SJ and Fleisher GR., "Duration of fever and its relationship to bacteremia in febrile outpatients three to 36 months old", The Occult Bacteremia Study Group. Pediatr Emerg Care, 13, pp. 317-9, 1997.

3- Lynn R. Goldman, Michael W. Shannon and the Committee on Environmental Health, "Technical Report: Mercury in the Environment: Implications for Pediatricians, Pediatrics", Vol. 108 No. 1, pp. 197-205, 2001.

4- Kyriacou P. A., "Temperature sensor technology", Biomedical Sensors, pp. 1-38, 2010.

5- IEC 60601-1, Medical Electrical Equipment- Part I: General Requirements for Safety and Essential Performance, International Electrotechnical Commission, Geneva, Clause 8.7,1995.

6- A. Tzavaras, and B. Spyropoulos, " Development of a low-cost wireless monitoring System supporting the Continuity of Medical Care of the Patient at home", 35th Annual International Conference of the IEEE EMBS Osaka, Japan, 3 - 7 July, 2013.

7- Mehta B. et al. "Real Time Patient Tele- monitoring System Using LabVIEW", International Journal of Scientific & Engineering Research (IJSER), Volume 3, Issue 4, April-2012.

Design and Development of in-home Wireless Crucial Events Logging and Alarming System for Elderly and Disabled People Care

Ali S. Al-Mejrad
Department of Clinical Laboratory Sciences, Faculty of Applied Medical Sciences University of Hail, Hail, Kingdom of Saudi Arabia.
E-Mail: a.almejrad@uoh.edu.sa

Abstract

This paper aims at developing a system that will help in in-home care for elderly or disabled people. The system can record the patient vital parameters such as temperature, SpO2, ECG, heart rate, bed moisture and fall off sensor which could be installed at the bathroom or shower area at the patient terminal and transmit the critical situations such as high or low values to an alarming system. The system will be developed using the National Instrument Wireless Sensor Network (NI-WSN) to control the crucial events transmission and LabVIEW software to design the user interface.

Keywords: Elderly; Disabled; In-home care; NI-WSN**;** PSoC**;** LabVIEW.

1. Introduction

The growth in the number and proportion of elderly and disable people are significantly increase which lead to the need of a special care system to give these people a chance to live depending on themselves [1]. In addition, the resources in in-patient facilities are limited and expensive. Elderly people are suffering many health problems, mobility limitation, and chronic physical problems and, other disorders compared to younger age people. All of which are requiring special continuous attention and care. The family and/or other help provider often shares the task of caring for these people. However, the elderly usually remain independent and alone at home and this becomes a critical issue due to non-availability of their help provider. While at home, smart sensors such as body temperature and ECG sensors can be used to wirelessly connect through NI-WSN to a Computer, which will record and analyze different parameters such as, body temperature, blood pressure, cardiac function and many others [2]-[4]. A connection to the Internet keeps the Elderly and disabled people within reach of their help providers and family. This method of transmitting selected data only when required rather than streaming all data continuously, is perfect method for monitoring critical situations and activating alarms for these critical cases [5]-[8]. Elderly people in Saudi Arabia are suffering from chronic health conditions such as chronic cardio-respiratory diseases, which is one of the major causes of death and admission to hospital in addition to many other concurrent problems [9]. Due to frequent occurrences of such diseases to these people who in need for continuous care will lead to costly hospitalization of such patients. In this paper a system will be developed that will help in in-home care for elderly or disabled people by recording the patient vital parameters such as SPO2, heart rate, ECG, temperature, bed moisture, and fall off sensor on the chair or the bed. In addition, these sensors could be installed at the bathroom or shower area at the patient terminal and transmit the critical situations such as high or low values to an alarming system. The system will be developed using the National Instrument Wireless Sensor Network (NI-WSN) to control the curial events transmission and LabVIEW software to design the user interface.

2. System Development

The designed system in the block diagram consists of different components as shown in Figure.

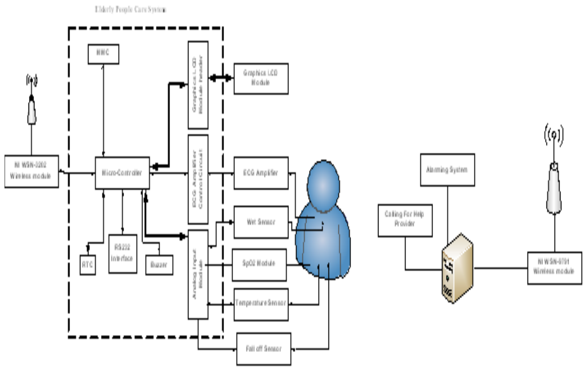


Figure 1. Block Diagram of the Designed System

The system has two major components; the console that is fixed at the patient chair, bed, or even held with the patient with the NI-WSN-3202 wireless module [10] and the other component is the NI-WSN-9791 Ethernet gateway [11] connected to the PC that is also connected to an alarming system to trigger alarm on critical situations. A user interface program using LabVIEW is developed to control the wireless module and collect the data of patient events such as high heart rate, high temperature, and fall off the chair or bed [12].

Measurement node NI-WSN-3202:

The measurement nodes connect directly with sensor and are able to transmit wirelessly a 2.4 GHz radio data to the WSN Ethernet gateway. Each measurement node has four analog input channels and four digital Input/output channels.

The measurement nodes and the NI WSN-9791 Ethernet gateway:

The measurement nodes and the Ethernet gateway communicate wirelessly using a radio frequency of 2.4 GHz based on the reliable IEEE 802.15.4 NI WSN protocol. The network can accommodates up to 36 nodes per gateway and provide an outdoor range of up to 300 m.

Figure 2. NI Wireless Sensor Network

Microcontroller PSoC (CY8C29466-24PVXI)

Programmable System on Chip (PSoC) is a programmable embedded system on chip which integrating on a single chip configurable digital and analog peripheral functions, memory and a microcontroller [13]. PSoC is designed to replace components in traditional MCU-based system with a single-chip programmable device. PSoC device has configurable blocks of digital and analog, as well as programmable interconnections. This architecture enables designer for creating customized peripheral configurations. This microcontroller is responsible for data collecting from all the analog modules connected to the patient and preparing the data to be transferred to/from the wireless module. Also to keep a complete trend in the MMC module as a data logger.

3. Results and Discussion

First, a prototype test system of three main signals including pulse oximeter, temperature and ECG using PSoC is developed to ensure the success of the data collection. The block diagram of the prototype test system with LCD shown in Figure 3 consisting of three main traces are described below:

Figure 3. Prototype test system of three main signals including pulse oximeter, temperature and ECG using PSoC

3.1 Pulse Oximeter:

First module is Pulse Oximeter, the frequency of this signal in the range of 1Hz to 40Hz the sensor used in this module is photo sensor sender and receiver, the finger of the patient located between the them, the infrared opt coupler will detect the volume change in finger's capillaries and to obtain blood flow profile.

Programmable Gain Amplifier (PGA)

The received light is converted into an electric signal by a photodiode and then enters the first block inside the PSoC controlled is PGA here which act as a buffer to match the circuit input output impedance, and also amplify with gain of 4.

Filters

The second block is filter we use two types of filter the first is second order Low pass filter, this filter is used to cut the high frequency component, the corner frequency of the filter is 100Hz, the other filter is Band pass filter is used to cut plethysmograph signal band (1-40Hz)

Pulse Peak detector and pulse counter

The function of this block is to detect the peaks of the signal and pass it to the pulse counter, which is a digital counter in order to count the heart rate.

ADC Converter

The analog to digital converter used to acquire the signal in digital form in order to display it on the LCD; here we use the PSoC 12 bits ADC converter with sampling rate of 50.

Now the signal will be handled by to the PSoC internal controller in order to display it on the LCD.

3.2 Temperature

PGA

The PGA here amplifies the sensing signal from the bridge circuit and pass it to the ADC converter.

ADC Converter

Converts the thermistor voltage into digital form to the PSoC internal controller to check the calibration tables and display the temperature on the LCD by using the graphic LCD routine.

3.3 ECG

Instrumentation Amplifier

In our system, some external circuits (outside the PSoC controller) were made such as Instrumentation amplifier and isolator circuit because of some limitation in the ranges of the PSoC.

Filters

The second block is filter. Two types of filter were used the first is second order Band Pass filter, this filter is used to cut the high and low noise and pass only the ECG signal (1-150 Hz), the other filter is Notch filter is used to cut mains noise 60Hz. The response of the amplifier circuit has been verified using Orcad Pspice simulation program and the response is illustrated in Figure 4 bellow

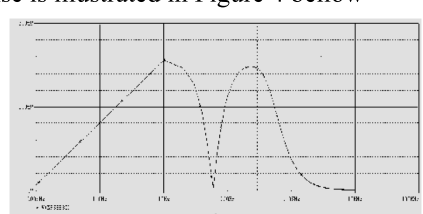


Figure 4. The total ECG filter response

PGA

Act here as the main amplifier of the ECG signal, and uses a gain of 16.

ADC Converter

The analog to digital converter used to acquire the signal in digital form in order to display it on the LCD; here we use the PSoC 12 bits ADC converter with sampling rate of 50.

Now the signal will be handled by to the PSoC internal controller in order to display it on the LCD.

3.4 Designed PSoC Circuit Diagram

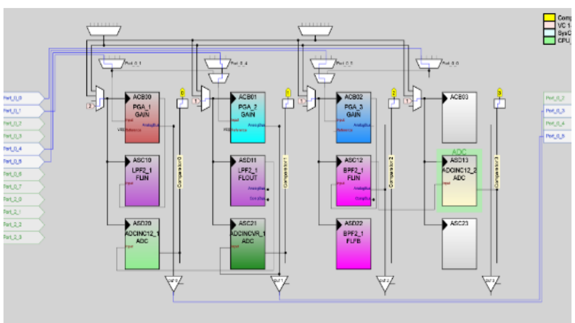


Figure 5. Designed PSoC circuit diagram

The analog signals of the three signals sampled in PSoC micro-controller designed as shown in Figure 5 are then stored in the system EEPROM and/or the MMC card on the system shown in Figure 1 for later analysis. The digital data is wired to the measurement node NI-WSN-3202 that is responsible for the wireless transfer of the data to the NI WSN-9791 Ethernet gateway receiver. The receiver node is connected to the PC through the Ethernet port for data collection and alarm triggering. Experimentation utilizing the prototype based on wireless technology with the PSoC system has shown that the communication range varies from 30 to 50 meters depends on the home walls that is sufficient.

3.5 Designed System and its User Interface

The complete system components is shown in Figure 6 and hardware test setup is shown in Figure 7.

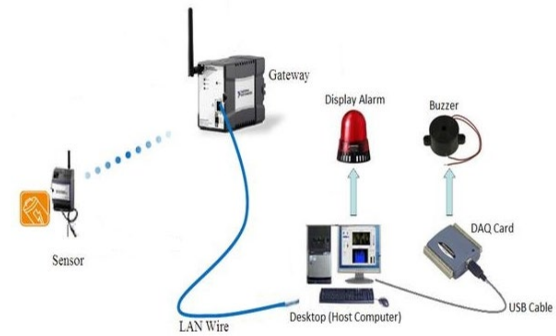


Figure 6. The complete system components

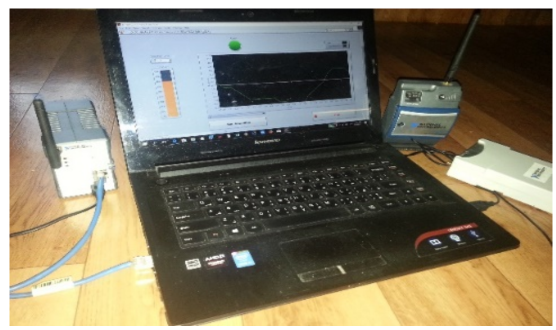


Figure 7. Hardware test setup with user interface

The system includes main components including a user interface program as shown in Figure 8. The system is developed to control the whole system by logging in the system and then configuring the events with minimum and maximum limits and configuring wireless module. When running, the system starts patient data acquisition, and display of the data and events of the patient such as high temperature, high heart rate, low oxygen saturation, and fall off the chair or bed. Finally, patient's data and events can be analyzed for follow-up and diagnosis purposes.

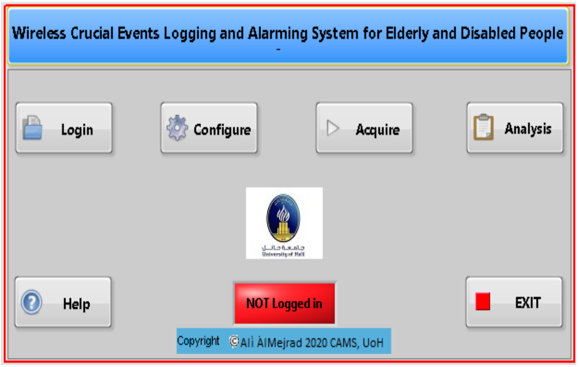


Figure 8. User Interface System

4. Conclusion and Future Work

The designed system goal is to provide in home, self-dependent, monitoring for critical events, and alarming system. As the majority of elderly people prefer to stay in their own homes and may require help and supervision. The designed system is monitoring the vital signs of the patient such as SpO2, heart rate, temperature, ECG, and the critical status (wet clothes and fallen off the chair) which are added easily as ON/OFF in order to trigger alarms for such events.

The system is connected to a PC that can easily transfer the collected data on the web for long distance further data analysis and/or call for help. This continuous remote monitoring and support will increase the elderly people freedom and safety; and prevent patient condition deterioration. This in turn will improve the elderly life quality. The designed system was developed as a prototype test system including three signals including pulse oximeter, temperature and ECG using PSoC with LCD. The future work is all about reducing the component using the Analog Front End (AFE) chips to improve power consumption and reduce the cost.

Acknowledgment

I like to thank the Deanship of Scientific Research of University of Hail, for their support of my research.

References

1- United Nations, Department of Economic and Social Affairs, Population Division. World Population Ageing 2015 (ST/ESA/SER.A/390), 2015.

2- Eduardo Freitas and Amandio Azevedo, “Wireless Biomedical Sensor Networks: The Technology”, Proceedings of the 2nd World Congress on Electrical Engineering and Computer Systems and Science (EECSS'16) Budapest, Hungary – August 16 – 17, 2016.

3- Panicker N. and A. Sukeshkumar, “Wireless sensor based systems for biomedical monitoring: A review”, AMSE JOURNALS –2014-Series: Modelling C; Vol. 75; No. 1; pp 65-78, 2014.

4- Demongeot, J. et al.,”Multi-sensors acquisition, data fusion, knowledge mining and alarm triggering in health smart homes for elderly people”, C. R. Biologies 325:673–682, 2002.

5- Clark, A. and K. EMMA, “A comprehensive alarm system for the elderly and infirm and disabled who live at home”, Res. Soc. Health Care, 15:205–27, 1979.

6- Chan, M. et al.,”Smart house automation system for the elderly and the disabled”, IEEE Int. Conf. Syst. Man Cybern, 2:1586–1589, 1995

7- Peeters, P. H., “Design criteria for an automatic safety-alarm system for elderly”, Technol. Health Care, 8:81–91,2000.

8- A. Tzavaras, and B. Spyropoulos, “ Development of a low-cost wireless monitoring System supporting the Continuity of Medical Care of the Patient at home”, 35th Annual International Conference of the IEEE EMBS Osaka, Japan, 3 - 7 July, 2013

9- Khan J. et al., “The Saudi Guidelines for the Diagnosis and Management of COPD”, Ann Thorac Med. 9(2): pp. 55–76, 2014.

10- NI-WSN-3202 wireless module – User guide and Specifications, 2010.

11- NI WSN-9791 Ethernet Gateway – User guide and Specifications, 2010.

12- National Instruments Corporation, “LabVIEW User Manual”, Edition 2014.

13- Cypress Semiconductors, PSoC 5LP Architecture Technical Reference Manual, Document No. 001-78426 Rev. *F, 2016.

Virtual bird's-eye for remote operation of unmanned construction machinery

Noritaka Sato, Akihiro Fukuda*
Morita and Sato Laboratory, Department of Electrical and Mechanical Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Aichi, 466-8555 Japan
Email: a.fukuda.975@nitech.ac.jp
http://watt.web.nitech.ac.jp/index_2.html

Abstract

Recently, remote operation of construction machinery is required to release humans from dangerous places. However, the risk of collisions and operation time increases because of lacking a sense of distance. To solve those problems, we propose a method to display a virtual bird's-eye view. Experiments are carried out to verify the effectiveness of the proposed method which can reduce the number of collisions and operation time comparing to the conventional method.

Keywords: virtual reality, remote operation, construction machinery, UI

1. Introduction

Recently, remote operation of construction machinery is required to release humans from dangerous places. One of the situations where the remote operation of construction machinery is required is "the sorting operation of slags" in iron factories shown in Fig. 1. "The sorting operation of slags" is the work of sieving slags with a power shovel equipped with a bucket and consists of the following three processes: shoveling up slags, sieving slags, and releasing the remains. The workplace is harsh for workers because of flying dust, the high temperature, and the high humidity.

The operation by the remote control of a construction machine was carried out by using the active robot SAM[1] in the past. However, problems such as a significant increase in working time and the collision of the power shovel with obstacles have occurred. It is considered that the reason for them is lacking the sense of distance because of only using a monocular camera.

Fig. 1 The sorting operation of slags

The research by Shiroma et al. has proved that a bird's-eye view of the operation target improves the sense of distance[2]. However, it is difficult to install cameras in the workplace and take bird's-eye view images of the construction machine due to problems such as the enormous amount of data communication, limited installation position, and poor visibility due to dust. The researchers also proposed a system that creates a virtual bird's-eye view by superimposing a CG model of the robot on an image obtained while the robot is moving[3].

In the previous paper, we proposed a new method to support the remote operation of construction machinery based on the system which creates virtual bird's-eye view images and verified the effectiveness of it on the sorting operation of slags in the real environment with a professional operator[4]. However, there was a problem which was an insufficient number of trials as it was hard to have the professional operator to make time. In this paper, we did an additional experiment to get confirmation of the effectiveness of the proposed system.

2. Proposed system

2.1. *System overview*

In the proposed system, virtual bird's-eye view images are created by superimposing a CG of the power shovel on images. Since the position of the power shovel did not change during the operation, the position and the posture of the CG without the posture of the body of the power shovel are changed based on measured values obtained from sensors attached to the power shovel.

2.2. *Real-time video*

In addition to virtual bird's-eye view images, real-time videos are displayed. In the experiment using the actual power shovel, cameras are mounted as shown in Fig. 2 for displaying slags and the inside of the bucket during sieving.

In the experiment in this study, a camera for slags shown in Fig. 2 is used, because the motion of the sieving is not required in the experiment.

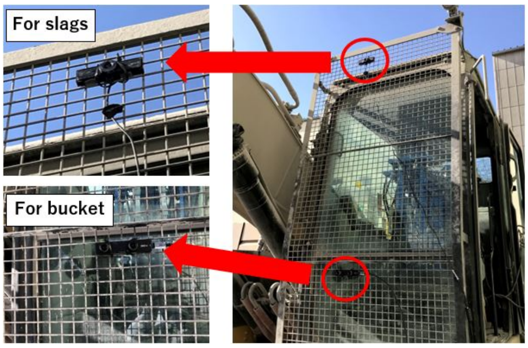


Fig. 2 Cameras on the power shovel

2.3. *Sensing position and posture*

To sense the position and the posture of the power shovel, we mounted some sensors on it. Fig. 3 shows mounted sensors. The angle of each link is calculated from the acceleration measured by the IMU attached to them. The position and the posture of the power shovel are measured by taking a picture of the AR markers and using ARToolKit[3] to get the relative position and posture from the external camera.

In the experiment in this study, the angle of each link is measured by an encoder on the servo motor, and the position and the posture of the body are fixed, therefore there is no sensor to measure them.

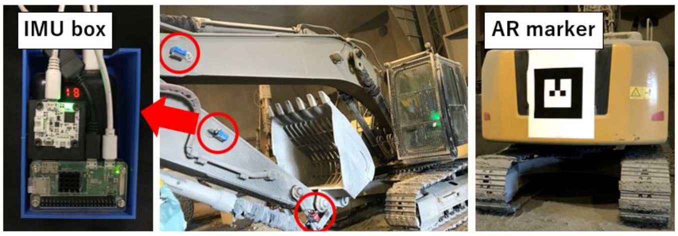


Fig. 3 Installed sensors

2.4. *Creating virtual bird's-eye view images*

Virtual bird's-eye view images were created using a game engine, Unity by the following steps in shown Fig. 4:

(i) Putting the images taken in advance inside the virtual space.
(ii) Putting a 3D CG model of the power shovel inside the virtual space based on its measured position and posture.
(iii) Getting a view from the position where the image was taken. (The CG model is virtually superimposed)

Fig. 4 Generating virtual bird's-eye view

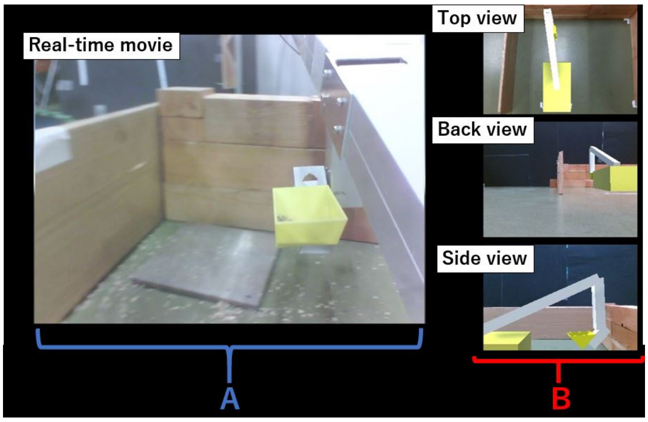


Fig. 5 The image shown to operators

Fig. 5 shows the video displayed to the operator. The left-side video (A) is a real-time video and the right-side videos (B) are virtual bird's-eye view images.

3. Experiments

3.1. *Experimental details*

The simulation was conducted in the environment and procedure shown in Fig. 6. Two male subjects in their 20s have no operation experience in the robot and were given 10minutes to practice. The sorting operations were conducted 10 times each using the real-time video only ("A" shown in Fig. 5) and the proposed system ("A" and "B" shown in Fig. 5). They were conducted in an alternating cycle of five times to remove the effect of the order.

The data recorded were operation time, number of collisions, and subjective evaluation by questionnaire.

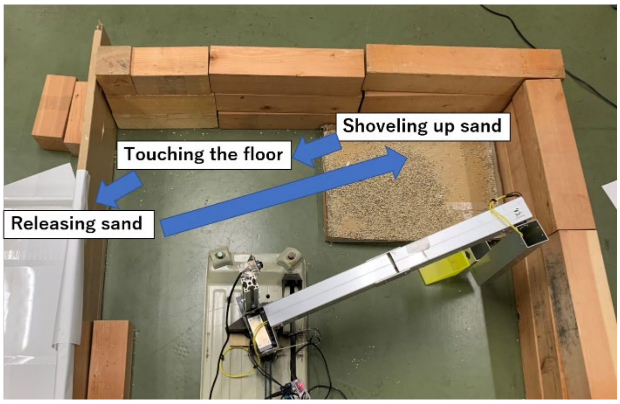


Fig. 6 The environment of the simulation

3.2. *Result*

Figs. 7, 8 show each subject's results: average operation time of 10 times and standard deviation. In subject 1, an average operation time was 53.7 sec when using the only real-time video and 48.4 sec when using the proposed system, and it was reduced by 9.87%. The standard deviation was 9.58sec when using the real-time video and 7.39 sec when using the proposed system, and it was reduced by 22.8%. In subject 2, the average operation time was 46.7 sec when using the real-time video and 39.1 sec when using the proposed system, and it was reduced by 16.3%. The standard deviation was 12.0 sec when using the real-time video and 6.10 sec when using the proposed system, and it was reduced by 49.1%.

On the other hand, an average of the number of collisions was 0.36 times per one cycle when using the real-time video and 0.16 times per one cycle when using the proposed system in subject 1. 0.23 times per one cycle when using the real-time video and 0 times per one cycle when using the proposed system in subject 2.

In the questionnaire, the usability of the virtual bird's-eye view images was surveyed. Subject 1 answered that the side view and the top view were useful to recognize the sense of distance between the bucket and walls and the back view was not useful. On the other hand, subject 2 answered that only the side view was useful.

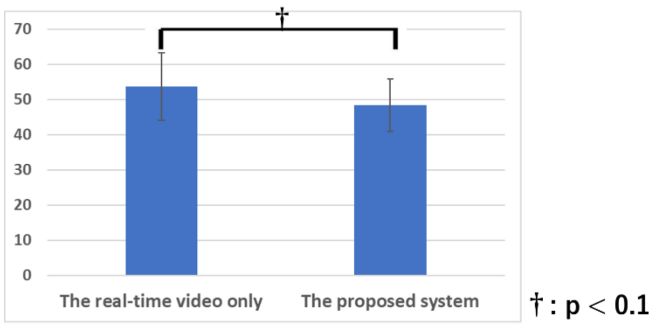


Fig. 7 Operation time (Subject1)

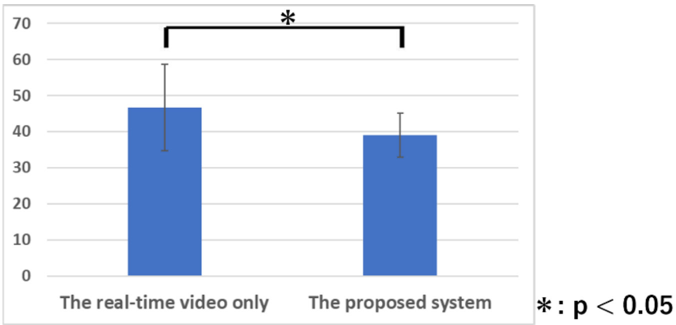


Fig. 8 Operation time (Subject2)

3.3. *Discussion*

Fig.7 and Fig. 8 show that the proposed system is effective for the reduction of the operation time. The results of the t-test showed that the proposed system is statistically significant against the real-time video system for each subject. The reason for the difference in the significant trend between the subjects is lack of practice.

On the other hand, the average number of collisions was reduced. Therefore, it is considered that the proposed system shows effective views to help the operation.

Moreover, there are some feedbacks that they are effective in the questionnaire. However, the images that are found to be effective or ineffective are different for each subject.

4. Conclusion

The purpose of this study is to reduce the operation time and risk of collisions when the power shovel was remotely operated to sort slags. In the previous paper, the efficacy of the proposed system was verified in the real environment. In this paper, we carried out an additional experiment to get confirmation of the effectiveness of the proposed system.

Experiments showed that the proposed system was effective for the operation and it was statistically significant for subjects. Furthermore, the superiority of the proposed system is also confirmed by subjective evaluation.

In the future, it is necessary to increase the number of subjects and conduct a between-subjects design.

References

1. T. Toyota, Development of the Rubber Artificial Muscle Drive Type Remote Operation Robot for Disaster Restoration, *Journal of the Robotics Society of Japan.Vol.33*, 2015, pp. 600-606.
2. N. Shiroma et al. *13th IEEE International Workshop on Robot and Human Interactive Communication*. 2004, pp. 107-112.
3. N. Shiroma et al. *Transaction of the Society of Instrument and Control Engineers.* Vol. 41, No. 12, 2005, pp. 1036-1043
4. A. Fukuda et al. Virtual bird's-eye view for remote operation of unmanned construction machinery, *Reiwa 2nd year Tokai-Section Joint Conference on Electrical, Electronics, Information, and Related Engineering,* 2020.

Domain Transformation of Chest CT Images Using Semi-Supervised Cycle GAN for Opacity Classification of Diffuse Lung Diseases

Masashi Miyake
Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Tokiwadai2-16-1
Ube, Yamaguchi 755-8611, Japan

Shingo Mabu
Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Tokiwadai2-16-1
Ube, Yamaguchi 755-8611, Japan

Shoji Kido
Graduate School of Medicine, Osaka University, Yamadaoka2-15
Suita, Osaka 565-0871, Japan

Takashi Kuremoto
Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Tokiwadai2-16-1
Ube, Yamaguchi 755-8611, Japan

Abstract

The aim of this research is to perform domain translation of chest CT images so that medical institutions can effectively use a computer-aided diagnosis (CAD) system trained at a different institution. We propose a semi-supervised Cycle GAN for domain transformation by combining the standard Cycle GAN and the trained CAD. In the experiment, we classified opacities of diffuse lung diseases in CT images and clarified the effectiveness of the proposed method for domain transformation.

Keywords: Deep Learning, Domain transformation, GAN, chest CT, Classification, Semi-supervised

1. Introduction

Recently, with the development of digital medical imaging devices, it has become possible to acquire an enormous amount of medical images. Therefore, studies on computer-aided diagnosis (CAD) have been actively conducted, where machine learning is applied to build CAD systems using a large number of medical images. Image processing and image recognition techniques are used to construct CAD systems. Recently, the research of convolutional neural networks has been rapidly developed in the research field of image recognition and has been applied to classifiers used in CAD systems[1]. However, training a classifier requires a large amount of annotated medical image data. Also, there are possibilities that the accuracy of diagnosis may fluctuate when the photography conditions of the images are different depending on the medical institutions. For example, different CT systems and their settings can change the pixel values. Therefore, a CAD system showing good diagnostic performance in a certain medical institution does not always show the same performance in other medical institutions. In that case, the classifier needs to be retrained, and a large amount of training data for each medical institution needs to be prepared for training. The current system is not easy to be used widely in various medical institutions, and it is necessary to solve this problem.

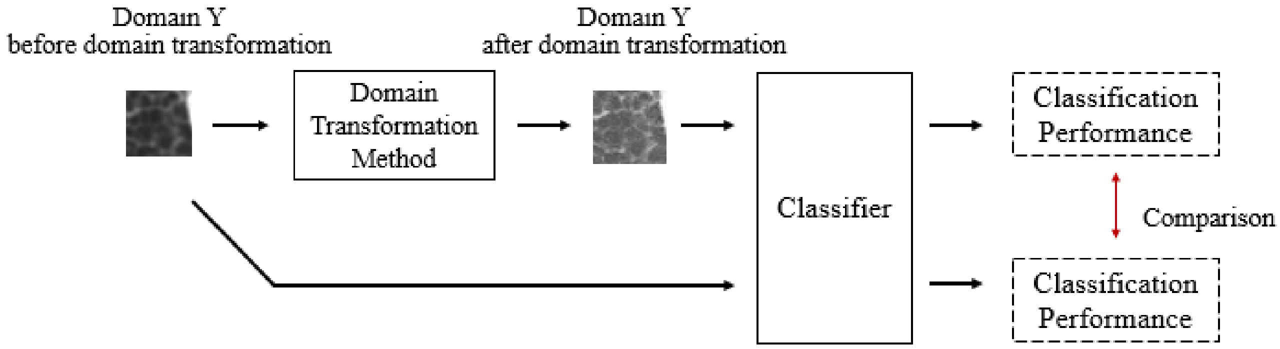


Fig. 1. The flow of the validation of the proposed method. We compare the classification performance of several domain transformation methods including the proposed method and without domain transformation. Refer to Table 1 for the details of domain transformation methods.

A domain transform method to classify opacities in chest CT images using Cycle GAN[2] has been proposed as a method to solve the above problem. However, since the training of domain transformation does not use labeled data, the transform is not always suitable for the classification of the lung opacities. Hence, we propose a semi-supervised Cycle GAN that combines the standard Cycle GAN with a trained classifier that is trained in other medical institutions. In detail, 1) the trained classifier classifies a small number of images of another medical institution after the images are domain transformed by Cycle GAN. 2) We examine whether the domain transformation by Cycle GAN is appropriately executed or not. 3) Cycle GAN re-train based on the results. The above 1), 2) and 3) are repeated to make the domain transformation suitable for the opacity classification.

2. Proposed Method

This study aims to classify opacities of diffuse lung diseases in chest CT images, where we also aim to diverse a classifier based on the domain transformation of semi-supervised Cycle GAN. In detail, we regard the conditions of the CT images taken at Osaka University Hospital (hereafter referred to as Domain O) as the standard domain and convert the CT images taken at Yamaguchi University Hospital (hereafter referred to as Domain Y) to the conditions of Domain O using semi-supervised Cycle GAN. This domain transformation from domain Y to domain O is defined as YtoO transformation.

The semi-supervised Cycle GAN used for YtoO transformation is a structure that combines a standard Cycle GAN with a trained classifier. The structure of the classifier is based on a Residual Network[4] (ResNet), and the ResNet is trained with domain O image data. ResNet classifies the opacity of a small number of domain Y images after the domain transformation YtoO by Cycle GAN has performed. Based on the classification result, Cycle GAN is re-trained.

The flow of the validation of the classification methods is shown in Fig. 1 and the methods for the validation are listed in Table 1. We measure classification accuracy averaged over 20 independent trials and apply the t-test between the proposed method and the other methods. After the validation in the case of YtoO, OtoY transformation is also verified.

2.1. *Creating ROI Images*

In this study, 503 chest CT images taken at Yamaguchi University Hospital and 636 chest CT images taken at Osaka University Hospital are used. Examples of the chest CT images taken at each hospital are shown in Fig. 2. The chest CT images of domains Y and O contain six opacities: consolidation (CON), diffuse nodular (DN), emphysema (EMP), ground-glass opacity (GGO), honeycombing (HCM), and normal (NOR). The CT images of domains Y and O are divided into 32×32 [pixels] regions of interest (ROIs), which are used for training the Cycle GAN. Each CT image has a mask image that shows the location of each opacity. The mask images were created by three radiologists. 32×32 [pixels] regions are scanned while striding from the upper left

Table 1. Condition of each method. We compare four different methods including the proposed method. The architecture of the proposed model is a combination of a standard Cycle GAN and a ResNet trained on the target domain: classifier trained at the other medical institution. "ResNet trained with source domain" represents ResNet trained with only a few teacher labels for semi-supervised learning.

Domain transformation method	Cycle GAN	With ResNet	
		Trained with source domain	Trained with target domain
No domain transformation			
Cycle GAN	✓		
Method for comparison	✓	✓	
Semi-supervised Cycle GAN (Proposed method)	✓		✓

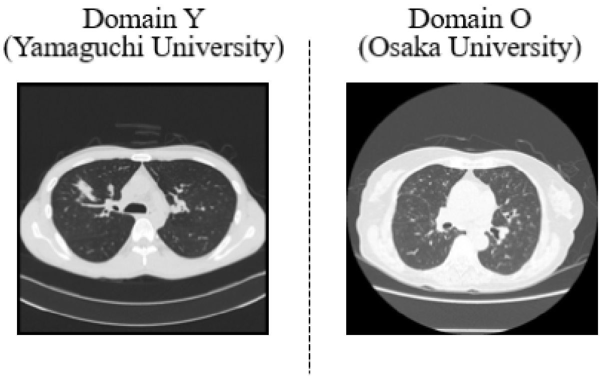


Fig. 2. Examples of chest CT images. The left is a chest CT image taken at Yamaguchi University Hospital, and the right is a one taken at Osaka University Hospital.

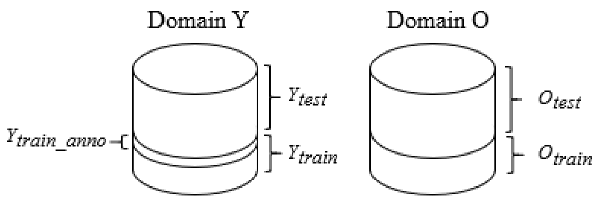


Fig. 3. Training data and testing data of domain Y and O. The created ROI images are split into a set of training data and test data. Y_{train} and O_{train} are used to train Cycle GAN, while Y_{test} and O_{test} are used to verify the effectiveness of the domain transformation. $Y_{train\_anno} \subset Y_{train}$ is a dataset of a small number of Y_{train} with class labels. Specifically, CT images of five patients are annotated for each opacity.

corner of the mask image to the lower right corner. If the regions contains more than 50% of the masked area indicating each opacity, the same areas are extracted from the corresponding CT images and used as ROI images.

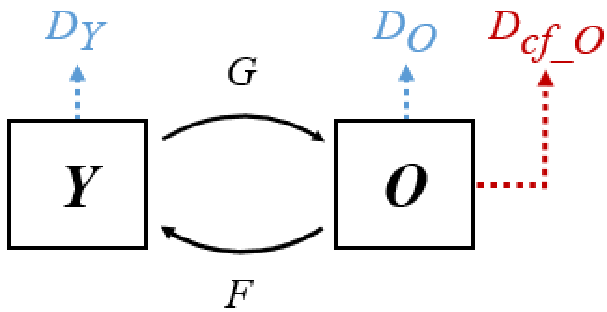


Fig. 4. Structure of semi-supervised Cycle GAN. Semi-supervised Cycle GAN contains a standard Cycle GAN consisting of two generators and two discriminators, and also contains a classifier D_{cf_o} that classifies the opacity. Discriminator D_O, which discriminates real data {o} and transformed data $G(y)$, learns to achieve high discrimination accuracy. In other words, D_O learns to maximize $log\left(1 - D_O\left(G(y)\right)\right)$ and $logD_O(o)$. While generator G, which converts data from domain Y to domain O, learns to minimize $log\left(1 - D_O\left(G(y)\right)\right)$. F and D_Y are trained in the same way as G and D_O. D_{cf_o} is a classifier that classifies the opacity $G(y)$ where D_{cf_o} is trained to achieve high classification accuracy.

2.2. *Domain Transformation by Semi-supervised Cycle GAN*

We apply domain transformation to the ROI images created in Section 2.1 using semi-supervised Cycle GAN. As shown in Fig. 3, Y_{train} and O_{train} are used to train Cycle GAN, and the domain transformation is performed on Y_{test} and O_{test}.

As shown in Fig. 4, the structure of the semi-supervised Cycle GAN consists of a combination of the standard Cycle GAN and a trained classifier. The purpose of this model is to learn domain transformation between two domains Y and O with given training samples $\{y_i\}_{i=1}^{N}$

where $y_i \in Y_{train}$ and $\{o_j\}_{j=1}^{M}$ where $o_j \in O_{train}$. N and M are the numbers of samples in set Y_{train} and O_{train}, respectively.

In the standard Cycle GAN, the loss functions Eq. (1) through (3) are used for training *G, F, D_Y, D_O*.

$$\begin{aligned} L_{GAN}(G, D_O, Y, O) &= E_{o\sim P_{data}(o)}[logD_O(o)] \\ &+E_{y\sim P_{data}(y)}\left[log\left(1 - D_O\big(G(y)\big)\right)\right] \\ L_{GAN}(F, D_Y, O, Y) &= E_{y\sim P_{data}(y)}[logD_Y(y)] \\ &+E_{o\sim P_{data}(o)}\left[log\left(1 - D_Y\big(F(o)\big)\right)\right] \end{aligned} \tag{1}$$

$$\begin{aligned} L_{cyc}(G, F) &= E_{y\sim P_{data}(y)}\left[\left\|F\big(G(y)\big) - y\right\|\right] \\ &+E_{o\sim P_{data}(o)}\left[\left\|G\big(F(o)\big) - o\right\|\right] \end{aligned} \tag{2}$$

$$\begin{aligned} L_{identity}(G, F) &= E_{o\sim P_{data}(o)}[\|G(o) - o\|] \\ &+E_{y\sim P_{data}(y)}[\|F(y) - y\|] \end{aligned} \tag{3}$$

We denote the data distribution as $y\sim p_{data}(y)$ and $o\sim p_{data}(o)$. The domain transformation is learned by minimizing the loss function of Eq. (1), but since this loss function alone will learn to map the same output pattern to any input images, the loss functions Eq. (2) and (3) are introduced. Eq. (2) is called Cycle Consistency Loss, which constrains the original data y and o to match the generated data $F\big(G(y)\big)$ and $G\big(F(o)\big)$, respectively. Eq. (3) is called Identity Mapping Loss, which constrains the generator not to convert any data that have belonged to the target domain.

In the semi-supervised Cycle GAN, we introduce a new constraint to re-train Cycle GAN using ResNet. First, we generate (fake) domain O data $G(y)$ from domain Y by Cycle GAN. ResNet D_{cf_o} trained on domain O is used to identify the data $G(y)$. The classification results are fed back to the Cycle GAN for re-train. In the re-training, only the ROI images $\left\{y_k^{(anno)}\right\}_{k=1}^{L}$ are used, where $y_k^{(anno)} \in Y_{train\_anno}$ and Y_{train_anno} contains annotated CT images of five patients per opacity class. L is the number of samples in dataset Y_{train_anno}. For learning G in the semi-supervised Cycle GAN, the loss function shown in Eq. (4) is used in addition to Eqs. (1) through (3).

$$\begin{aligned} &L_{opacity}\big(G, D_{cf\_O}\big) \\ &= E_{y^{(anno)}\sim P_{data}(y^{(anno)})}[logD_{cf\_o}(G(y^{(anno)}))] \end{aligned} \tag{4}$$

Our full loss function is:

Table 2. Stride width and the number of ROI images set for each opacity.

Opacity	The number of ROI images		Width of stride	
	Domain Y	Domain O	Domain Y	Domain O
CON	3071	3447	8	11
DN	3023	3311	16	14
EMP	3122	3021	24	27
GGO	3460	3273	12	18
HCM	3236	3434	13	13
NOR	3117	3035	29	32

$$\begin{aligned} L\big(G, F, D_A, D_B, D_{cf\_o}\big) = L_{GAN}&(G, D_B, A, B) \\ +L_{GAN}&(F, D_A, B, A) \\ +\lambda_1 L_{cyc}&(G, F) \\ +\lambda_2 L_{identity}&(G, F) \\ +\lambda_3 L_{opacity}&\big(G, D_{cf\_o}\big) \end{aligned} \tag{5}$$

where λ_1, λ_2, and λ_3 are bias terms, which are set to 40, 5, and 0.2, respectively. *G* and *F* are trained to satisfy the objective function expressed by Eq. (6).

$$G^*, F^* = arg \min_{G,F} \max_{D_Y\, D_O} L(G, F, D_Y, D_O, D_{cf\_o}) \tag{6}$$

YtoO transformation is realized by the above procedure.

3. Experimental Results

The effectiveness of the domain transformation YtoO and OtoY is verified using the trained ResNet. In YtoO transformation, after domain Y data are transformed fake O by YtoO transformation, they are input to ResNet trained on domain O, and six-class classification is performed. Then, we measure classification performance. The classification performance of the other methods shown in Table 1 is also examined to compare. We perform t-test on the F-measure of the proposed method and other methods to verify whether there are significant differences in terms of p-values.

3.1. *Creating Dataset*

If the stride width of each class is kept constant when generating ROI images, the number of ROI images of each class will be highly biased. Therefore, in this study, we adjusted the stride width for each opacity to keep the number of ROI images as around 3000 for each class. Table 2 shows the stride width and the number of ROI images for each opacity. The generated ROI images have opacity (class) labels based on the annotation by the radiologists.

Table 3. The number of ROI images used for training and testing. The number of training data for each opacity is about 1000, and that of testing data is about 2000. The same training data and testing data are used for both YtoO and OtoY transformation.

Opacity	Training data		Testing data	
	Domain Y	Domain O	Domain Y	Domain O
CON	1022 (95)*1	1027 (104)*2	2049	2420
DN	1018 (107)	1020 (98)	2005	2291
EMP	1020 (105)	962 (105)	2102	2059
GGO	989 (108)	996 (108)	2471	2277
HCM	1003 (96)	1021 (96)	2233	2413
NOR	1003 (105)	1024 (101)	2114	2011

*1 The values in parentheses are the number of annotated data used in the training of YtoO transformation.

*2 The values in parentheses are the number of annotated data used in the training of OtoY transformation.

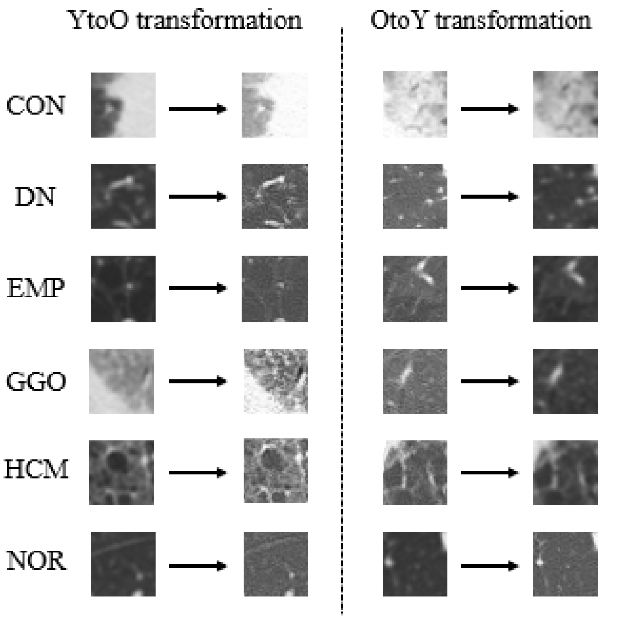


Fig. 5. Example of domain transformation using semi-supervised cycle GAN. The left side shows ROI images after YtoO transformation, and the right side show the one after OtoY transformation.

3.2. *Results of ResNet Training*

We trained ResNet using the generated ROI images. ResNet learned the ROI-based opacity classification for domains Y and O, respectively. We adopted ResNet34, where the original filter size 7 x 7 of the first convolution layer was changed to 3 x 3. Then we normalized the value range to [-1,1] as a preprocessing of ROI images used for the training of ResNet. The parameters for training were 20 epochs and mini-batch size 16, and we used Adam[5] as the optimization method. As a result of the training, the accuracy for the training data in domain Y was 98.8%, and that for domain O was 98.1%. The semi-supervised learning of the domain transformations is performed by a model that combines the trained ResNet with a standard Cycle GAN. In the case of YtoO transformation, the ResNet trained on domain O is incorporated into the Cycle GAN.

3.3. *Results of Domain Transformation by Semi-supervised Cycle GAN*

In this study, the number of residual blocks in the generator was set to 3 because the ROI image size input to the Cycle GAN is small, i.e., 32 x 32 [pixels]. Since the data are grayscale images, the number of input channel is one. Cycle GAN was trained using the training data shown in Table 3. The obtained images before and after the domain transformation using the semi-supervised Cycle GAN are shown in Fig. 5.

3.4. *Performance Evaluation*

We measure the classification performance of ResNet on the domain transformed ROI images. In the case of YtoO transformation, ResNet trained with domain O was used to classify the opacity of the transformed data from domain Y. For the performance evaluation, we used precision, recall, and F-measure. The results averaged over 20 independent trials for YtoO and OtoY transformation are shown in Table 4 and Table 5, respectively. The proposed method shows the best

Table 4. Precision, recall and F-measure obtained by each method and the results of t-test in the case of YtoO transformation. The values are the average over 20 trials. P-value is calculated by t-test on F-measure between the proposed method and other methods.

Method	Precision (%)	Recall (%)	F-measure	p-value
No domain transformation	36.5±8.6	28.6±6.4	0.228±7.8	1.83×10^{-17}
Cycle GAN	67.9±3.6	65.8±4.2	0.655±4.5	4.73×10^{-7}
Method of comparison	71.3±2.0	68.4±2.6	0.682±2.6	1.13×10^{-7}
Semi-supervised Cycle GAN	**74.0±1.0**	**72.7±0.9**	**0.727±0.9**	-

Table 5. Precision, recall and F-measure obtained by each method and the results of t-test in the case of OtoY transformation. The values are the average over 20 trials. P-value is calculated by t-test on F-measure between the proposed method and other methods.

Method	Precision (%)	Recall (%)	F-measure	p-value
No domain transformation	35.0±4.5	37.0±2.2	0.319±1.9	5.80×10^{-42}
Cycle GAN	75.2±1.6	73.4±1.6	0.730±2.0	6.00×10^{-3}
Method of comparison	73.4±2.7	70.1±5.9	0.692±6.4	1.13×10^{-3}
Semi-supervised Cycle GAN	**76.4±1.4**	**74.7±1.4**	**0.745±1.4**	-

performance for both YtoO and OtoY transformation. Besides, the standard deviation of the results obtained by the proposed method is the smallest, confirming that the classification performance is stable when the ResNet is diverted.

4. Conclusion

In this study, we proposed a semi-supervised Cycle GAN that combines a standard Cycle GAN with a trained classifier to perform domain transformation of chest CT images. In the experiment, we classified the normal opacity and five abnormal opacities of diffuse lung diseases in the CT images after the domain transformation and clarified the effectiveness of the proposed method.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP19K12120.

References

1. Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Geert Litjens, et al., Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks, *IEEE Transactions on Medical Imaging*, 2016, pp. 1160-1169
2. Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros, Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, *International Conference on Computer Vision*, 2017
3. Masashi Miyake, Shingo Mabu, Shoji Kido, Takashi Kuremoto, Yasushi Hirano, Domain Transformation of Chest CT Images Using Cycle GAN and Its Application to Classification Systems, *The 38th JAMIT Annual Meeting*, 2019, pp. 108-115
4. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep Residual Learning for Image Recognition, *IEEE Conference on Computer Vision and Pattern Recognition*, 2016
5. Diederik P. Kingma, Jimmy Lei Ba, ADAM: A method for stochastic optimization, *Proceedings of the 3rd International Conference on Learning Representations*, 2015

Anomaly detection of lung sounds using DAGMM

Ryosuke Wakamoto
Graduate School of Sciences and Technology for Innovation, Yamaguchi University,
2-16-1 Tokiwadai, Ube, Yamaguchi 755-8611, Japan

Shingo Mabu
Graduate School of Sciences and Technology for Innovation, Yamaguchi University,
2-16-1 Tokiwadai, Ube, Yamaguchi 755-8611, Japan

Shoji Kido
Graduate School of Medicine, Osaka University,
2-2 Yamadaoka, Suita, Osaka 565-0871, Japan

Takashi Kuremoto
Graduate School of Sciences and Technology for Innovation, Yamaguchi University,
2-16-1 Tokiwadai, Ube, Yamaguchi 755-8611, Japan
E-mail: b093vg@yamaguchi-u.ac.jp, mabu@yamaguchi-u.ac.jp
kido@radiol.med.osaka-u.ac.jp, wu@yamaguchi-u.ac.jp

Abstract

There are only a few small-scale benchmark datasets of lung sounds that are annotated for the training of machine learning. Therefore, we aim to build an anomaly detection system that only uses normal data that can be obtained more than abnormal data. We propose an algorithm that improves the Deep Autoencoding Gaussian Mixture Model (DAGMM), where various types of neural networks are applied to DAGMM as the compression networks. Experimental results show that the proposed methods obtain effective classification performance.

Keywords: lung sounds, deep learning, anomaly detection, DAGMM

1. Introduction

In recent years, deep learning algorithms based on large amounts of data have been proposed and applied in various fields[1]. In the field of image recognition, in particular, various applications have been made using convolutional neural networks (CNNs). Because of its availability, CNNs have been applied to the medical field where high accuracy is required.

For example, various approaches have been developed for medical images, such as feature extraction specific to a lesion or 3D modeling of an organ. Most of the research on computer-aided diagnosis using deep learning has been attempted for image diagnosis, which has greatly contributed to improve the accuracy of medical image diagnosis such as lung diseases[2] which are dealt with in this study. However, diagnosis with only images is not sufficient because, in addition to the limitations of qualitative diagnosis in imaging, there are multiple ways to diagnose lung diseases. In the diagnosis of lung diseases, the patient is interviewed and physically examined, and the condition of the disease is estimated to some extent, and respiratory function tests, image diagnosis, electrocardiography, and gait assessment are performed[3]. Because of this, the burden on the physicians making the diagnosis is heavy, requiring precise treatment and enough experiences.

As mentioned above, because doctors need to conduct various tests for diagnosis, it is necessary to develop diagnostic support techniques for not only visual information such as images, but also for other senses such as hearing. Therefore, we consider applying deep learning-based speech recognition to auscultation which is one of the diagnostic methods for lung diseases. If we can construct a system for discriminating sounds, we can improve the accuracy of diagnosis by combining it with other examination results such as imaging.

Although deep learning requires a large amount of sound data to learn the features of lung sounds, there are few small benchmark datasets with annotation (class labeling), which is necessary for machine learning. In addition, it takes a great deal of effort and time for each medical facility to acquire a sufficient amount of sound data. When training is performed with a small amount of data, it may not obtain generalized performance because it is difficult to capture essential features that are important for the diagnosis due to the large individual differences of lung sounds. There have been some research related to lung sounds classification, such as the analysis using histogram statistics[4] and the lung sounds classification using deep learning[5], however, the number of data to generate classification models is limited in both methods. Therefore, this study focuses on normal data, which is easier to obtain than abnormal data, and constructs a system for detecting abnormalities of pulmonary auscultation sounds using deep learning with the aim of capturing generalized features even with small data.

In this paper, we propose a deep learning algorithm for feature extraction of lung sounds that improves the Deep Autoencoding Gaussian Mixture Model (DAGMM)[6], which is an anomaly detection algorithm capable of learning feature extraction and clustering at the same time. Specifically, it uses Mel-Frequency Cepstral Coefficients (MFCC)[7] for the feature extraction and three types of networks (auto-encoders) based on: 1) CNN, 2) Long short term memory (LSTM), 3) Convolutional LSTM (C-LSTM) are applied to DAGMM as compression networks to design an algorithm for efficient feature extraction. C-LSTM has features of both CNN and LSTM, so it can consider the features of both peripheral and time series information. These three types of networks have been shown to be useful in discriminating lung sounds in previous studies, and since the standard Multi-Layer Perceptron (MLP) is used as a compression network in the conventional DAGMM, the above three networks are expected to show better performance in DAGMM. In our experiments, we compare the conventional DAGMM and the proposed methods, and better classification performance is obtained by improving the compression networks in DAGMM.

The paper is organized as follows. Section 2 describes the related work and Section 3 presents the proposed method. Section 4 presents the experimental conditions and results, and Section 5 presents conclusions and future work.

2. Related work

There are two types of research related to sound and speech: speech analysis and speech recognition. Speech analysis is a process of extracting features from the sampled sound data for speech recognition. Speech recognition is the process of identifying the target sound or speech based on the features obtained by the speech analysis. In addition to these processes, this section also describes the anomaly detection.

2.1. *Speech Analysis*

Speech analysis improves recognition accuracy by eliminating unnecessary information as noise as well as extracting only the features necessary for speech recognition. In this study, we use MFCC, a feature extraction method to extract the cepstrum by applying a filter based on the human hearing (Mel-filter bank), which separates the formant frequencies, which is necessary for speech recognition, from the pitch frequencies, which contains individual differences. Since Formant frequencies of MFCC appear in the low-frequency regions, important information for sound classification can be obtained by extracting the lower dimensional components of MFCC. Fig. 1 shows an example of MFCC that shows the 20-dimensional features of MFCC for 5-second of lung sounds. The horizontal axis of Fig. 1 represents the 20 segments of the 5-second data and the vertical axis represents the dimensions of MFCC.

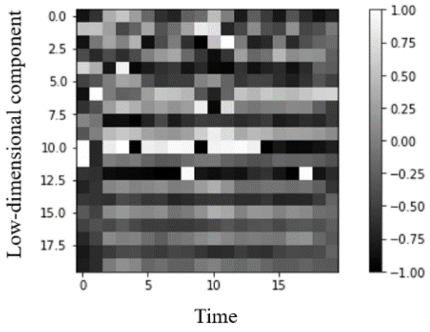


Fig. 1. MFCC with 20 dimensions

2.2. *Speech Recognition*

Before deep learning is actively studied, dynamic programming (DP) matching[8] and hidden Markov models (HMM)[9] were used for speech recognition, however, once deep learning was applied to speech recognition, Recurrent Neural Network (RNN) were applied and false recognition rate of speech was significantly decreased. However, it was difficult to learn long data stably due to the problem of gradient loss. Then, LSTM shown in Fig. 2 were used. LSTM has three gates: an input gate (i. in Fig. 2), a forget gate (ii. in Fig. 2), an output gate (iii. in Fig. 2), and a cell that holds past information. The problem of gradient loss is solved by using gates to select the information in the cells.

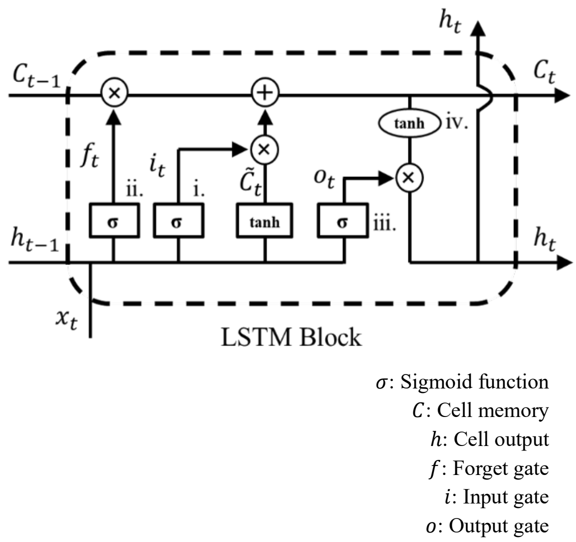


Fig. 2. Structure of LSTM

In addition, image recognition methods such as CNNs are often utilized in speech recognition[10]. CNNs are widely used in image recognition, but they also have excellent performance as a local feature extractor in speech recognition. Both CNNs and LSTMs are useful for feature extraction of lung sounds, and in the previous study, classification models using C-LSTM that possesses both CNN and LSTM features have shown higher accuracy[11].

2.3. *Anomaly Detection*

Anomaly detection is a method that creates a distribution of normal data by learning only normal data, then, regards the data that do not belong to the distribution of normal as abnormal data. In this section, we describe the outlier (anomaly) detection methods using machine learning. In general, unsupervised outlier detection methods execute feature extraction and distribution creation separately. For example, auto-encoder is first used to extract features from the input data, then clustering methods such as k-means and gaussian mixture model (GMM) are applied to create distributions. One of the problems in such learning structures is that the learning of feature extraction and clustering are executed separately. The features extracted independently from the clustering process may not be useful. In order to accurately detect outliers, a coherent feature extraction is required for clustering.

The structure of DAGMM is shown in Fig. 3. DAGMM is an outlier detection method that performs feature extraction and clustering simultaneously. The feature extraction is realized by a compression network using auto-encoder and the clustering is realized by combining an estimation network and GMM. The compression network outputs the encoded feature and the reconstructed data of the inputs, and the estimation network outputs the probability that the input x belongs to each cluster. In the compression network, input x is encoded to feature Z_c. Furthermore, the feature Z_c is decoded to obtain the reconstructed image x'. In the compression network, auto-encoder learns that x and x' become the same, where the reconstruction error Z_r is represented as follows.

$$Z_r = (d_1, d_2) = \left(\frac{\|x-x'\|_2}{\|x\|_2}, \frac{x \cdot x'}{\|x\|_2\|x'\|_2}\right). \quad (1)$$

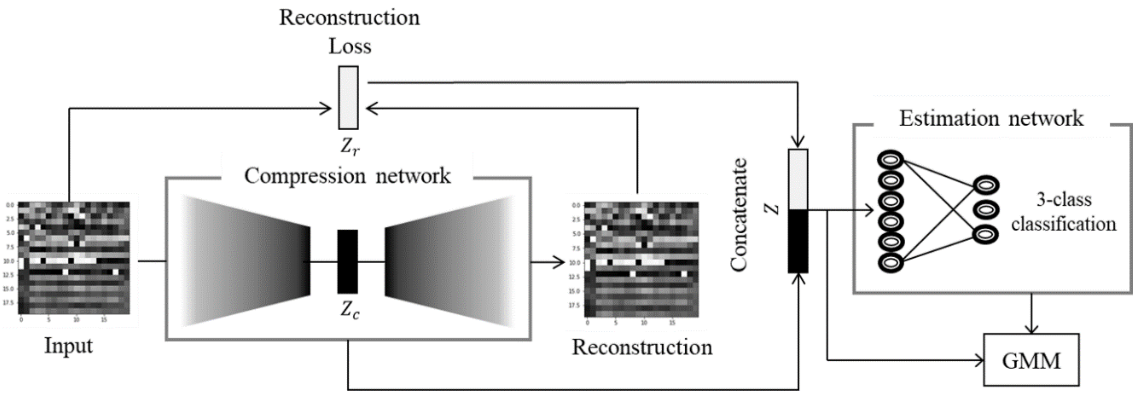


Fig. 3. Structure of DAGMM

By using the two measures d_1 and d_2, we can measure the reconstruction error from different perspectives. Then, the concatenated feature of Z_c and error Z_r generated by the compression network forms a new feature Z. With this as the input to the estimation network, the estimation network then outputs the affiliation probability $\hat{\gamma}$ of how well Z is matched to each cluster. Then, GMM is generated by the feature Z and the affiliation probability $\hat{\gamma}$. GMM requires three parameters: the mixture ratio phi, the mean matrix mu, and the covariance matrix $sigma$. The formula for each is as follows.

$$phi_k = \sum_{i=1}^{N} \frac{\hat{\gamma}_{ik}}{N}. \quad (2)$$

$$mu_k = \frac{\sum_{i=1}^{N} \hat{\gamma}_{ik} z_i}{\sum_{i=1}^{N} \hat{\gamma}_{ik}}. \quad (3)$$

$$sigma_k = \frac{\sum_{i=1}^{N} \hat{\gamma}_{ik} (z_i - mu_k)(z_i - mu_k)^T}{\sum_{i=1}^{N} \hat{\gamma}_{ik}}. \quad (4)$$

DAGMM calculates the energy of the feature Z after the generation of the GMM: when a data is located in the center of the distribution of the GMM, the energy of Z is small, and the energy of the off-centered Z is large. The energy function is represented as follows.

$$E_{(Z)} = -log\left(\sum_{k=1}^{K} phi_k \frac{exp\left(-\frac{1}{2}(z-mu_k)^T \sum_k^{-1}(z-mu_k)\right)}{\sqrt{|2\pi sigma_k|}}\right). \quad (5)$$

DAGMM can learn feature extraction suitable for clustering by the above learning structure. In addition, when outlier detection is performed, the energy of abnormal data is expected to be larger than that of normal data.

3. Proposed methods

In this study, we deal with an anomaly detection algorithm for lung sounds. Since previous studies in lung sounds have been concerned with the lack of data, an anomaly detection algorithm is suitable because normal data is easier to obtain than abnormal data and obtains generalized features even with a small number of data. In this study, the conventional DAGMM is regarded as the conventional method that realizes anomaly detection performing both clustering and feature extraction simultaneously.

From the next subsection, we propose a deep learning algorithm that improves the feature extraction of DAGMM for lung sounds. Specifically, we replace the compression network in DAGMM in Fig. 3 with various types of auto-encoders based on CNN, LSTM and C-LSTM. Since the three auto-encoders based on CNN, LSTM and C-LSTM have different structures and characteristics, the details of each structure are described one by one.

3.1. *Convolutional auto-encoder (CAE)*

The structure of a CAE is shown in Fig. 4. The input data is converted to an image as shown in Fig. 1 by applying MFCC to 5-second unlabeled sound data. Therefore, it is possible to perform convolutional processing on sound data as the same way as on images. CAE encodes the input by an Encoder consisting of two convolutional and two pooling layers, and then reconstructs the input by a Decoder consisting of two deconvolutional and two upsampling layers. The structure of the compression network with convolution and pooling captures the local features of MFCC making it easier to ignore the noise that is likely to have an influence on speech recognition.

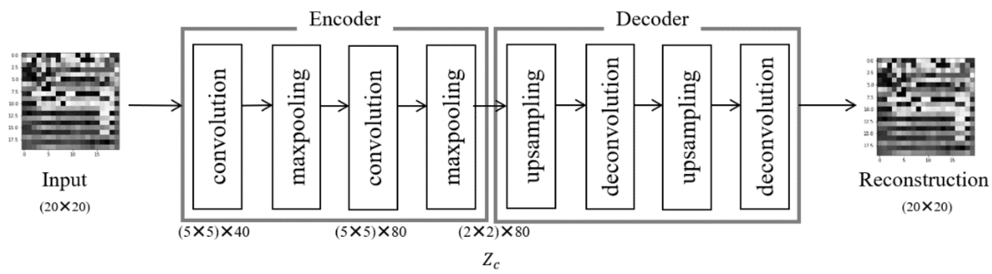


Fig. 4. Structure of CAE

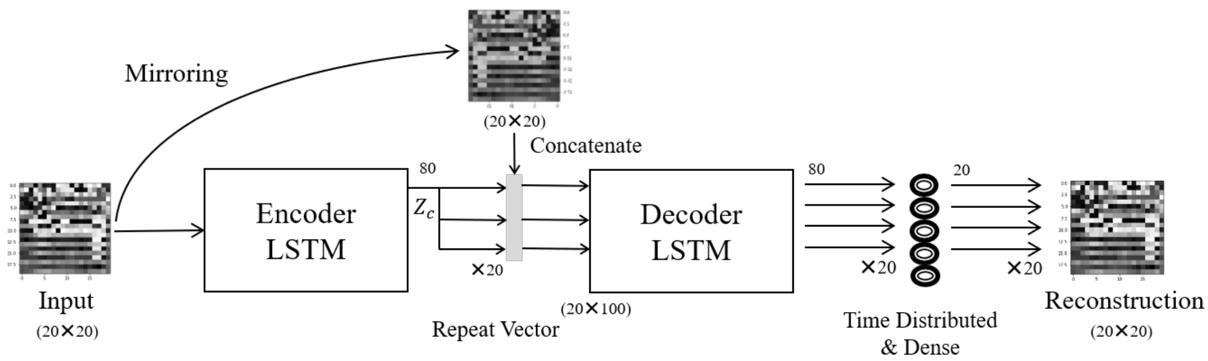


Fig. 5. Structure of LSTM-AE

3.2. *LSTM-based auto-encoder (LSTM-AE)*

The structure of a LSTM-AE is shown in Fig. 5 The LSTM-AE dealt with in this study is based on the structure proposed in the literature[12]. As in the case of CNN, we use unlabeled data to train Encoder LSTM and Decoder LSTM. In this case, the details of the network structure are as follows. First, Encoder LSTM outputs 80 features from the hidden layer of LSTM. Then, the input data to the Encoder LSTM (20×20) is mirrored and added to the generated features. This process is carried out by duplicating the generated one-dimensional features (80) 20 times to form a two-dimensional features (20×80) as shown in Fig. 5, and the generated two-dimensional features is combined with the mirroring input (20×20). This yields a new feature (20×100). Here, we explain the meaning of combining the generated features and mirroring input. Since the features (80) obtained by the Encoder LSTM has lost the information on the time axis, it is difficult to reconstruct the input image at the output layer. Therefore, the mirroring input serves as a flag for reconstructing the input from the features, and also has the effect of giving information on the inverse time axis. The combined features described above (20×100) are input to the Decoder LSTM row by row in time order (1×100), and the Decoder LSTM outputs the features (80) corresponding to each time. The output of one-dimensional features (80) is converted to a size (20) equivalent to the number of columns, i.e., 20 time segments of the input data, where the conversion is implemented by MLP. Here, the features (20) generated at all the time (20) are joined in sequence to obtain data with the same size as the input data (20×20). The purpose of an auto-encoder using LSTM is to make the output be the same as the input.

3.3. *C-LSTM-based auto-encoder (C-LSTM-AE)*

C-LSTM is a neural network that uses convolution layers when inputs to the sigmoid function of each gate of LSTM (i., ii., and iii. in Fig. 2) are calculated, and when the input to the tanh layer (iv. in Fig. 2) is calculated. Thus, LSTM can implement data processing considering peripheral information. The output shape is the same as the CNN, which is (the size of the feature map after

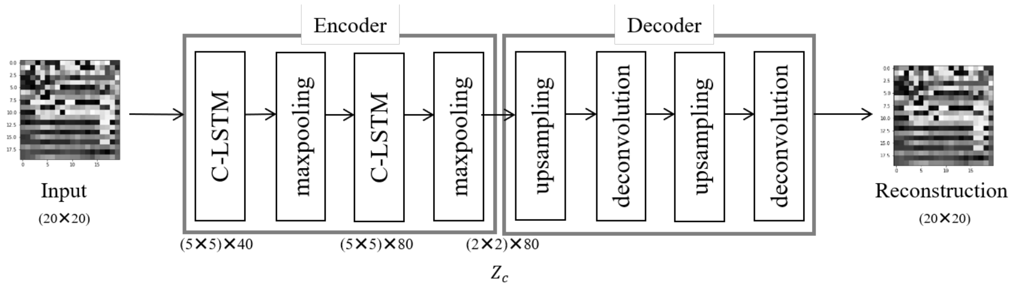


Fig. 6. Structure of C-LSTM-AE with convolutional decoder

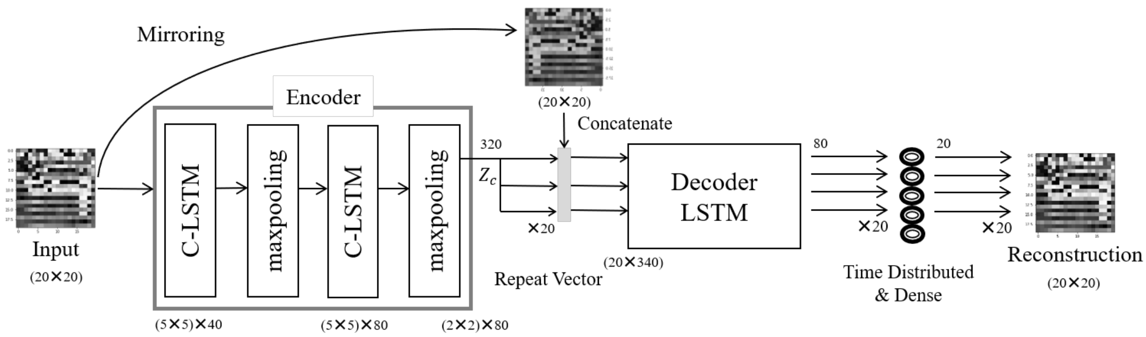


Fig. 7. Structure of C-LSTM-AE with LSTM decoder

convolution) x (the number of feature maps). We propose two structures of C-LSTM-AEs: the first is the C-LSTM-AE with convolutional decoder and the second is the C-LSTM-AE with LSTM decoder. The former structure is shown in Fig. 6 and the latter structure is shown in Fig. 7. The difference between the two is the structures of the decoders, which is described in detail in the next sub-subsection.

3.3.1. Convolutional decoder

In Fig. 6, the decoder consists of two deconvolutional layers and two upsampling layers, which is similar to the decoder used in the CAE. As the feature map of C-LSTM is two-dimension like CNN, the network structure is similar to that of CAE using deconvolution layers in the decoder. This structure is expected to capture local features while taking into account time-series information.

3.3.2. LSTM-Decoder

In Fig. 7, the decoder has the same structure as that used in the LSTM-AE, which converts the features (2×2×80) obtained by the C-LSTM's encoder into a one-dimensional features (320). In addition, the mirroring input is generated for combining the one-dimensional features and the mirroring input. Then, Decoder LSTM outputs a one-dimensional features (80) every time step, and it is converted by MLP into the array with the size (20) which is the same as the number of columns, i.e., 20 time segments of the input data. Therefore, the shape of the final output becomes the same as the input. Since we use C-LSTM and pooling layers in the Encoder, it is expected to learn the features considering time series information while mitigating the effects of noises in the input.

4. Experiments and results

4.1. *Data overview*

In this experiment, we used lung sounds data provided by Yamaguchi University Hospital, Japan to discriminate between discontinuous rale[3] (abnormal sounds) and normal sounds. The data used in this study are 5-second data, which were cut out from the data recorded during auscultation in a private room with the subjects in a seated position, and judged to be normal, fine crackle and coase crackle by the doctor. Here, the 12 areas were used for auscultation, that is, the chest area was firstly divided into four areas: upper and lower portions of both left and right lungs, respectively, and furthermore, the three areas : anterior, lateral, and posterior portions of the lungs are considered. Therefore, auscultation is done at totally 12 areas of lung, and the lung sounds of each area were recorded for more than three respiratory phases (inhalation plus exhalation) and more than about 15 seconds. Although the data were acquired from the three areas in the front, side, and back, the number of channels was 1 because the diagnosis was made independently from each area. The sampling rate was set to 11 kHz and the digital stethoscope (Power stethoscope, Starkey Japan) was connected to a voice recorder (ICD-MS1, Sony) and the sounds were recorded in a flash memory in 16-bit WAV file format. It is noted that noises caused during data acquisition depend on the way of breathing, subtle shifts in the stethoscope's position, and differences in the loudness of the sound. If an user is not skilled in handling the stethoscope, a lot of noises are picked up. In this study, the doctor who has enough experiences of auscultation recoded the lung sounds. The discontinuous rale in this study are discontinuous rales of short duration, and can be classified into two types: fine crackle and coarse crackle. Table 1 shows the characteristics of each auscultation sound including the normal sound. The characteristics of the abnormal intermittent sounds show sudden appearance, high sound pressure level, and very short duration. In addition to the paucity of data, the frequency bands of normal and abnormal sounds are overlapped, and the individual differences are large, thus, frequency analysis is difficult for the classification. In this study, to deal with the anomaly detection algorithm, two types of discontinuous rale are grouped together as one abnormal class.

Table 1. Characteristics of lung sounds[4]

Class	Abnormal		Normal
	Coarse	Fine	
Frequency [Hz]	250-500	200-500 700-1000	150-600
Duration [ms]	10-15	Less than 5	-
Estimated diseases	Bronchitis, Pneumonia, Pulmonary tuberculosis	Interstitial pneumonia, Pulmonary fibrosis	-

Table 2. Overview of data of each class

Class	Abnormal	Normal
Number of data	79	140
Number of patients (male, female)	24 (16, 6)	12 (12, 0)

*Abnormal includes two of unknown sex patients.

4.2. *Experimental conditions*

The anomaly detection method was applied to two classes of auscultation sounds: abnormal and normal. The number of data of each class, the number of patients including their sex are shown in Table 2. 130 of the normal data were extracted as training data, and 10 normal data and 79 abnormal data that were not included in the training data were extracted as test data. In this study, we conducted an experiment with 14-fold cross validation. The evaluation index is Area under the Curve (AUC) which is area under the ROC curve, and the thresholds of the classification boundary of normal and abnormal when calculating ROC is determined by dividing the energy range of normal data into 11 ranges. In order to achieve high performance with this evaluation method, it is necessary to give lower energy to normal data and higher energy to abnormal data. In other words, it is necessary to obtain generalized features of normal data in training.

For the input data, pre-processing of data sampling (sampling frequency of 2000 Hz), calculation of MFCC, and normalization (mean 0, variance 1) were applied prior to the training of each method. Because lung sounds are diagnosed mostly at frequencies below 2000 Hz[13] clinically, the sampling rate was set at 2000 Hz.

4.3. *Results*

Table 3 shows the AUC scores obtained by the conventional and proposed methods after 14-fold cross validation. From Table 3, we can see that the mean AUC

Table 3. AUC scores of each method obtained by 14-fold cross validation

Folds	Methods				
	DAGMM	DAGMM with CAE	DAGMM with LSTM-AE	DAGMM with C-LSTM-AE (conv)	DAGMM with C-LSTM-AE (LSTM)
1	0.9316	0.9791	0.8949	0.9911	**0.9924**
2	0.8665	0.8722	0.8930	0.8962	**0.9038**
3	0.9386	0.9633	0.9032	0.9741	**0.9766**
4	0.8633	0.8399	0.7241	0.8437	**0.8918**
5	0.8506	0.9101	0.6987	**0.9380**	0.9323
6	0.7949	0.8310	0.7633	0.8715	**0.8892**
7	0.8519	0.8696	0.8886	0.8930	**0.8949**
8	0.8962	0.9557	0.9570	**0.9899**	0.9861
9	0.9665	0.9804	0.9873	0.9880	**0.9930**
10	0.9146	**0.9867**	0.9608	0.9684	0.9633
11	0.8842	0.9006	0.9386	0.9342	**0.9462**
12	0.8709	0.9595	0.8842	**0.9728**	0.9646
13	0.8506	0.9177	0.8842	0.9449	**0.9532**
14	0.7411	0.9044	0.8823	**0.9399**	0.9272
Mean	0.8730	0.9193	0.8757	0.9390	**0.9439**

(conv): convolutional decoder, (LSTM): LSTM decoder

Table 4. Standard deviation of each methods by 14-folds cross validation

Method	DAGMM	DAGMM with CAE	DAGMM with LSTM-AE	DAGMM with C-LSTM-AE (conv)	DAGMM with C-LSTM-AE (LSTM)
Standard deviation	0.0560	0.0508	0.0841	0.0452	**0.0365**

score (0.9439) obtained by C-LSTM-AE with LSTM decoder is the best. The standard deviation obtained by each method is shown in Table 4, where C-LSTM-AE with LSTM decoder is also show the lowest standard deviation, therefore, the AUC scores are not deviated by a wide margin compared to other models in each validation.

Here, we discuss the usefulness of the proposed method. While the average AUC score of the conventional DAGMM without the proposed method is 0.8730, the average AUC scores of all the proposed methods are higher than that of the conventional DAGMM.

The results of the t-test of all methods are shown in Table 5. We can see from Table 5 that there is significant differences between the proposed methods and the conventional DAGMM, except for LSTM-AE, where the significance level is 5% (<0.05). It can be said that the performance of DAGMM can be improved by improving the compression network. Since the difference is not significant for LSTM but significant for the other proposed methods, it can be said that the peripheral information considered by convolution layers and the mitigation of the influence of noises by the pooling layers are related to the superiority of the performance in the feature extraction of lung sounds. When comparing C-LSTM-AE(conv) and LSTM-AE, C-LSTM-AE shows the better result with a significant difference, thus, the combination of both the peripheral information obtained by convolution and the time series information obtained by LSTM is useful in the feature extraction of lung sounds. Although time series information includes the important features in the lung sounds, it is susceptible to noises, thus it is necessary to use convolution and pooling together. When comparing C-LSTM-AE(LSTM) and C-LSTM-AE(conv), LSTM decoder is better because it recovers time series information more accurately.

Table 5. P-values of the results of t-test obtained by each method

	DAGMM	DAGMM with CAE	DAGMM with LSTM-AE	DAGMM with C-LSTM-AE (conv)	DAGMM with C-LSTM-AE (LSTM)
DAGMM	-	1.06×10^{-3}	4.48×10^{-1}	1.58×10^{-4}	1.43×10^{-5}
DAGMM with CAE	-	-	1.43×10^{-2}	2.29×10^{-4}	3.39×10^{-4}
DAGMM with LSTM-AE	-	-	-	1.76×10^{-3}	1.41×10^{-3}
DAGMM with C-LSTM-AE (conv)	-	-	-	-	1.20×10^{-1}
DAGMM with C-LSTM-AE (LSTM)	-	-	-	-	-

5. Conclusions

In this paper, various types of compression networks are proposed and evaluated by the AUC scores. The experimental results showed that utilizing convolution and pooling was effective in learning lung sounds, and utilizing time-series information together improved the performance of the model. In the future, we would like to compare these methods with other anomaly detection methods and improve the generalization abilities.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP19K12120.

References

1. Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature* **521**(7553), 2015, pp. 436-444
2. M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network, *IEEE Transactions on Medical Imaging* **35**(5), 2016, pp. 1207-1216
3. T. Ishihara, T. Kawashiro, T. Abe, K. Kikumaru, and M. Yonemaru, Auditory training by CD, Nankodo, 1993 (in Japanese)
4. Y. Igami, H. Shouno, and S. Kido, Identification of pulmonary auscultation data using statistics of adjacent waveform intervals, *The 9th IEEE Hiroshima Student Symposium (HISS2007)* **B53**, 2007 (in Japanese)
5. D. Bardou, K. Zhang, and S. M. Ahmad, Lung sounds classification using convolutional neural networks, *Artificial Intelligence in Medicine* **88**, 2018, pp. 58-69
6. B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, H. Chen, Deep autoencoding gaussian mixture model for unsupervised anomaly detection, *International Conference on Learning Representations*, 2018
7. B. Logan, Mel Frequency Cepstral Coefficients for Music Modeling, *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)* , 2000
8. H. Sakoe, Two-level DP-matching-A dynamic programming-based pattern matching algorithm for connected word recognition, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **27**(6), 1979, pp. 588-595
9. A. P. Varga and R. K. Moore, Hidden Markov model decomposition of speech and noise, *Proceedings of the 1990 International Conference on Acoustics, Speech, and Signal Processing* **2**, 1990, pp. 845-848
10. R. Narasimhan, X. Z. Fern, and R. Raich, Simultaneous segmentation and classification of bird song using CNN, *Proceedings of the 2017 International Conference on Acoustics, Speech, and Signal Processing*, 2017, pp. 146-150
11. R. Wakamoto, S. Mabu, S. Kido, and T. Kuremoto, Lung Sound Classification Using Deep Neural Networks with Pre-training - Comparison of the Performance between CNN, LSTM and Convolutional LSTM -, *IEEJ Transactions on Electronics, Information and Systems* **140**(12), 2020, pp. 1402-1409 (in Japanese)
12. N. Srivastava, E. Mansimov, and R. Salakhutdinov, Unsupervised learning of video representations using LSTMs, *Proceedings of the 32nd International Conference on Machine Learning* **37**, 2015, pp. 843-852
13. V. Gross, A. Dittmar, T. Penzel, F. Schüttler and P. V. Wichert, The Relationship between Normal Lung Sounds, *American Journal of Respiratory and Critical Care Medicine* **162**(3), 2000, pp. 905-909

A ground reaction force analysis in walking and running gaits in horse leg model on viscoelastic hoof-ground contact

Dondogjamts Batbaatar
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu 808-0196, Japan

Hiroaki Wagatsuma
Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu 808-0196, Japan
RIKEN Center for Brain Science, Saitama, Japan
E-mail: batbaatar.dondogjamts581@mail.kyutech.jp, waga@brain.kyutech.ac.jp
www.kyutech.ac.jp

Abstract

In this paper, we focused on the hoof-ground interaction in the simplified horse leg model because walking and running gaits are known to be different in trajectories; however, the force analysis still remains as unsolved issues. The computational experiments in Matlab, elastic and inelastic impact with the ground was resolved by using the dissipative contact force model and the ground reaction force was clearly examined in four different conditions from the combination of walking/running and elastic/inelastic contact.

Keywords: Closed-loop linkage; Multibody dynamics; Dissipative contact force model; hoof-ground interaction.

1. Introduction

In the human gait analysis, three different foot-rocker mechanisms are known as heel rocker, ankle rocker and forefoot rocker, which are generated from pendulum dynamics with three fulcrum points of heel, ankle and toe [1]. In a gait cycle, there are two phases as swing stance phases and particularly in the human gait analysis, the stance phase is decomposed to the initial contact for touching with the ground from the heel, the loading response for shifting of the center of mass from back to front, the mid-stance for lifting the heel up, the terminal stance for transition to the swing phase, and then pre-swing [2]. Such as bipedal gait cycle can be simplified to a mechanical system based on a coupled pendulum as demonstrated by McGeer (1990) [3], called a passive dynamic walking robot without any electromechanical actuator to mimic three rockers. In this case, the compliant human leg dynamics has been studied with a series of contact points on the rolling surface [4-6] and with the effective roll-over geometry of leg motion [7-9]. The fact suggests that the utilization of the simple skeleton model provides a large benefit to compensate the limitation of experimental measurements in the gain analysis to realize the actual phenomenon of the ground reaction force, because there are evidences in the walking condition [10,11], such as demonstrated two peaks in the temporal evolution of the ground reaction force during the stance phase, while there is a lack of evidences in other gait patterns. In robotic applications inspired from biological mechanisms, an articulated leg with hoof was inspired from the horse [12] and a linkage mechanism to mimic the horse leg trajectory by Batbaatar & Wagatsuma (2019) [13] and similar approaches were found in closed-loop linkage mechanisms [14,15]. In the animal locomotion, there are various gait patterns known as walk, trot, gallop and so on [16]. Therefore, the establishment of the detail analytical method for the ground reaction force is highly important to elucidate the principle to differ gaits in the sense of the energy consumption. Focusing on the ground contact phenomenon, the hoof part of the horse leg mechanism can be modelled in the form of the triangle structure to connect three representative points as the toe, heel and ankle, which reproduces the rocker mechanisms. If it is possible to provide kinematic/kinetics of the hoof-ground contact, the viscoelastic effect is also considerable to analyse how much the leg absorbs the impact force when touching with the ground. It realizes the detail analysis of the amount of the energy storage discussed in the distal leg movement and landing motions [17,18]. The purpose of the present study is to establish the basement analytical method for the ground reaction forces in gaits and demonstrate the efficacy of the method in the simplified leg model to focus on the hoof-ground interaction. This paper is divided into following sections. Section 2 introduces the general Multibody dynamics (MBD) formulation applied to the proposed leg mechanism as well as the viscoelastic contact force introduced in order to simulate realistic hoof-ground contact. Section 3 contains the kinematic and kinetic analysis focusing on the hoof-ground interaction, while Section 4 summarizes the main results and discusses the potential improvements and limitation of the study.

2. Methods

2.1. *Multibody system formulation for horse leg mechanism*

Kinematics and dynamic analysis of horse leg mechanism with multibody dynamics was presented in the article by Batbaatar & Wagatsuma (2019) [13] in which simplified model of leg mechanism generated a flexible coordinated movement of hoof with constraint closed-loop linkage structure. Furthermore, in applying appropriate driver constraint in system as principle movement strategy in running animal functional and animal-like end trajectories were generated and evaluated with duty factor which defines the specific locomotor behavior with respect to the stance and swing timing in one locomotive cycle. Mathematical model for horse leg mechanism can be stated by using multibody dynamics (MBD), the vector $\boldsymbol{q}$ with 42 elements including position and orientation of each body in leg mechanism expressed as generalized coordinates can be written

$$\boldsymbol{q} = \left[q_1^T, q_2^T, q_3^T, \ldots, q_{14}^T \right]^T. \tag{1}$$

The vector of q contains 42 (= 14 × 3) elements which are 14 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. A set of kinematic algebraic constraint equation according to the given initial configuration can be written

$$\boldsymbol{\Phi}(q,t) = \begin{bmatrix} \Phi^K(q) \\ \Phi^D(q,t) \end{bmatrix}_{42\times 1} = 0. \tag{2}$$

where the first 41 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 element $\Phi^D(q,t)$ defines the driver constraint (Crankshaft) of the proposed leg mechanism. The partial derivative of kinematic constraint equation respect to the generalized absolute Cartesian coordinates $\boldsymbol{q}$ is Jacobian matrix $\boldsymbol{\Phi}_q$ is obtained as

$$\boldsymbol{\Phi}_q = \left[\frac{\partial \boldsymbol{\Phi}(\boldsymbol{q},t)}{\partial \boldsymbol{q}} \right]_{42\times 42}, \tag{3}$$

where it allows us to investigate placement, velocity and acceleration analyses kinematically. The forward dynamics analysis introduces the mass matrix $\mathbf{M} = (42\times 42)$, and the generalized external force vector $\mathbf{h^{(a)}} = (42\times 1)$, as follows:

$$\mathbf{M} = diag(M_1, M_2, \ldots, M_{14}), \tag{4}$$

$$\{\mathbf{M}_i = [m_i, m_i, J_i]^T \mid i = 1,2,\ldots,14\}, \tag{5}$$

$$\mathbf{h^{(a)}} = \left[h_1^{(a)^T}, h_2^{(a)^T}, \ldots, h_{14}^{(a)^T} \right]^T \tag{6}$$

$$\{\mathbf{h_i^{(a)}} = [0, -m_i g, 0]^T \mid i = 1,2,\ldots,14\}, \tag{7}$$

where m_i is the mass of rigid link to point i, $J_i = 2l_i/3$ 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

$$\begin{bmatrix} \boldsymbol{M} & \boldsymbol{\Phi}_q^T \\ \boldsymbol{\Phi}_q & \mathbf{0} \end{bmatrix} \begin{bmatrix} \ddot{\boldsymbol{q}} \\ \lambda \end{bmatrix} = \begin{bmatrix} \boldsymbol{h}^{(a)} \\ \gamma - 2\alpha\dot{\boldsymbol{\Phi}} - \beta^2 \boldsymbol{\Phi} \end{bmatrix}. \tag{8}$$

where $\ddot{q}$ is the acceleration vector, $h^{(a)}$ denotes the generalized external force vector, λ is the Lagrange multipliers with vector $\boldsymbol{\Phi}_q\ddot{\boldsymbol{q}} = -\left(\boldsymbol{\Phi}_q\dot{\boldsymbol{q}}\right)_q\dot{\boldsymbol{q}} - 2\boldsymbol{\Phi}_{qt}\dot{\boldsymbol{q}} - \boldsymbol{\Phi}_{tt} = \gamma$. The parameter α and β are the feedback compensator in Baumgarte stabilization method [19] used to guarantee accuracy of the numerical solution. In the forward dynamic analysis, new coordinates and velocities require two arrays for $\dot{\boldsymbol{q}}$ and $\ddot{\boldsymbol{q}}$ for the time step $t+\Delta t$ as

$$\boldsymbol{u} = \begin{bmatrix}\boldsymbol{q}\\ \dot{\boldsymbol{q}}\end{bmatrix},\quad \dot{\boldsymbol{u}} = \begin{bmatrix}\dot{\boldsymbol{q}}\\ \ddot{\boldsymbol{q}}\end{bmatrix},\quad \dot{\boldsymbol{u}}\xrightarrow{yields}\boldsymbol{u}\left(t+\Delta t\right), \tag{9}$$

At the starting point of the numerical simulation, initial configurations of target mechanisms are given according to the primary operation in forward dynamics analysis.

2.2. *Formulation of the contact-impact process in planar multibody system*

When describing the contact impact event between the hoof and ground, we need to determine the state of the contact and accurately acquire the relative deformation and speed at the instant of contact.

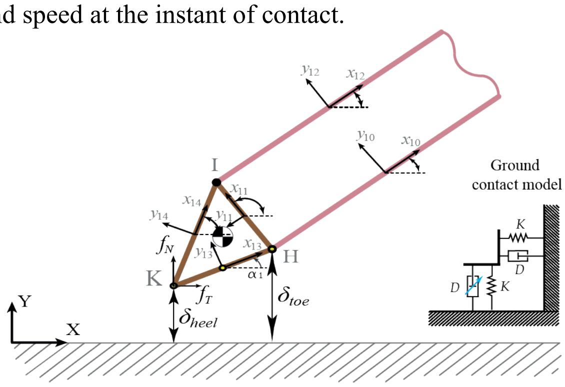


Fig. 1. Representations of foot-ground interactions by kinematic model of hoof-ground contact.

As shown in Fig. 1, only non-contact scenarios is considered in the kinematic model. When the leg contacts with the ground, the deformation or penetration is estimated as

$$\delta = y_G - P_n, \tag{10}$$

where the P_n is the normal component of the position vector or the potential contact point as $K = \left[K_x, K_y\right]$ (toe-first) or $H = \left[H_x, H_y\right]$ for the case of heel first contact. The ground plane y_G is consistently levelled from the lowest point of the locomotive trajectory in the sense of the relative indentation.

$$\delta = \begin{cases}\delta & \left(\delta > 0 \quad F_N = 0\right)\\ -\delta & \left(\delta \le 0 \quad F_N \ge 0\right)\end{cases} \tag{11}$$

In Eq. 11, the positive value of δ is that distance represent a separation, while negative values denote relative penetration of the contacting bodies. Therefore, the sign of penetration indicates the phase transition from swing to stance and vice versa for the case of hoof-ground contact. By using Eq. 10, it is clear to discriminate walking phases and contact forces as continuous function of deformation. Based on updated position and velocity of potential contact points on the hoof $K = \left[K_x, K_y\right]$ or $H = \left[H_x, H_y\right]$ from the multibody kinematic analysis, relative normal and tangential velocity among the contact points are estimated as

$$v_n = \dot{\delta} = \left(\dot{r}_j - \dot{r}_i\right)^T n, \tag{12}$$

similarly,

$$v_t = \left(\dot{r}_j - \dot{r}_i\right)^T t, \tag{13}$$

where t is the unit vector in tangential direction obtained by rotating unit normal vector n counter-clockwise by 90 degrees. When bodies come into contact, the normal contact force can be estimated with contact force model introduced by Lankarani & Nikravesh (1990) [20]. In the damping factor model, the crucial parameter known as coefficient of restitution was introduced which defines the whether contact is fully elastic $cr = 1$ (or plastic) $cr = 0$, resulting impact force model can be written

$$F_N = K\delta^n\left[1 + \frac{3\left(1-cr^2\right)}{4}\frac{\dot{\delta}}{\dot{\delta}^{(-)}}\right]. \tag{14}$$

where cr denotes the coefficient of restitution which is defined as the ratio of relative approach velocity $\dot{\delta}^{(+)}$ and relative departing velocity of $\dot{\delta}^{(-)}$. The tangential friction force in contact is calculated using modified Coulomb friction law [21]. The dynamic friction forces F_f in the presence of sliding [22] can be written as

$$F_f = -\mu_k F_N c_f c_d. \tag{15}$$

where μ_k the kinetic friction coefficient. In order to take the friction force effect into account, the direction of the tangential velocity vector $c_f = |v_T|$ with the target body velocity v_T, which is opposite to direction of motion. c_d is the dynamic correction coefficient given by [21]. If the normal force F_N is obtained from a contact with a force-deformation (Eq. 14.) of contact force model, a logical point-to-point spring-damper element [23] is only active during the period of contact in normal and tangential direction as shown in kinematic model in Fig. 1. Finally, absolute values of the force vector in normal and tangential defines the resultant ground reaction forces $GRF = \sqrt{F_N + F_f}$ exerted when the hoof press against the ground.

3. Results

According to the MBD formulation above, the constitutive contact force model is calculated numerically with kinematic and dynamic analyses, which allow to visualize temporal evolutions of the locomotive trajectory, resultant ground reaction force of the leg mechanism.

Table 1. Parameters used in the numerical simulation.

Kinematic/Dynamic analysis		
Gravitational acceleration [m/s²]	g	9.81
The velocity of the driving crank [rad/s]	ω	2π
Total simulation time [s]	t	$0 \le t \le 1$
Baumgarte parameter	α	15
Baumgarte parameter	β	$\sqrt{2\alpha}$
Time step [s]	dt	1.0×10^{-3}
Contact force analysis		
Stiffness [N/m]	K	2×10^4
Coefficient of restitution	cr	$0 \le cr \le 1$
Coefficient of kinetic friction	μ_k	0.4

Table 1 displays the parameters used in the numerical simulation for contact force analyses of horse leg mechanisms. Matlab-based numerical simulation was performed with a combination of the Euler method with the time step of $1\times10^{-3}s$. For the contact analysis to prevent an unnatural rise of ground reaction force, the stiffness is no smaller than 2×10^4 N/m for the contacting bodies is considered according to simulation in [24]. It's simply that the ground deformation is equivalent to the resultant motion of the stiffness-damping system under the compression. Two trajectories were selected as typical trajectories to reproduce walking and running gaits according to the mechanics of the simple model, which are generated by changing the control parameters associated with driver constraint in the horse leg mechanism.

As shown in Fig.2, the 1*st* trajectory considered to be a walking in term of duty factor and step length, the relation was evaluated in previous study. The 2*nd* trajectory which has longer step length compared to the 1*st* trajectory considered to be running gait in which leg orientation controlled by swinging motion with respect to the body, which may potentially generate the propulsive ground reaction force due the intrinsic high rate of angular oscillation at the hoof.

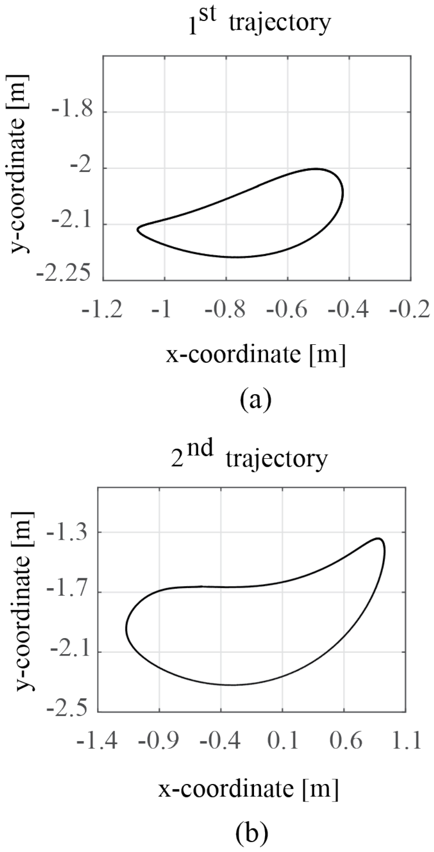
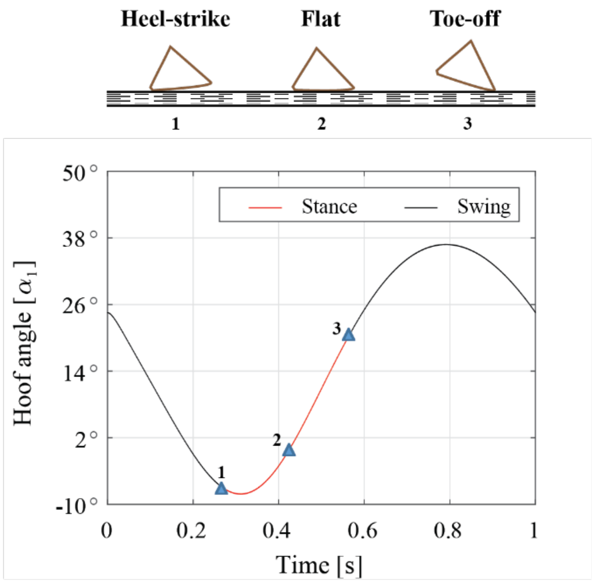
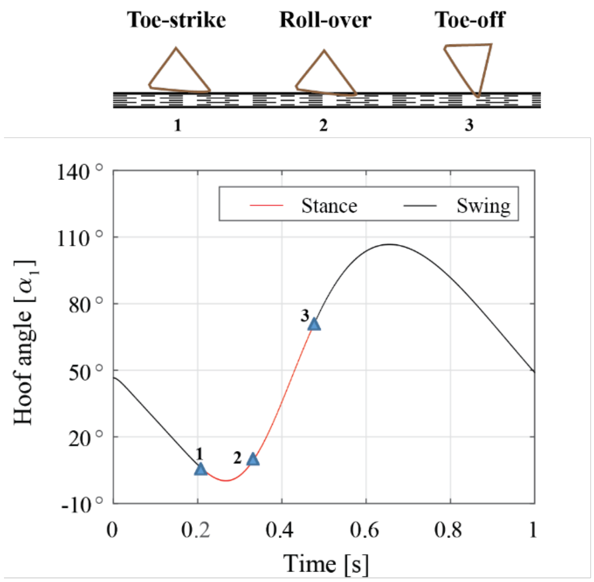


Fig. 2. Gait trajectories generated from the horse leg motion. Walking (a), Running (b).

In the temporal evaluation of the hoof angle as shown in Fig. 3, walking and running gaits were generated in accordance with the first and second trajectories in Fig.



(a)

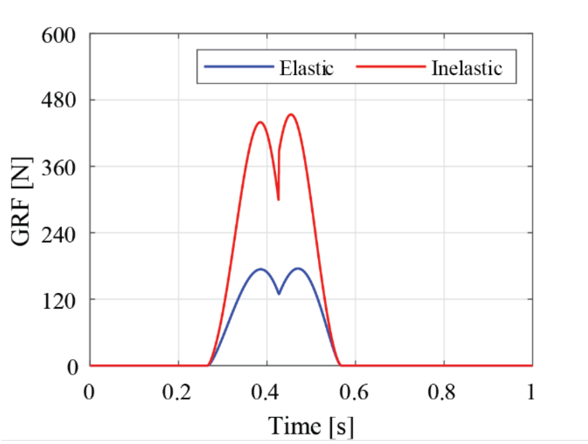


(b)

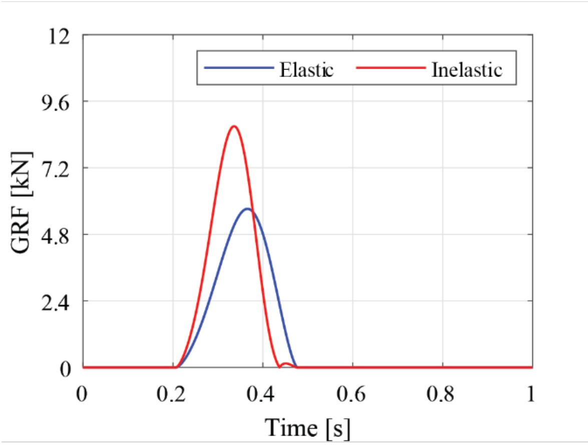
Fig. 3. Hoof angle variations with respect to the locomotive trajectory. Angular rotation of hoof in one locomotive cycle of walking gait (a) and running gait (b).

2. In the waling gait, the transition of sub-phases commonly occurred in the rolling motion of the triangle structure in both cases. In the walking gait, those sub-phases from heel-strike, flat and toe-off were gradually generated (Fig. 3(a). top), which is consistent with the gradual change of the hoof angle (Fig.3 (b). top), while in the running gait, the transition started from the toe-strike that causes the roll-over sub-phase quickly and then reached to the toe-off sub-phase with a large gap from the ground rather than that in the walking gait. The result was numerically examined as the first contact occurs at instant time $t_1 = 0.267s$ and $\alpha_1 = -6.83^{\circ}$ leaves the ground at $t_3 = 0.569s$ and $\alpha_1 = -20.9^{\circ}$ for the 1^{st} (walking) trajectory. In the case, the hoof rolling over the ground and during the mid-stance multiple contact points were examined at $\alpha_1 = -0.23^{\circ}$ degree in which hoof was almost parallel with the ground. When the hoof leaves the ground, angular variation during the contact phase was reached 26° degrees. For the 2^{nd} (running) trajectory, the contact phenomenon occurred at the instant time $t_1 = 0.208s$ and $\alpha_1 = 6.20^{\circ}$ firstly, shifted to leaving from the ground at $t_3 = 0.478s$ and $\alpha_1 = 71.08^{\circ}$. Finally, the hoof angle was reached 64.8° degrees, which is approximately 2.5 times larger than the result in the walking gait. Interestingly, the hoof-ground interaction was significantly influenced by the initial impact phenomenon, which differentiate the grounding part either heel or toe and it reflects to successive sub-phases too. Even in the simple model, the differentiation was clearly observed not only in the trajectory level as a posture and kinematic analysis but also in the kinetic level as the dynamics analysis, which is easily extended to the energy analysis. All these sub-phases are characterized as representative sub-phase to reconstruct the target gait and it can provide the parametric analysis with stiffness and damping, which is associated with an actual parameter from soft tissues in the distal limb. It is because that the compliant contact force model was theoretically described as shown in the method section, which allow to change those parameters related to reaction forces easily.

In the Matlab based numerical simulation, ground reaction forces were evaluated by using Lankarani & Nikravesh [20] model by changing restitution coefficients as $cr = 1$ for elastic and $cr = 0$ for inelastic contact cases. According to the analysis, the maximum contact force was obtained as 453.65N for the 1^{st} (walking) trajectory in the inelastic contact condition and 175.34N in the elastic case. Average normal force was 80.83N and 34.88N respectively. For the 2^{nd} (running) trajectory, the maximum GRF was 8.694kN for the



(a)



(b)

Fig. 4. Resultant values of ground reaction forces under the different contact condition. Force pattern of walking gait (a) and running gait (b).

inelastic contact condition and 5.719kN in the elastic case. Average normal force was 978N and 800N respectively. Result showed that the significant reduction of ground reaction force with respect to the viscoelastic contact was estimated as 61.35% in 1*st* trajectory and 34.22% in 2*nd* trajectory. In the walking trajectory, multiple point support occurring in which hoof loaded the body weight at mid-stance as shown in Fig. 4(a), the maximum pressure or force peak indicated in only when the leg first impact and leaves the ground which was consistent with the force pattern observed in the human walking gait [25,26].

4. Conclusion and Discussion

For the sake of the establishment of the generalized analysis of the viscoelastic effect with respect to various gait patterns, we introduced a novel theoretical method as the integrative framework of the multibody dynamics and contact force model and then the computational analysis clearly demonstrated results in four different conditions from the combination of walking/running and elastic/inelastic contact. In the simplification of the hoof mechanism by using a triangle structure to focus on three nodes as heel, ankle and toe, the result revealed not only the two peak generation in the temporal evolution of the ground reaction force during the stance phase of the walking gait, but also a single peak observation in the case of the running gait. The phenomenon was generated from that wheel-like rolling motion of the hoof support and stabilizes the body in mid-stance, and partially consistent with the observation in the human foot-ground interaction [1,10]. In spite of the existence of the complexity in the horse leg structure, our computational result suggested a possible way to analyze the spring-damper effect embedded in principle to the biological mechanism for absorbing the ground reaction force flexibly as a common tendency of the walking and running gaits. The analysis with damping factor model by Lankarani & Nikravesh (1990) realized the fact that a soft grounding effectively reduces the ground reaction force. In the actual phenomenon due to the interaction in the musculoskeletal systems such as a vibration from the sensory feedback loop [27,28], the further detail modeling is crucial. Phenomenologically in the level of the force generation, we successfully observed a time delay in a peak in the force time profile specifically in running gaits, which indicates an intrinsic behavior due to the damping factor is highly important in the kicking motion of the hoof. The phenomenon is also need to be verified with more detail leg models and evidences from the biological system [29,30]. Further considerations are possible to compare in the energy consumption due to the elasticity [18] and robotic application [31]. Geometric characteristics of the hoof is of interest to researchers studying on the biological nature mechanical in the hoof shape and its influence in joint angles and contact timing [32,33].

Acknowledgements

This work was supported in part by Mongolia-Japan higher Engineering Education Development (MJEED-

JICA) through the joint research project "Development of industrial, service and intelligent systems based on advanced technologies" (J23A16), JSPS KAKENHI (16H01616, 17H06383) and the New Energy and Industrial Technology Development Organization (NEDO), and Project on Regional Revitalization Through Advanced Robotics (Kyushu Institute of Technology/Kitakyushu city, Japan).

References

1. D. Joshua Mayich, Alison Novak, Daniel Vena, Timothy R. Daniels, and James W. Brodsky. Gait analysis in orthopedic foot and ankle surgery - Topical review, part 1: Principles and uses of gait analysis. *Foot and Ankle International*, 35:80–90, 2014.
2. Logan J. Dockery. Analysis of Gait Propulsion During the Ankle and Forefoot Rocker Sub-phases. *PhD thesis*, 2019.
3. T. Mcgeer. Passive Dynamic Walking. *The International Journal of Robotics Research*, 9(2):62–82, 1990.
4. P. Mahmoodi, R.S. Ransing, and M.I. Friswell. Modelling the effect of 'heel to toe' rollover contact on the walking dynamics of passive biped robots. *Applied Mathematical Modelling*, 37(12):7352–7373, 2013.
5. D. Torricelli, J. Gonzalez, M. Weckx, R. Jiménez-Fabián, B. Vanderborght, M. Sartori, S. Dosen, D. Farina, D. Lefeber, and JL. Pons. Human-like compliant locomotion: state of the art of robotic implementations. *Bioinspiration and Biomimetics*, 11(5), 2016.
6. Eduardo Corral, M. J. Gómez García, Cristina Castejon, Jesús Meneses, and Raúl Gismeros. Dynamic modeling of the dissipative contact and friction forces of a passive biped-walking robot. *Applied Sciences* (Switzerland), 10, 2020.
7. A. H. Hansen, D. S. Childress, and E. H. Knox. Roll-over shapes of human locomotor systems: effects of walking speed. *Clinical biomechanics* (Bristol, Avon), 19(4):407–414, 2004.
8. A. H. Hansen and D. S. Childress. Investigations of roll-over shape: implications for design, alignment, and evaluation of ankle-foot prostheses and orthoses. *Disability and Rehabilitation*, 32(26):2201–2209, 2010.
9. Bellusci G. Schepers H. M. de Zee M. Andersen M. S. Karatsidis, A. and P. H. Veltink. Estimation of Ground Reaction Forces and Moments During Gait Using Only Inertial Motion Capture. *Sensors*, 17(1) (75), 2016.
10. Bonnefoy-Mazure A. and Armand S. *Normal gait*, pages 200–211. Amsterdam: Elsevier, 2015.
11. Alberto Leardini, John J. O'Connor, and Sandro Giannini. Biomechanics of the natural, arthritic, and replaced human ankle joint. *Journal of Foot and Ankle Research*, 7, 2014.
12. W. Lee and M. Raibert. Control of hoof rolling in an articulated leg. *In Proceedings. 1991 IEEE International Conference on Robotics and Automation*, pp. 1386–1391.
13. D. Batbaatar and H. Wagatsuma. A proposal of the kinematic model of the horse leg musculoskeletal system by using closed linkages. *In IEEE International Conference on Robotics and Biomimetics*, pp. 869–874, 2019.
14. T. Jansen. *The Great Pretender*. Netherlands:010, 2007.
15. K. Komoda and H. Wagatsuma. Energy-efficacy comparisons and multibody dynamics analyses of legged robots with different closed-loop mechanisms. *Multibody System Dynamics*, 40:123–153, 2017.
16. Sandra D. Starke, Justine J. Robilliard, Renate Weller, Alan M. Wilson, and Thilo Pfau. Walk-run classification of symmetrical gaits in the horse: A multidimensional approach. *Journal of the Royal Society Interface*, 6:335–342, 2009.
17. Jeffrey J. Thomason and Michael L Peterson. Biomechanical and mechanical investigations of the hoof-track interface in racing horses. *Veterinary Clinics: Equine Practice*, 24(12):53–77, 2008.
18. R. Mcneill Alexander. Elastic energy stores in running vertebrates. *Integrative and Comparative Biology*, 24(1):85–94, 1984.
19. J. Baumgarte. Stabilization of constraints and integrals of motion in dynamical systems. *Computer Methods in Applied Mechanics and Engineering*, 1:1–16, 1972.
20. H. M. Lankarani and P. E. Nikravesh. A contact force model with hysteresis damping for impact analysis of multibody systems. *Journal of Mechanical Design, Transactions of the ASME*, 112:369–376, 1990.
21. J A C Ambrósio. *Impact of Rigid and Flexible Multibody Systems: Deformation Description and Contact Models*, pp. 57–81. Springer Netherlands, Dordrecht, 2003.
22. V. P. Chervinskii. Dissipative function of coulomb friction on elliptical contact area. *Journal of Friction and Wear*, 32:386–393, 2011.
23. Parviz E Nikravesh. *Planar multibody dynamics: formulation, programming with Matlab, and applications*. CRC Press, second edition, 2019.
24. Gerald Ristow. Simulating granular ow with molecular dynamics. *Journal de Physique I, EDP Sciences*, 2:649–662, 1992.
25. R. McNeill Alexander. Energetics and optimization of human walking and running: The 2000 Raymond Pearl Memorial Lecture. *American Journal of Human Biology*, 14(5):641–648, 2002.
26. R. McNeill Alexander and A. S. Jayes. Fourier analysis of forces exerted in walking and running. *Journal of Biomechanics*, 13(4):383–390, 1980.
27. Daniel F. B. Haeufle, Katrin Stollenmaier, Isabelle Heinrich, Syn Schmitt, and Keyan Ghazi-Zahedi. Morphological Computation Increases From Lower- to Higher-Level of Biological Motor Control Hierarchy. *Frontiers in Robotics and AI*, 7:1–13, 2020.

28. Z. H. Shen and J. E. Seipel. A fundamental mechanism of legged locomotion with hip torque and leg damping. *Bioinspiration and Biomimetics*, 7, 2012.
29. Ian Abraham, Zhuohua Shen, and Justin Seipel. A Nonlinear Leg Damping Model for the Prediction of Running Forces and Stability. *Journal of Computational and Nonlinear Dynamics*, 10:1–8, 2015.
30. Alan M. Wilson, M. Polly McGuigan, Anne Su, and Anton J. Van den Bogert. Horses damp the spring in their step. *Nature*, 414(6866):895–899, 2001.
31. E. Garcia, J. C. Arevalo, G. Múnoz, and P. Gonzalez-de Santos. On the biomimetic design of agile-robot legs. *Sensors*, 11:11305–11334, 2011.
32. C. Johnston and W. Back. Hoof ground interaction: When biomechanical stimuli challenge the tissues of the distal limb. *Equine Veterinary Journal*, 38:634–641, 2006.
33. Angelos Karatsidis, Giovanni Bellusci, H. Martin Schepers, Mark de Zee, Michael S. Andersen, and Peter H. Veltink. Estimation of ground reaction forces and moments during gait using only inertial motion capture. *Sensors*, 17:1–22, 2017.

Satellite Image-based UAV Localization using Siamese Neural Network

Seong-Ha Ahn
Department of Electronic Engineering, Pusan National University,
2, Busandaehak-ro 63beon-gil Geumjeong-gu
Busan, 46241, Republic of Korea

Ho-Sun Kang
Department of Electronic Engineering, Pusan National University,
2, Busandaehak-ro 63beon-gil Geumjeong-gu
Busan, 46241, Republic of Korea

Jang-Myung Lee
Department of Electronic Engineering, Pusan National University,
2, Busandaehak-ro 63beon-gil Geumjeong-gu
Busan, 46241, Republic of Korea

E-mail: seongha7379@pusan.ac.kr, hosun7379@pusan.ac.kr, jmlee@pusan.ac.kr
www.pusan.ac.kr

Abstract

We present a method for UAV localization using pre-existing satellite images. The use of Unmanned Aerial Vehicles (UAVs) has rapidly increased in several applications such as surveillance, search, and defense. When in GPS-denied situations, however, the onboard GPS signal may be noisy or inaccurate. The proposed method is based on a Siamese Neural Network that contains two instances of the same neural architecture and weights. Siamese Neural Network learns the similarity metric so that can recognize the same place from two raw images. Convolutional Neural Network is used as a backbone in Siamese Neural Network to overcome variation due to differences such as perspective, shadow angle, and presence of vehicles. We describe UAV localization pipeline and a dataset for training and testing our networks. Finally, the performance of the proposed method was shown in accuracy.

Keywords: visual localization, uav, satellite image, siamese neural network, image retrieval

1. Introduction

Unmanned Aerial Vehicles (UAVs) are aircraft that remotely controlled or fly autonomously with an onboard manipulation system. Nowadays, the use of UAVs has rapidly increased in several applications such as surveillance, search, and defense [1]. Localization is an essential task to utilize UAVs in such tasks. Currently, the onboard navigation system is mostly relied on Global Positioning System (GPS). When in GPS-denied situations, however, the onboard GPS signal may be noisy or inaccurate. Visual localization, which is the problem of estimating a camera pose, can be a useful alternative in GPS-denied environments. Image retrieval-based approach is one of the common visual localization methods [2]. In image retrieval tasks for UAVs visual localization, satellite imagery can be utilized as a geodatabase. In recent year, deep neural networks have

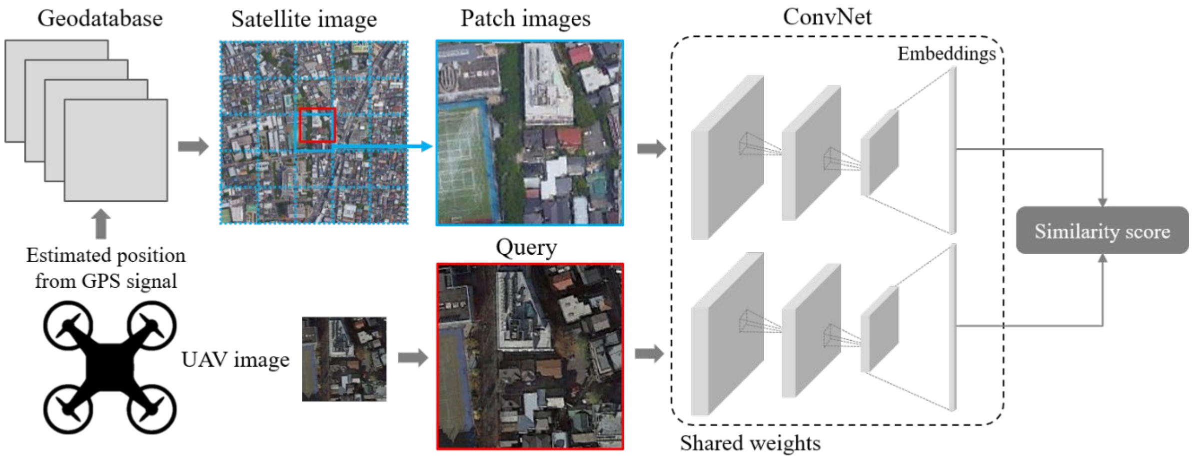


Fig. 1. Image retrieval pipeline of visual localization for UAVs. georeferenced satellite imagery is divided into patches for search and UAVs image is used as query image. The closest patch image will be retrieved by comparing similarity of every patch and query image pairs.

been successfully applied in various image processing and analysis tasks. We apply deep neural architecture in the image retrieval system to analyze images efficiently. In this paper, we propose a satellite image-based UAVs visual localization method with image retrieval approach. In section 2, we describe the deep neural architecture, called Siamese Neural Networks that learns metric for similarity. Dataset preparation for training neural architecture will be covered in section 3. We train and evaluate the proposed method in section 4. Finally, we summarize our study in section 5.

2. Image Retrieval using Siamese Neural Networks

2.1. *Siamese neural networks*

Siamese neural networks (SNN) is a neural architecture that has two networks which share its weights. Siamese neural networks employ a unique structure to learn a similarity metric between two input [3]. To extract features effectively from input images, Convolutional Neural Networks (CNN) is used in feature extraction layers of siamese neural networks. Figure 1 shows the image retrieval pipeline using CNN-based siamese neural networks. The proposed method uses semi-supervision of GPS system. The georeferenced satellite image is selected according to the noisy GPS signal from database. Satellite image is divided into few numbers of patches. The network takes two input: patch of satellite image and UAVs image. Input images are projected onto the low dimensional space which is called embedding space. By training siamese neural networks with UAVs and satellite patch imagery, the networks are trained to extract discriminative features to measure similarity. It helps the networks to recognize if two input images are taken from the same place or not. Every patch images and UAVs image are compared by measuring the distance in the embedding space.

2.2. *Triplet loss*

One of the options for training a siamese neural networks to learn similarity metric is to use triplet loss. The input data are projected into the embedding space via the networks. When Anchor, Positive and Negative images are input, triplet loss is calculated as the euclidean distance between three embedded points. (see Figure 2) The loss function can be described as follows:

$$L(A,P,N) = \max\left(\left\|f(A) - f(P)\right\|^2 - \left\|f(A) - f(N)\right\|^2 + \alpha, 0\right) \quad (1)$$

Where α is a margin between positive and negative pairs, and f is an embedding. In our implementation, we set α as 1.

Fig. 2. The distance from the Anchor to Positive is minimized and the distance from the Anchor to Negative is maximized.

3. Dataset

To simulate UAVs and satellite imagery, we used two satellite imagery sources. As shown in Figure 3, the difference in characteristics of satellite imagery sources causes various visual representations.

3.1. *Database properties*

We build two databases: UAVs image database and satellite imagery database. These databases are consist of total 88 correspondence images. Database image pairs from 3 different cities in New York and Florida. Each image covers a 0.16km^2 area of land. UAVs database image size with 0.125m resolution is 3200*3200 pixels, satellite image size with 0.5m resolution is 800*800 pixels. In our experiments, we divide database images into 16 patches per each. For training and validating, we generate approximately 30k and 0.5k, 0.5k image pairs respectively with the augmented images generated by rotating and adding a random value to channels.

3.2. *Anchor, Positive and Negative image*

As described section 2, triplet loss is calculated by measuring euclidian distance between Anchor, Positive and Negative images embedded in embedding space. For positive images, we create 4 shifted images from the original patch. The pixel range for the shift is limited to ±25 percent of total pixels in the image. The Anchor and Positive image pairs are created by combining 8 images. The Negative images are sampled in the same database image except the same patch and the other images.

Fig 3. Various visual representation of the same area caused by different characteristic of satellite imagery sources.

4. Implementation

We implement the proposed system in Python and PyTorch. Experiments have been done using 2 RTX2080ti. Stochastic Gradient Descent (SGD) optimizer is employed to train the networks. Since the only image with the shortest Euclidean distance is assumed to be a *Positive*, we evaluate our methods by calculating top- k precision score. The proposed method achieves as Table 1. It retrieves correct images for most queries, however, it performed poorly in specific cases, such as when the area has too many structures. (see Figure 4) It shows that the dataset requires more images that contain various visual features for generalization.

Table 1. The experimental results on our test set.

Precision score

Top-k precision	score
k=1	0.65
k=3	0.74

Fig 4. Example image pairs with good retrieval results (left column) and bad retrieval results(right column).

5. Conclusion

We proposed and implemented an image retrieval system that queries UAV images and retrieves the most similar image from pre-existing satellite images using CNN-based Siamese neural networks. To simulate UAVs and satellite imagery, we created a dataset using two different satellite imagery sources. The proposed system performed decently on our test set, however, the performance should be improved in some specific cases. In order to deploy the system for practical usage, the more diverse dataset is necessary for generalization. Moreover, further research of the additional post-process is required to acquire global camera pose with high accuracy.

Acknowledgements

This research was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No.2019R1A2C2088859)

References

1. B. H. Y. Alsalam et al., "Autonomous UAV with vision based on-board decision making for remote sensing and precision agriculture," 2017 IEEE Aerospace Conference, Big Sky, MT, 2017, pp. 1-12, doi: 10.1109/AERO.2017.7943593.
2. Pion, Noé et al., (2020). Benchmarking Image Retrieval for Visual Localization.
3. F. Schroff et al., "FaceNet: A unified embedding for face recognition and clustering," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 815-823, doi: 10.1109/CVPR.2015.7298682.

Analysis based on CNN for Automated Vehicle Parking Occupancy

Muksimova Shakhnoza
Department of Computer Engineering, Gachon University, Gyeonggi-do 461-701
Seongnam, Korea

Young Im Cho
Department of Computer Engineering, Gachon University, Gyeonggi-do 461-701
Seongnam, Korea
E-mail: muksimova1986@gmail.com, yicho@gachon.ac.kr
www.gachon.ac.kr

Abstract

The remarkable growth of the automobile industry, along with a lack of urban planning, exacerbated problems such as traffic congestion, air pollution and driving difficulties. The majority of car manufacturers have a pre-installed or aftermarket navigation device in vehicles to make it easier for customers. This helps drivers to travel to their destination with ease. However, a rough estimate of the remaining distance to the destination is given by navigation systems. With the latest advancement in technology, companies are interested to find out whether these advancements can assist in reducing the time spent to search for empty parking space. Locating a vacant parking spot is a big predicament that drivers face because looking for it is a tiring operation. Our investigations have shown that, in contrast with previous methods, for the purpose of classifying given parking spaces as empty or occupied, the current solution is more robust, stable and well-generalized for unseen images taken from entirely separate camera points of view, which has clear signs that it can easily generalize to other parking lots.

Keywords: Convolutional neural network, parking, sensors, lightweight network

1. Introduction

The value of unoccupied parking space detection systems is growing dramatically, as preventing traffic congestion and the time-consuming task of locating empty parking space is a big challenge for urban center drivers. However, current parking space occupancy monitoring technologies are either costly hardware or not well-generalized for differing images taken from multiple camera views. As a solution, we take advantage of an effective visual identification approach that is made possible by the fact that camera tracking is already present in most parking spaces. The current challenge, on the other hand, is a difficult vision job due to outdoor conditions such as lighting variance, occlusions, distortion of focus, multiple camera views, and adjustments due to the different seasons of the year. In order to address these limitations, we suggest an approach based on the Convolutional Neural Network specially constructed to detect parking space occupancy in a parking lot, provided only the picture of a single parking spot as data.

In addition, the computer vision for traffic applications is an evolving field of study, with major developments over the last few years. For example, intelligent transport management systems, such as those examined in [1], aim to accomplish tasks such as monitoring, detecting and identifying vehicles, as well as higher-level tasks such as recognizing vehicle actions and revealing irregularities. In a vision-based device, any visual node that contains a camera and a transmitter will simultaneously observe several cars, minimizing the cost per parking stand. As it can be used for other activities, such as security, the device base is normally already configured and can be handled by other applications after implementation. However, the literature review [3] has demonstrated that image-based parking reservation tracking policies can be enforced using existing security cameras that are already connected to the main monitoring system. On the opposite, those who support sensor-based approaches, such as the authors of [2], contend that video cameras are significantly more costly than sensors and generate a significant quantity of data that cannot be readily disseminated over a wireless network.

In this paper, to resolve the above problems, we are presenting a CNN framework that uses dilated convolution neural networks to show the status of parking space occupancy. The PKLot dataset [4], which is freely available, was used to test the efficiency of our model. Using pictures of various situations and parking lots, we were able to test the generalizability of our strategy. We learned our model on a subset of images and checked it on a totally different one. Several methods already exist, such as the one explored in the analysis mentioned in [5], AlexNet, where this sort of generalization property has been experimentally tested; however, we have shown that our method outperforms them. Most of the related studies contrasted their methodology to Alex Net [6], which showed a high degree of success in this mission. Thus, in conjunction with Alex Net, as well as with other well-known deep learning structures, we tested our method

2. Related Work

To address the problem mentioned above, there are currently different schemes, such as those suggested in [2, 7]. In several studies, the issue with the parking management scheme was resolved by the use of smart devices. Sensor-based solutions include sensors for identification, such as ultrasonic sensors, which are mounted in each parking room. Huang and Wang [9] held that image-based approaches can be divided into two classes: space-driven and car-driven. Space-driven approaches are targeted at finding empty parking spaces rather than vehicles. In the case of images taken from static (such as surveillance) cameras, the technique most commonly used is background subtraction, which implies that the background deviation is statistically immobile over a small time period. Considering that this presumption does not hold outdoor or outdoor scenes, the limitations of this method are instantly apparent. The authors used Gabor filters to train a classifier with images of unoccupied parking spaces under different lighting conditions. In the category of car-driven process, algorithms are learned to identify vehicles that are objects of interest. Due to the distortion of perspective found in most of the parking lot images, as seen in Figure 1, cars positioned at a substantial distance from the camera occupy a small area of the scene, and thus less information can be observed, which greatly decreases the efficiency of the object detection algorithms. The technique presented in [5] is based on a deep Convolutional Neural Network specially developed for smart cameras. They have obtained reasonably high results, and we will equate our results with that. The method, named mAlexNet, is a variant of AlexNet and provides a decentralised approach for visual space occupancy detection. Very strong results were obtained, but only for the same subset of images.

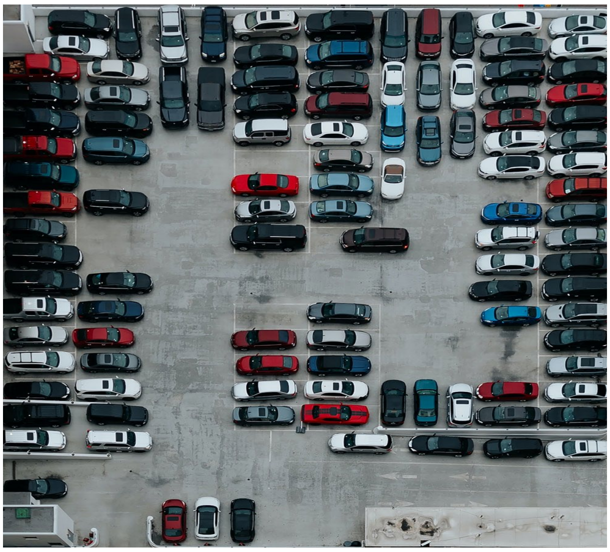
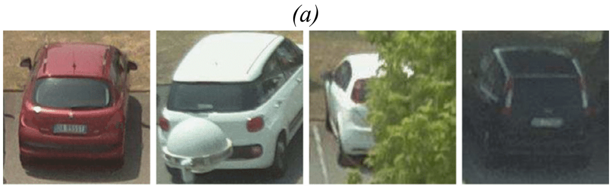


Figure 1. Example for a car parking image from PKLot dataset [4].

(a)



(b)

Figure 2. Segmented images of dataset for occupied and vacant space.

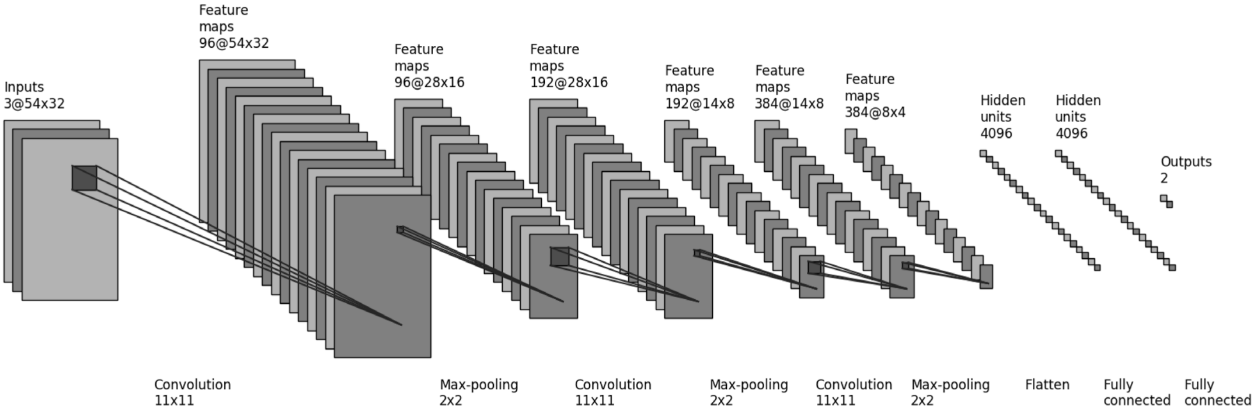
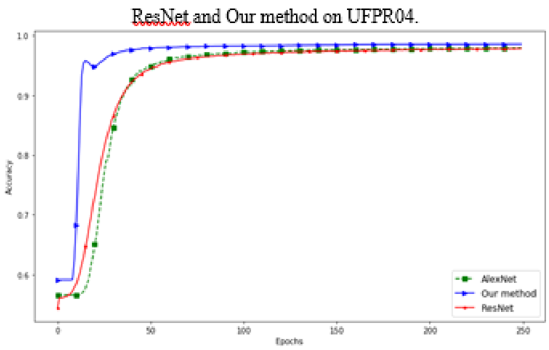


Figure 3. Comparison of our model with respect to other models

3. Proposed Approach

There are three subsets of this PKLot dataset [4] and during this experiment, we only utilized two of them in



our work. As we mentioned above on this task, AlexNet [6] model was achieved highest performance and so far a number of models were compared to this model in this reason we also compared our work to AlexNet [6] model. Our model is five layers deep and output layer. However, AlexNet [6] model is eight layers deep that five convolutional and three fully connected layers and learning of. parameters are also fairly many than our network. While increasing the number of learnable parameters makes the model bigger and increases training time. Below, in Table 1 and Figure 3 is shown experimental results of the models.

Table 1. Comparison of training and validation scores for AlexNet, ResNet and Our method on UFPR04.

Name of Method	Training Accuracy	Validation Accuracy
Our Method	97.92%	97.87%
Alex Net[6]	96.99%	97.91%
Res Net[5]	96.51%	97.80%

4. CONCLUSION

An efficient solution-Our model, for visual detection of a parking status, was presented which uses Convolutional Neural Networks. Our novel model is provided as a robust design for employees to classify images of parking spaces taken from a camera as occupied or vacant. In this proposed approach, we used three contributions: a little number of layers, small window sizes. We showed by experiments that the current task can be tackled effective by using these contributions. To assess the performance of the model and to be able to compare it with respect to other architectures, we made our experiments on PKLot dataset since prior attends were evaluated on this dataset as well.

Acknowledgements

This research was supported the National Research Foundation (NRF)—2018R1D1A1A09084151.

References

1. Tian, B.; Morris, B.T.; Tang, M.; Liu, Y.; Yao, Y.; Gou, C.; Shen, D.; Tang, S. *Hierarchical and Networked Vehicle Surveillance in Its: A Survey*. ***IEEE Trans. Intell. Transp. Syst.*** 2015, ***16***, 557–580.
2. Tang, V.W.; Zheng, Y.; Cao, J. *An Intelligent Car Park Management System Based on Wireless Sensor Networks. In Proceedings of the 1st International Symposium on Pervasive Computing and Applications*, Urumchi, Xinjiang, China, 3–5 August 2006; pp. 65–70.
3. Ichihashi, H.; Notsu, A.; Honda, K.; Katada, T.; Fujiyoshi, M. *Vacant Parking Space Detector for Outdoor Parking Lot by Using Surveillance Camera and FCM classifier. In Proceedings of the IEEE International*

Conference on Fuzzy Systems, Jeju Island, Korea, 20–24 August 2009; pp. 127–134.

4. Amato, G.; Carrara, F.; Falchi, F.; Gennaro, C.; Meghini, C.; Vairo, C. *Deep Learning for Decentralized Parking Lot Occupancy Detection*. ***Expert Syst. Appl.*** 2017, ***72***, 327–334. [CrossRef
5. De Almeida, P.R.; Oliveira, L.S.; Britto, A.S., Jr.; Silva, E.J., Jr.; Koerich, A.L. *PKLot-A Robust Dataset for Parking Lot Classification*. *Expert Syst. Appl.* **2015**, *42*, 4937–4949. [CrossRef].
6. Amato, G.; Carrara, F.; Falchi, F.; Gennaro, C.; Meghini, C.; Vairo, C. *Deep Learning for Decentralized Parking Lot Occupancy Detection*. ***Expert Syst. Appl.*** 2017, ***72***, 327–334. [CrossRef].
7. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012;* Curran Associates, Inc.: Red Hook, NY, USA, 2012.
8. Ashok, V.G.; Gupta, A.; Shiva, S.; Iyer, H.; Gowda, D.; Srinivas, A. *A Novel Parking Solution for Metropolitan Parking Garages. In Proceedings of the 3rd WSEAS International Conference on Urban Planning and Transportation*, Sozopol, Bulgaria, 22–24 July 2010; pp. 153–159.
9. Huang, C.C.; Wang, S.J. A *Hierarchical Bayesian Generation Framework for Vacant Parking Space Detection. IEEE Trans. Circuits Syst. Video Technol.* 2010**,** *20***,** 1770–1785**.** **[CrossRef**]

On the Boundary Layer in Singularly Perturbed Problem with An Initial Jump

Assiya Zhumanazarova
Department of Computer Engineering, Gachon University, Gyeonggi-do 461-701
Seongnam, Korea

Young Im Cho
Department of Computer Engineering, Gachon University, Gyeonggi-do 461-701
Seongnam, Korea
E-mail: assiya_mukhamaddin@mail.ru, yicho@gachon.ac.kr
www.gachon.ac.kr

Abstract

In this study, we consider the foundations of the boundary layer theory for studying the boundary layer part of the asymptotics of a singularly perturbed integro-differential boundary value problem with an initial jump. A part of the boundary layer is defined as an expansion in powers of a small parameter, and the coefficients of this expansion are solutions to differential initial value problems. The purpose of this study is to describe in detail the properties of the boundary layer part of the asymptotics based on the principles of boundary layer theory.

Keywords: *boundary layer, singular perturbation, initial jump, exponential estimates.*

1. Introduction

The boundary layer theory studied by Prandtl has been applied to characterize the problems of external flow around a body at high flow degree of velocity or low viscosity. Various fundamental problems of mechanics and physics can be formulated using a significantly simplified model using the approximate theory of the boundary layer. The results of analytical solutions to these problems are provided with a high degree of accuracy. The practical significance of the study of boundary layers lies in the fact that they take place in many important processes, where the effect of fluid viscosity is significant in the neighborhood of the boundary surface. One of the specific directions of the boundary layer theory is the friction layer in aerodynamics, where it is characterized by a strong gradient of flow velocity. To the theoretical study of the flow of low-viscosity fluids, the entire field can be divided into two parts: the region of a thin boundary layer near the wall, where friction forces must be considered, and the region outside the boundary layer, where the friction forces can be ignored due to their smallness.[1] The approximate character of the considered theory is that it is based on the assumption of a small thickness of the boundary layer.

The boundary layer theory can be considered from a physical and mathematical point of view. The theory of differential equations containing small parameters at the highest derivatives is important in mathematical modeling of physical processes. In the theory of singularly perturbed boundary value problems, a boundary layer phenomenon is observed when the solution to the perturbed problem differs significantly from the solution of a degenerate problem for arbitrarily small parameters. The behavior of the solution of a singularly perturbed system in the boundary layer is described by exponentially decaying functions.[2] The system describing a viscous fluid is an example of a singularly perturbed system. In hydrodynamics, the equation of an ideal fluid even at low viscosity is not suitable for describing a process near the boundary, called the boundary layer.[2] In perturbation theory the method of matched asymptotic expansions was developed to find an approximate solution by constructing different asymptotic solutions inside and outside the boundary layer of rapid changes and then, matching them together to obtain a unique approximate solution.[3]

In present section, we considered the basic concepts of the boundary layer theory to find association between their practical and theoretical importance. In the following section, we consider a singularly perturbed problem with an initial jump where the foundations of the theory of boundary layers are applied.[4]

The remainder of this paper is organized as follows. Section 2 is devoted to describe the main points based on the results of previous studies on singularly perturbed boundary value problem with an initial jump and to the formation of boundary layer part of asymptotics of the solution. Finally, Section 3 provides some concluding remarks.

2. The Boundary Layer Part of Asymptotics

Consider a singularly perturbed integro-differential boundary value problem with an initial jump.[5,6] In these studies, singularly perturbed and unperturbed problems, as well as the convergence of solutions using asymptotic methods are investigated. Since the presence of an integral term in the original equation significantly affects the solution, the phenomenon of an initial jump was observed in the problem at the left point of the considered segment. To form a uniform asymptotic expansion considering the asymptotic properties of solutions, the method of boundary functions was used. Thus, an algorithm was formulated for constructing asymptotic expansion of solution, which was characterized by a rapid change within the boundary layer. The asymptotics was sought as the sum of two parts - regular and boundary. A uniform asymptotic expansion of the solution is obtained by solving a sequence of simpler problems for finding the composite parts of the asymptotics. That is, a uniform asymptotic expansion has the form[6]

$$y(t,\varepsilon)=\sum_{m=0}^{n}\varepsilon^m y_m(t)+\sum_{m=0}^{n+2}\varepsilon^m\rho_m(\tau)+r_n(t,\varepsilon). \quad (1)$$

Here, $y_m(t)$ and $\rho_m(\tau)$ denote the regualar and boundary layer terms, respectively, and $r_n(t,\varepsilon)$ denotes the remainder term of the asymptotics.

Consider the differential initial value problems defined for finding the boundary layer terms

$$\begin{aligned}&\dddot{\rho}_m(\tau)+A(0)\ddot{\rho}_m(\tau)+B(0)\dot{\rho}_m(\tau)=\psi_m(\tau),\ m\ge 0,\\&\psi_0(\tau)\equiv 0,\end{aligned} \quad (2)$$

$$\begin{aligned}&\rho_0(0)=-\Delta_0,\ \dot{\rho}_0(0)=0,\ \ddot{\rho}_0(0)=B(0)\Delta_0,\\&\rho_1(0)=-\Delta_1,\ \dot{\rho}_1(0)=\beta-y_0'(0),\\&\ddot{\rho}_1(0)=-A(0)(\beta-y_0'(0))+B(0)\Delta_1-\int_0^\infty\psi_1(p)dp,\\&\rho_m(0)=-\Delta_m,\ \dot{\rho}_m(0)=-y_{m-1}'(0),\\&\ddot{\rho}_m(0)=A(0)y_{m-1}'(0)+B(0)\Delta_m-\int_0^\infty\psi_m(p)dp,\ m\ge 2.\end{aligned} \quad (3)$$

By direct calculations, we obtain the solution for the 0-th order approximation in the form

$$\begin{aligned}\rho_0(\tau)=&-\frac{B(0)\Delta_0}{\mu_2(0)-\mu_1(0)}\left(\frac{1}{\mu_1(0)}e^{\mu_1(0)\tau}-\frac{1}{\mu_2(0)}e^{\mu_2(0)\tau}\right)+\\&+\frac{\left(B(0)-\mu_1(0)\mu_2(0)\right)\Delta_0}{\mu_1(0)\mu_2(0)}.\end{aligned} \quad (4)$$

It follows immediately from the solution that

$$\left|\overset{(j)}{\rho_0}(\tau)\right|\le \mathrm{M}e^{-\delta\tau},\ j=\overline{0,2},\ \tau\ge 0. \quad (5)$$

Here, M, δ are positive constants.

To find the solutions to problems (2), (3), $m\ge 1$ that is, for the non-homogeneous differential equations (2), $m\ge 1$ we have used the Lagrange's method of variation of constants. After finding the solutions $\rho_m(\tau),\ m\ge 1$ and their derivatives of first and second order, we have obtained their exponential estimates. Using the method of mathematical induction, that is, supposing that the estimates are valid up to the m-1-th order approximation, we prove them for the m-th order approximation.

The boundary layer phenomenon can occur when solving a boundary value problem for an integro-differential equation with a small parameter. The degenerate equation obtained by equating a small parameter has an initial jump in the integral term, as well as in the boundary conditions, modification is required, in consequence of which a boundary layer occurs.

3. Conclusion

In this study, we considered the main concepts of the boundary layer theory for studying a singularly perturbed integro-differential boundary value problem with an initial jump. To describe their properties, the boundary layer part of the asymptotic expansion of the solution was considered. The study revealed that the terms of the boundary layer part of the asymptotics are characterized by exponentially decreasing functions. In conclusion, having studied the foundations of the theory of the boundary layer, it was possible to characterize the properties of the boundary layer part of the asymptotic expansion of the solution. Further work is required to expand this study, including fully proving boundary layer term estimates and providing detailed solutions to non-homogeneous differential equations.

Acknowledgements

This research was supported the National Research Foundation (NRF)—2018R1D1A1A09084151.

References

1. H. Schlichting, *Boundary-Layer Theory*, Nauka, Moscow, 1974.
2. A.B. Vasilyeva and V.F. Butuzov, *Asymptotic Expansion of Solutions of Singularly Perturbed Equations,* Nauka, Moscow, 1973.
3. E.J. Hinch, *Perturbation Methods,* Cambridge University Press, Cambridge, 1991.
4. K.A. Kasymov, *Singularly Perturbed Boundary Value Problems with Initial Jumps,* Sanat, Almaty, 1997.
5. A. Zhumanazarova and Y.I. Cho, Asymptotic convergence of the solution of a singularly perturbed integro-differential boundary value problem, *J. Mathematics*, **8**(213), 2020, pp. 1–17.
6. A. Zhumanazarova and Y.I. Cho, Uniform Approximation to the Solution of a Singularly Perturbed Boundary Value Problem with an Initial Jump, *J. Mathematics*, **8**(2216), 2020, pp. 1–14.

A Design and Implementation Intelligent flowerpot

Songyun Shi , Yizhun Peng*,
College of Electronic Information and Automation,
Tianjin University of Science and Technology, Tianjin, 300222, China;
*E-mail: * pengyizhun@tust.edu.cn*
www.tust.edu.cn

Abstract

The original intention of the device is that people often neglect the care of potted plants at home, resulting in plants unable to grow and survive unattended. This intelligent flowerpot is a device that can make potted plants survive and grow better without supervision. Smart home products based on Internet of things technology. The data of temperature sensor, humidity sensor, soil humidity sensor, harmful gas sensor, photosensitive sensor and other sensors are collected by STM32 single chip microcomputer, and the data are cooperated with intelligent tracking system and automatic irrigation system. To achieve the purpose of potted cultivation, beautify and improve the living environment. In view of the disadvantages of traditional family life of artificial cultivation and potted plants, the maintenance of scientific intelligence is realized, and the intelligent flowerpot system is designed.

Keywords: Internet of things technology, singlechip, Smart home, ecology, WIFI, flowerpot.

1. Introduction

With the accelerating pace of people's life, people usually ignore the care of potted plants, and potted plants can not play their due role. Gradually, people have lost the habit and interest of planting potted plants at home[1]. There are many advantages of planting potted plants at home, such as raising flowers, which can convey feelings and cultivate sentiment; green leaf potted plants can purify the air and make people happy. So potted plants have been loved by many people[2]. However, due to the fast pace of people's life and lack of time management, taking good care of potted plants has gradually become a luxury. Many people want to take some time to take good care of their flowers and plants, but they can't find the right time, so they stay at the stage of "just living". Therefore, this paper designed a kind of intelligent flowerpot which can be placed at home and take care of potted plants independently[3,4]. Users can not only not take care of the potted plants, but also know the situation of potted plants at any time[5]. Due to the rapid development of Internet of things technology, multiple intelligent flowerpots are connected to form a family ecosystem[6].

2. Embedded hardware system design

2.1. Overall structure of flowerpot hardware

When the signal of each sensor changes, the MCU will give instructions to the motion system

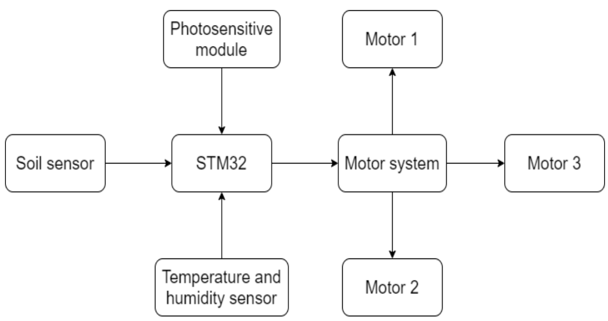


Fig.1. Hardware block diagram

2.2. Main control chip

The core microcontroller of the embedded hardware system is stm32f767igt6 chip, and the core is arm cortex-m7, which is introduced by STMicroelectronics. The operating frequency is up to 216mhz. On chip, 512KB flash memory and 256Kb SRAM memory are integrated; there are up to 21 communication interfaces, including 4 UARTS, 4 USART interfaces running at 12.5 Mbit / s, 5 SPI interfaces running at 50 Mbit / s, 3 I 2C interfaces, one can, two SDIO and one USB with PHY on chip 2.0 full speed device / host / OTG controller; 1 USB 2.0 High Speed / full speed device / host / OTG controller with built-in dedicated DMA controller, up to 18 synchronous 16 bit timers, and up to 140 I / O ports with interrupt capability. Compared with other MCU, it has high performance, low cost and low power consumption, which can fully meet the system requirements.

Fig.2. STM32F767IGT6 chip

2.3. Main components and sensors

2.3.1 Bluetooth serial communication module

The Bluetooth hc05 is a master-slave Bluetooth serial port module. In short, when the Bluetooth device and Bluetooth device are paired and connected successfully, we can ignore the internal communication protocol of Bluetooth and directly use Bluetooth as a serial port. When a connection is established, two devices share a channel, that is, the same serial port. One device sends data to the channel, and the other device can receive the data in the channel.

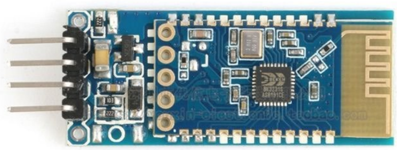


Fig.3. Hc05 Bluetooth serial communication module

2.3.2 Photosensitive module

The photosensitive resistance module is the most sensitive to the ambient light, which is generally used to detect the brightness of the ambient light and trigger the microcontroller or relay module. When the brightness of the ambient light fails to reach the set threshold, the do terminal outputs the high level. When the brightness of the external environment light exceeds the set threshold value, the do terminal outputs the low-level do The output end can be directly connected with the single-chip microcomputer, through which the high and low-level level can be detected, so as to detect the change of the light brightness of the environment; the analog output Ao of the small board can be connected with the ad module, and the more accurate value of the ambient light intensity can be obtained through AD conversion

.

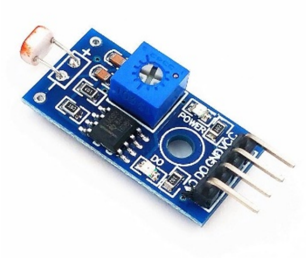


Fig.4.The motor drive module

2.3.3 Soil moisture sensor

The soil moisture sensor consists of two detectors to measure the amount of water in the soil. The two probes allow current to pass through the soil and measure the moisture content of the soil based on its resistance. When there is more water, the soil conducts more current, which means that the resistance will be smaller. So the moisture content will be higher. Dry soil reduces electrical conductivity. Therefore, when there is less water, the soil conducts less electricity, which means that it has a greater resistance. Therefore, the moisture content will decrease.

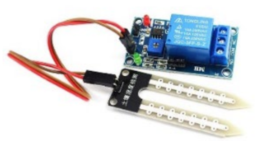


Fig.5. Soil moisture sensor

2.3.4 Integrated temperature and humidity sensor

The integrated temperature and humidity sensor uses digital integrated sensor as probe and digital processing circuit to convert the temperature and relative humidity in the environment into corresponding standard analog signal, 4-20mA, 0-5V or 0-10V. The temperature and humidity integrated analog sensor can transform the change of temperature and humidity value into the change of current / voltage value at the same time, and can be directly connected with various standard analog input secondary instruments.

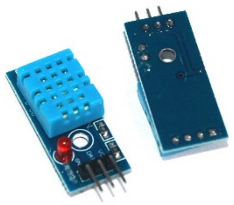


Fig.6. Integrated temperature and humidity sensor

3. Software design

STM32 microcontroller judges the growth environment of potted plants by reading the change of the level value of each sensor connecting pin. Each sensor detects the external environment in real time and transmits the collected signals to the single chip microcomputer. The Bluetooth module is connected with the mobile phone Bluetooth, and receives the information sent by the mobile phone Bluetooth for corresponding operation. The user's operation information can be transmitted to the flowerpot through WiFi.

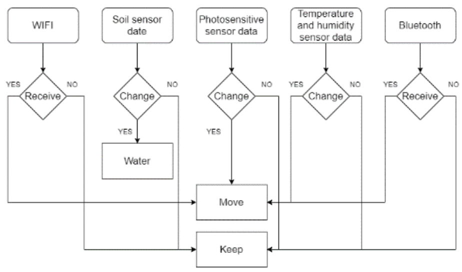


Fig.7. Software design flow chart

4. Testing and conclusion

4.1. Test plan

We put the seeds of Yushu in the flowerpot for cultivation, add a certain amount of water and fertilizer, put them on the windowsill with suitable light intensity, and let the flowerpot carry out self-cultivation for one month

Fig.8. Autonomic growth test

4.2. Test results

A month later, when we observed the flowerpot again, it had already sprouted. We extracted and analyzed the nutrient and water of the pot, and found that it met the normal standard of plant growth

Fig.9. Self growth test results

References

1. Zhang Liya; informatization and intelligent application of residential quarters [a]; science and education first and sustainable development in Western Development
2. Yang Weimin; discussion on the influence of environmental factors on flower growth and development [J]; horticulture and seedlings; issue 12, 2019
3. Tian Jie. Research on detection device of traffic sensing system based on Bluetooth [J]. Heilongjiang transportation science and technology, 2008,41 (05): 216-217.
4. Ni Tianle; Cheng Ying; Li Linfei; Xu Xinyi; Huang lifeI; research on indoor ecological and practical development of intelligent flowerpots [J]; rural economy and science and technology; issue 16, 2019

5. Yuan Zhenghong; Wang sixu; Li Jipeng; Huo Fengcai; an unattended intelligent home flowerpot design; science and technology wind, issue 05, 2017
6. Cen Hua; Zhang Na; modular design: an innovative thinking of green design; ecological economy; issue 05, 2008

Research of Attention-LSTM Model for Baby Cry Detection Robot

Tianye Jian[1,2], Yizhun Peng [1,2*], Wanlong Peng[1,2], Zhou Yang[1,2]
[1]*College of Electronic Information and Automation, Tianjin University of Science and Technology, Tianjin, 300222, China*
[2]*Advanced Structural Integrity International Joint Research Centre, Tianjin University of Science and Technology, Tianjin, 300222, China*
* *pengyizhun@tust.edu.cn*

Abstract

In order to achieve the effective acquisition of frame-level speech features under different emotional needs of baby, a speech emotion recognition model for baby based on improved long-term and short-term memory (LSTM) network is established. The frame-level speech features are used instead of the traditional statistical features to preserve the temporal relationships in the original speech, and the traditional forgetting and input gates are transformed into attention gates by introducing an attention mechanism, in order to improve the performance of speech emotion recognition, the depth attention gate is calculated according to the self-defined depth strategy. The results show that, in Fau Aibo Children's emotional data corpus and baby crying emotional needs database, compared with the traditional LSTM based model, the recall rate and F1 score of this model are 3.14% , 5.50% , 1.84% and 5.49% higher, respectively, compared with the traditional model based on Lstm and GRU, the training time is shorter and the speech emotion recognition rate of baby is higher.

Key words: children's emotional; time sequence relationship; frame-level speech feature; deep attention gate; Long Short-Term Memory(LSTM)

1. Introduction

Children's emotion recognition is an important part of affective computing[1]. Children are far less able than adults to act rationally in emotional outbursts and in response to different emotions, which can lead to emotional disorders if children are not able to act rationally and are directed in a timely manner, which can lead to anxiety disorders and other mental health problems. Therefore, it is of great significance to use appropriate algorithms or models to judge and guide children's emotions.

The researchers conducted in-depth research on children's emotion recognition from acoustic features, machine learning and deep learning[2] it is proposed to use Support vector machine and convolutional neural network to construct a system for detecting children's secondary emotional states[3] real-time emotional state of children is defined by multi-agent based interaction system[4] to establish the children's dual-modal emotion database and use the dual-modal emotion recognition method to measure the proportion of children's emotion contribution, and to point out that infants'(or young children's) emotion is more difficult to judge than older children's, babies usually cry to express their needs to their parents or guardians[5] the Mel-Frequency Cepstrum Coefficients (MFCC) of infant cries were extracted and classified based on Hidden Markov Model (Hmm) to identify whether the infant cries were healthy or not6. The Spectrogram was used as the feature vector, and the Convolutional Neural Network (CNN) was selected as the classification model to classify and recognize the crying of infants in pain, hunger and sleepiness[7]. The Support Vector Machine (SVM) was used as the classifier to classify the crying sounds of infants in the condition of hunger, pain and sleepiness, and the recognition effect was better than that of the Support Vector Machine (SVM).

Although the above algorithms have been successfully applied to children's emotion recognition, traditional machine learning algorithms, as well as self encoders and convolutional neural network in deep learning, can only accept data with fixed dimensions as input, this is in contradiction with the fact that the effective length of speech is constantly changing. To solve this problem, reference extracts emotion-related features (hereinafter referred to as frame-level features) from short-term speech frames, and applies static statistical functions (such as mean, variance, maximum, linear regression coefficient) to frame-level features, finally, the feature vectors with fixed dimensions are formed to represent the features of the frame speech[8-10]. Although this method solves the problem of model input, the time sequence information of the original speech is lost through the statistical analysis of the speech features.

2. Related work

2.1 LSTM network

The LSTM Network is a variant of Recurrent Neural Network (RNN), which is mainly used to process sequence information with long time difference. The LSTM network can solve the problem that the long-term information is difficult to store because of the gradient disappearance of RNN in the reverse propagation[11-13]. The LSTM network has been successfully applied in natural language processing (NLP) [14-16]. In order to enhance the ability of LSTM network to process data in specific tasks, the researchers further optimized the internal structure of LSTM network[17]. The fusion of the input gate and the forgetting gate of the LSTM network by the Gated Recurrent Unit (GRU) reduces the model parameters, the performance of LSTM network is better than that of GRU[18] in all machine translation tasks[19]. By using CONVLSTM network structure, the computing method of gate structure of LSTM is improved from Matrix multiplication to convolution[20]. Infinite Impulse Response Filter (IIR) Memory block of RNN is improved to Finite Impulse Response Filter (Fir) Memory block by Feedforward Sequential Memory Network (FSMN)[21], but FSMN usually needs to stack very deep layers, so FSMN has delay compared with one-way LSTM network[22]. Advanced short-term memory (Advanced LSTM) network, which uses attention mechanism to weight multiple cell states, can be effectively used for emotion recognition. However, reference[23] indicates that this method does not change the gate structure in the LSTM network and requires more training time. In addition, researchers are exploring how to stack LSTM structures to achieve more reliable emotion recognition[24]. End-to-end emotion recognition is achieved by extracting multi-channel speech features from a 6s-long speech waveform via a Convolutional Neural Network (CNN) as input to the LSTM Network[25]. In this paper, 1280

In this paper, we propose an improved Long Short-Term Memory (LSTM) based model for children's speech emotion recognition. Based on the LSTM network structure, the frame-level speech features are substituted for the traditional statistical features, in order to obtain better recognition performance, attention gates were used to replace the traditional forgetting gates and input gates, and to construct the deep attention gates by weighting attention in multiple cellular states.

kinds of abstract features were extracted from 6s-long speech waveform by CNN, and then fused with facial features as input of LSTM network.

The formula used in the traditional LSTM network is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \qquad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \qquad (2)$$

$$C_t = tanh(w_C . [h_{t-1}, x_t] + b_c \qquad (3)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C_t \qquad (4)$$

$$O_t = \sigma(w_o . [h_{t-1}, x_t] + b_o \qquad (5)$$

$$h_t = o_t . \tanh(C_t) \qquad (6)$$

σ Is the sigmoid activation function, $h_{t-1}$is the hidden layer output at $t-1$, x_t is the input at t, C_{t-1} is the cell state at $t-1$,$C_t$is the candidate value of the cell state at t, f_t、i_t、o_t. i_t and o_t are respectively forgetting gate, input gate and output gate.

2.2 Attention Mechanisms

The attention mechanism is based on the human visual attention mechanism. Attention makes people pay more attention to the important part of the information captured by vision, get as much detail information as possible, and reduce the attention to the irrelevant information around the target, that is, to suppress the irrelevant information[26]. In order to effectively utilize the information output from the LSTM network, the soft attention mechanism is introduced into the LSTM Network Model (hereinafter referred to as the LSTM model), and it is successfully applied to the field of machine translation, by weighting the attention of the LSTM model at different time, the relevance degree of the word to be translated and other words can be expressed. The encoding-decoder structure based on attention mechanism is proposed in reference[27], and its application in speech recognition is superior to HMM decoding system. Based on the encoder-decoder structure, a local attention model is proposed in reference [28], in which an alignment position is first predicted, and then a probability distribution similar to the soft attention model is obtained over the l window at the alignment position[29]. The single-head attention mechanism is improved to the multi-head attention mechanism, and the quality of machine translation is improved significantly by Transformer model.

In recent years, the mechanism of self-attention has become a hot topic[30]. Calculates the self-attention of the output of the LSTM model, calculates several fractions for different time steps, and then proposes a new LSTM model:

$$A = softmax(W_{s2} \cdot tanh(W_{s1} \times H^T)) \tag{7}$$

Where a is the attention fraction matrix, hi is the output of hidden layer cells at time I, and h is the stack result of hi in each time of LSTM MODEL:

$$H = (h_1, h_2, \dots, h_n) \tag{8}$$

The weighted output is expressed as:

$$M = A \times H \tag{9}$$

The introduction of the attention mechanism reduces the computational burden of processing Galway input data, and makes the task processing system more focused on finding information in the input data that is significantly related to the current output, thus improving the output quality. In recent years, researchers have applied attention mechanisms to the improvement of affective effects in speech[31]. The attention mechanism was used to screen the features and cross-link between the multilayer LSTM networks, and good effect of emotion recognition was obtained. In the output of RNN, a Local Attention mechanism is proposed to improve the effect of emotion recognition in multiple data sets[32].

3. Improved Deep Attention Gate

3.1 Attention Gate

In this paper, the attention mechanism is introduced into the internal structure of LSTM model, and the attention gate-based LSTM model is proposed, which greatly reduces the number of parameters of LSTM By introducing the concept of depth into the attention gate, the LSTM model can learn the input features better and avoid information redundancy. The attention-gate-based LSTM structure proposed in this paper enables the cell state at the last moment to determine the characteristics to be noticed when calculating each time step, at the same time, the attention gate is used to modify the traditional forgetting gate and the input gate is used to weight the features that need attention.

Because traditional input gates and forgetting gates are only implemented by a single fully connected layer, the model needs to be trained enough times to notice the cellular state information that needs to be left and the new input information that needs to be added, which causes it to converge and slow down[33]. On this basis, the peephole connection is added, the cell state is also taken as input, the cell state information is added into the three gates, and the increase of the parameters results in the increase of the training time and space complexity. In this paper, self-attention is paid to every cell state, and input candidate information is added to the parts of the cell state that do not need attention, the self-attention Algorithm replaces the three matrices needed for forgetting, input, and peeping connections with attention gates.

The attention gate at is defined as follows:

$$a_t = activation(V \cdot tanh(W \cdot C_{t-1})) \tag{10}$$

The activation is an activation function (the activation function can be selected according to the need, but its value range should be less than 1). The calculation formula for updating the Cell State is:

$$C_t = a_t \cdot C_{t-1} + (1 - a_t) \cdot C_t \tag{11}$$

Attention Gate can improve the recognition rate of the model while reducing the number of parameters and training time. Model distillation[34], 8-bit quantization[35], and shared parameter[36-37] are usually used in the existing reports. In this paper, an attention gate based on attention mechanism is proposed, which significantly reduces the parameters of LSTM model. In addition, the attention-gate-based model can reduce more training time for longer input sequences due to the modification within the LSTM model. For example, for a layer LSTM model with an input dimension of 512 and an output dimension of 256, if the bias is ignored, the commonly required parameters are: 1) the dimension weight of [512 + 256,256 × 4] for the three gate structures and the candidate values; 2) if the cell state at the last moment is taken into account in calculating the gate structure, a vector of 3 × [1,256] is added as the peephole[38]. In this paper, we do not need to introduce the peephole structure because we calculate the self-attention as the attention gate directly to the cell state. At the same time, the number of parameters required is reduced to [512 + 256,256 × 2] and the weight of 2 × [256,256] for calculating attention due to the combination of forgetting and input gates as attention gates. For this layer, the number of parameters was reduced from the original 787200 to 524288, reducing the number of parameters by 33.4%. For the LSTM model with deeper layers, more complex model and more data, the parameter quantity is reduced effectively.

3.2 Deep Attention Gate

The LSTM model is usually used to process time series information, but this information will increase with time, therefore, the calculation of the LSTM model at a certain time (that is, updating the cell state C and hidden layer output h) is only based on the external input and the cell state and hidden layer output at the previous time. Before the attention mechanism is put forward, if each moment is considered several times before, it will lead to too much information and the loss of important information, and increase the calculation amount and lead to the gradient explosion. However, the information of cell state at t-time is not only related to the information at t-1, but also closely related to the information at t-2, which is selectively forgotten at t-1. In this paper, the concept of deep forgetting gate is proposed and the corresponding input gate is designed.

The deep forgetting gate looks not only at the information about the state of the cell at the last moment (depth length = 1) , but also at the information about the state of the cell at the time t-2, t-3, ... , t-n (depth length = n) , which constructs the Deep-Attention-Lstm Cell structure, as shown in Fig.1

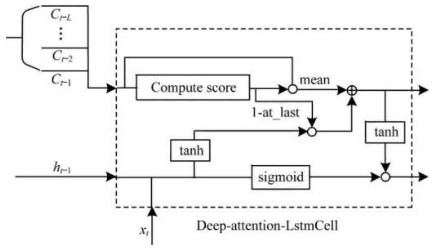


Fig.1 Deep-Attention-Lstm Cell internal structure sketch

It is worth noting that the introduction of “depth” will lead to an increase in training time. This is because in addition to forward increasing the number of attention gates in the loop for computing the state of multiple cells, reverse propagation also increases the number of chain derivations. From the point of view of the parameters of the model, although the depth will cause the increase of the training time, it will not cause the increase of the parameters of the model because the attention gate weights of each layer are shared by V and W.

In this paper, the aim of depth is to improve the performance of speech emotion recognition. In order to study the performance improvement, the following experiments were carried out:

In experiment 1 investigated the effect of depth performance on children’s emotion recognition rate. The attention-gate-based LSTM model (hereafter referred to as the attention-gate LSTM model) at Depths 1, 2, and 3 is used for comparison.

In experiment 2, the effects of decreasing the number of parameters and training time on speech emotion recognition performance were investigated. The attention gate LSTM model with depth 1 was compared with the traditional GRU model and LSTM model.

3.3 Training Framework

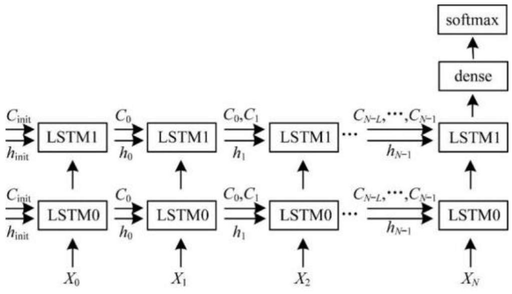


Fig. 2 Training framework of depth of attention gate LSTM model

The training framework for the deep attention gate LSTM model is shown in Fig.2. Among them, LSTM0 represents the LSTM model of the first-level deep attentional gate, and LSTM1 represents the LSTM model of the second-level deep attentional gate. X_t extracts speech features [8-10] from INTERSPEECH in frame t after framing and windowing, $\mathbf{h_t}$ and $\mathbf{C_t}$ are the hidden layer output and cell state of its corresponding LSTM model output. As can be seen from Fig. 2, the input state of the traditional LSTM model at time t is (h_{t-1}, C_{t-1}), while in the training of this paper, the input state at each time is expanded to (h_{t-1}, { C_{t-1}, C_{t-2}, ... ,C_{t-L} }), where L is the depth of the attention gate. The last state of the last layer of LSTM, which contains all the temporal information of the preorder, is input into the subsequent classification network for the identification of children's emotions.

4. Experimental Setup and Analysis

4.1 Experimental Setup

In the experiment, two databases with different emotional representation forms are used to verify the validity of the algorithm in this paper. In order to study the performance of this algorithm in dealing with other types of emotion recognition problems, and whether reducing the number of parameters can optimize the time or reduce the performance, Fau Aibo Children's affective Corpus, infant crying affective needs Corpus and CASIA Chinese affective Corpus were used to verify the results[39].

FAU AIBO EMOTIONAL Corpus for children: Fau Aibo used high performance wireless earphones to collect and record the vocalization of 51 children aged around 10 years old and their electronic pet Aibo during the game, with the most emotionally salient data retained. Natural language contains 48,401 words. In order to ensure the accuracy of the tagging, each sample of the Corpus is auditioned by 5 language majors and then marked by voting. This paper selects five categories of tags defined in the INTERSPEECH 2009 emotional challenge: A (Angry, Touchy, Reprimanding) , E(Emphatic) , N (Neutral) , P(Motherese、Joyful) and R (Rest) .

2) Infant crying emotional needs corpus: Since there is no universal crying emotional needs Corpus in the world, the author collaborated with a hospital in China to record and annotate the audio files of infant crying in five states: Angry, Hungry, Pain, Sad and Tired. In order to improve the quality of the Corpus, the author screened the emotional speech data of infant crying by artificial method, and got rid of the speech-related frames of parents comforting children, and the speech-related frames of two or more babies crying at the same time. The Corpus consists of 10 infants (5 boys and 5 girls respectively), each infant has 20 items in each state, which is 5 × 10 × 20 = 1000 items.

3) CASIA Chinese affective Corpus was recorded by Chinese Academy of Sciences, and the pronunciations of six emotions, namely angry, happy, fear, sad, surprise and neutral, were carried out by four related professionals. There are 9600 Corpora in this corpus.

4.2 Frame Level Feature Selection

The experiment selects part of the frame-level features based on the speech emotion features of INTERSPEECH. In reference [8], 16 low-level descriptors (LLD, zero-crossing rate, root-mean-square Frame Energy, pitch frequency and Mel Frequency Cepstrum Coefficients 1-12) and their difference coefficients are extracted. For each descriptor, 12 statistical functions are calculated, so the total eigenvector has 16 × 2 × 12 = 384 features. On this basis, the speech emotion characteristics of INTERSPEECH 2010(IS2010) have been increased to 38 kinds of LLD, so the total feature dimension has been expanded to 1582 dimensions. The feature dimension of the speech feature set is increased to 6373 dimensions.

4.3 Setting of Experimental Parameters

The original data is divided into two parts, the training set and the test set, which are separated from each other, and the ratio of the training set and the test set is 4:1. The experiments all adopt unidirectional two-layer LSTM stack structure and use one full connection layer and one $softmax$ layer as the training model. During the training, a small batch of gradient descent is used and $Tanh$ is used as the activation function, as shown in Table 1. In order to ensure the validity of experimental comparison, the same Corpus and model experimental parameters are identical.

5. Algorithm Performance Analysis

The traditional LSTM model gets rid of the redundant information through the forgotten gate and gets the new information through the input gate. In this paper, we use the self-attention and the basic structure of LSTM to do self-attention to the cell state, so as to compare the forgetting gate and the input gate of LSTM. At the same time, considering the correlation of time series information, the depth-based self-attention Gates are proposed and compared at the depth of 1, 2 and 3. Four kinds of models were compared: Traditional model, LSTM + deepf_1 model, LSTM + deepf_2 model, and LSTM + deepf_3 model. The depth distributions of these models are 0, 1, 2 and 3, as shown in Fig.3. As can be seen from Fig.3 and Fig.4 , after replacing the forgetting door and the output door of the traditional

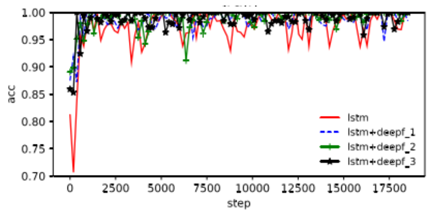


Fig. 3The performance of infant crying emotional needs corpus in different LSTM models

LSTM model with the attention door proposed by using

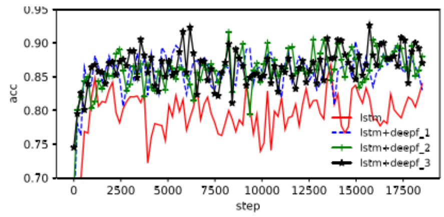


Fig. 4 Performance of different LSTM models using Fau Aibo children's emotion corpus database

the infant crying affective needs corpus and Fau Aibo children affective Corpus, the convergence rate of the attention gate LSTM model on the training set and the test set is much higher than that of the traditional LSTM model, when FAU AIBO is used in children's affective Corpus, the traditional LSTM model starts to converge at about 30000, while the attention gate model starts to converge at about 17000 When the model converges, the attention-gate LSTM model outperforms the traditional LSTM model in the average recognition rate

Table 1. Experimental parameters.

Name of Parameter	Baby crying needs corpus parameter values	Fau Aibo	CASIA
Eta	1e-3 beta2= 0.7	1e-4 beta2= 0.9	1e-4 beta2= 0.9
Batch size	64	128	128
Epochs	1 500	1 200	1 200
Lstm cells	[512, 256]	[512, 256]	[512, 256]
Dense layers	[256, 128]	[256, 128]	[256, 128]
Softmax layers	[128, 5]	[128, 5]	[128, 6]
L2	1e-4	1e-4	1e-4

of children's emotion.

According to the analysis above, the performance of attention gate LSTM model is improved because it modifies the forgetting gate and input gate of traditional LSTM model, it makes the LSTM model leave important information by self-attention to the cell state at the last moment, and supplement the unimportant information as the new input in the corresponding position, thus improving the performance of the LSTM model. The attention gate LSTM model introduces the concept of depth so that each forgetting operation is determined by multiple cell states rather than one of them.

In order to compare the performance of different models in the test set for each kind of emotion, the performance index of the model with the highest recognition rate from the beginning of training to the end of the test set was Quantitative analysis, the performance indicators

obtained from the infant crying affective needs Corpus and Fau Aibo children affective corpus are shown in tables 2 and 3. It can be seen that for the test set, the performance index of the attention gate LSTM model is better than the traditional LSTM model.

From Table 2, it can be seen that the recall rate of attention-gate LSTM model is better than that of traditional LSTM model except that the items of drowsiness and sleepiness are close to each other The F1 scores of attention gate LSTM model were better than those of traditional LSTM model in five kinds of emotion. In the aspect of depth, the performance of the attention-gate LSTM model of depth 3 and depth 2 is close to that of the attention-gate LSTM model of depth 1, except "sad", the recall rate and F1 score of the other 4 items of the model are better than those of depth 1.

From Table 3, it can be seen that when FAU AIBO is used in children's affective corpus, the recall rate and F 1 score of attention gate LSTM model are lower than those of traditional LSTM model except that the e class is lower than that of traditional LSTM model, and the other four items are better than traditional LSTM model. In the aspect of depth, the performance of the attentional gate LSTM model of depth 3 and depth 2 is close to that of the attentional gate LSTM model of depth 1, except for class R, the recall rate and F1 score of the other four items of the attentional gate LSTM model are better than those of depth 1.

Table 2. Performance indicators of different LSTM models using emotional needs corpus of infant crying database.

model	Measure	Angry	Hungry	Pain	Sad	Tired	AVG
sample size	—	40	45	34	35	46	200
LSTM	Recall	0.875	0.911	0.824	0.829	0.957	0.885
	F1 SCORE	0.875	0.901	0.848	0.906	0.889	0.885
LSTM+deepf_1	Recall	0.875	0.889	0.882	1.000	0.935	0.915
	F1 SCORE	0.909	0.941	0.822	0.946	0.945	0.916
LSTM+deepf_2	Recall	0.925	0.978	0.882	0.943	0.957	0.940
	F1 SCORE	0.949	0.957	0.923	0.930	0.936	0.940
LSTM+deepf_3	Recall	0.925	0.978	0.824	0.943	0.957	0.930
	F1 SCORE	0.949	0.926	0.875	0.943	0.946	0.930

Table 3. Performance indicators of different LSTM models using Fau Aibo children's emotion corpus database.

model	Measure	A	E	N	P	R	AVG
sample size	—	611	1 508	5 376	215	546	8 256
LSTM	Recall	0.352	0.373	0.755	0.088	0.081	0.594
	F1 SCORE	0.339	0.373	0.743	0.110	0.093	0.586
LSTM+deep f_1	Recall	0.326	0.300	0.814	0.153	0.119	0.621
	F1 SCORE	0.358	0.338	0.770	0.173	0.133	0.603
LSTM+deep f_2	Recall	0.339	0.289	0.826	0.158	0.081	0.625
	F1 SCORE	0.342	0.340	0.772	0.173	0.104	0.601
LSTM+deep f_3	Recall	0.360	0.327	0.800	0.191	0.095	0.619
	F1 SCORE	0.371	0.363	0.765	0.192	0.112	0.604

It should be noted that the sample size of Fau Aibo Children's affective corpus is uneven, with a maximum of 5,376 samples in n category and only 215 samples in P category. From the above analysis, with the increase of depth, the model can enhance the learning of a small number of samples. Compared with the traditional LSTM model, the recall rate of LSTM + deepf_1model was increased by 5.50% , and the F 1 score was increased by 5.49% , and the recall rate of LSTM + deepf_2 model was increased by 3.14% when Fau Aibo was used, the F1 score of LSTM + deepf_3 model was increased by 1.84% .

6. Conclusion

In this paper, a child speech emotion recognition model based on improved LSTM network is proposed frame-level speech features are used instead of traditional speech features, the attention mechanism is introduced into the forgetting gate and the input gate of the internal structure of the LSTM network model to form the attention gate. The experimental results show that the recognition rate of this model is significantly higher than that of traditional LSTM model, and the recognition rate of depth model is higher than that of shallow model. In CASIA database with other emotions, the training time of this model is shorter than that of LSTM model, and the recognition rate is higher than that of LSTM model and GRU model. The next step is to introduce this model into the fields of speech recognition, machine translation and lie detection, to test and study the continuous affective Corpus and to improve the model for calculating attention scores, to further improve children's speech emotion recognition rate.

References

1. LU Fang, CHEN Guopeng. An overview on the development of children's emotion regulation[J]. Psychological Science, 2003, 26(5): 928-929. (in Chinese)
2. GONG Y, POELLABAUER C.Continuous assessment of children's emotional states using acoustic analysis[C]//Proceedings of 2017 IEEE International Conference on Healthcare Informatics.Park City, USA: IEEE Press, 2017: 171-178.
3. DE SILVA P R, MADURAPPERUMA A P, MARASINGHE A, et al.A multi-agent based interactive system towards Child 's emotion performances quantified through affective body gestures[C]//Proceedings of the 18th International Conference on Pattern Recognition.Hong Kong, China: IEEE Press, 2006: 1236-1239.
4. DAI Weijia.Research on expression and speech bimodal emotion recognition of children[D].Nanjing: Southeast University, 2016.
5. LEDERMAN D, COHEN A, ZMORA E, et al.On the use of hidden Markov models in infants'cry classification[C]//Proceedings of the 22nd Convention on Electrical and Electronics Engineers in Israel.Tel-Aviv, Israel: IEEE Rress, 2002: 350-352.
6. CHANG Chuanyu, LI Jiajing.Application of deep learning for recognizing infant cries[C]//Proceedings of 2016 IEEE International Conference on Consumer Electronics-Taiwan.Nantou County, China: IEEE Press, 2016: 1-2.
7. CHANG C Y, CHANG C W, KATHIRAVAN S, et al. DAG-SVM based infant cry classification system using sequential forward floating feature selection[J]. Multidimensional Systems and Signal Processing, 2017,

28(3): 961-976.
8. SCHULLER B, STEIDL S, BATLINER A.Theinterspeech 2009 emotion challenge[C]//Proceedings of the 10th Annual Conference of the International Speech Communication Association 2009.Brighton, UK: IEEE Press, 2009: 1-7.
9. SCHULLER B, STEIDL S, BATLINER A, et al.The INTERSPEECH 2010 paralinguistic challenge[C]//Proceedings of the 11th Annual Conference of the Inter-national Speech Communication Association 2010.Makuhari, Japan: IEEE Press, 2010: 25-32.
10. SCHULLER B, STEIDL S, BATLINER A, et al.The INTERSPEECH 2016 computational paralinguistics challenge: deception, sincerity and native language[C]//Proceedings of Interspeech 2016.San Francisco, USA: ISCA, 2016: 2001-2005.
11. HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
12. HAN Wenjing, LI Haifeng. A brief review on emotional speech databases[J]. Intelligent Computer and Applications, 2013, 3(1): 5-7.
13. ATHIWARATKUN B, STOKES J W.Malware classification with LSTM and GRU language models and a character-level CNN[C]//Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing.New Orleans, USA: IEEE Press, 2017: 2482-2486.
14. MERITY S, KESKAR N S, SOCHER R.Regularizing and optimizing LSTM language models[EB/OL].(2017-08-07)[2019-10-20].
15. LI Wei, MAK B.Derivation of document vectors from adaptation of LSTM language model[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics.Stroudsburg, USA: Association for Computational Linguistics, 2017: 456-461.
16. LIU Yue, ZHAI Donghai, REN Qingning. News text classification based on CNLSTM model with attention mechanism[J]. Computer Engineering, 2019, 45(7): 303-308.
17. CHO K, VAN M B, GULCEHRE C, et al.Learning phrase representations using RNN encoder-decoder for statistical machine translation[EB/OL].(2014-06-03)[2019-10-20].
18. BRITZ D, GOLDIE A, LUONG M T, et al.Massive exploration of neural machine translation architectures[EB/OL].(2017-03-11)[2019-10-20].
19. SHI Xingjian, CHEN Zhourong, WANG Hao, et al.Convolutional LSTM network: a machine learning approach for precipitation nowcasting[EB/OL].(2015-06-13)[2019-10-20].
20. ZHANG Shiliang, JIANG Hui, WEI Si, et al.Feedforward sequential memory neural networks without recurrent feedback[EB/OL].(2015-10-09)[2019-10-20].
21. ZHANG Shiliang, LEI Ming, YAN Zhijie, et al.Deep-FSMN for large vocabulary continuous speech recognition[C]//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing.Calgary, Canada: IEEE Press, 2018: 5869-5873.
22. TAO Fei, LIU Gang.Advanced LSTM: a study about better time dependency modeling in emotion recognition[C]//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing.Calgary, Canada: IEEE Press, 2018: 2906-2910.
23. XIE Yue, LIANG Ruiyu, LIANG Zhenlin, et al. Speech emotion classification using attention-based LSTM[J]. ACM Transactions on Audio, Speech, and Language Processing, 2019, 27(11): 1675-1685.
24.. TRIGEORGIS G, RINGEVAL F, BRUECKNER R, et al.Adieu features end-to-end speech emotion recognition using a deep convolutional recurrent network[C]//Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing.Shanghai, China: IEEE Press, 2016: 5200-5204.
25. TZIRAKIS P, TRIGEORGIS G, NICOLAOU M A, et al. End-to-end multimodal emotion recognition using deep neural networks[J]. IEEE Journal of Selected Topics in Signal Processing, 2017, 11(8): 1301-1309.
26. BAHDANAU D, HYUN CHO K, BENGIO Y.Neural machine translation by jointly learning to align and translate[EB/OL].(2014-09-01)[2019-10-20].
27. CHOROWSKI J, BAHDANAU D, CHO K, et al.End-to-end continuous speech recognition using attention-based recurrent NN: first results[EB/OL].(2014-12-04)[2019-10-20].
28. LUONG M T, PHAM H, MANNING C D.Effective approaches to attention-based neural machine translation[EB/OL].(2015-08-17)[2019-10-20].
29. VASWANI A, SHAZEER N, PARMAR N, et al.Attention is all you need[C]//Proceedings of Advances in neural information processing systems 2017.Long Beach, USA: IEEE Press, 2017: 5998-6008.
30. LIN Z H, FENG M W, SANTOS C N, et al.A structured self-attentive sentence embedding[EB/OL].(2017-03-09)[2019-10-20].
31. XIE Yue, LIANG Ruiyu, LIANG Zhenlin, et al.

Attention-based dense LSTM for speech emotion recognition[J]. ICE Transactions on Information and Systems, 2019, 102(7): 1426-1429.

32. MIRSAMADI S, BARSOUM E, ZHANG C.Automatic speech emotion recognition using recurrent neural networks with local attention[C]//Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing.New Orleans, USA: IEEE Press, 2017: 2227-2231.
33. GERS F A, SCHMIDHUBER J.Recurrent nets that time and count[C]//Proceedings of Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks.Como, Italy: IEEE Press, 2000: 189-194.
34. HINTON G, VINYALS O, DEAN J.Distilling the knowledge in a neural network[EB/OL].(2015-03-09)[2019-10-20].
35. JACOB B, KLIGYS S, CHEN B, et al.Quantization and training of neural networks for efficient integer-arithmetic-only inference[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.Salt Lake City, USA: IEEE Press, 2018: 1-5.
36. HAN S, MAO H Z, DALLY W J.Deep compression: compressing deep neural networks with pruning, trained quantization and huffman coding[EB/OL].(2015-10-01)[2019-10-20].
37. DAI Dawei, YU Liping, WEI Hui. Parameters sharing in residual neural networks[J]. Neural Processing Letters, 2020, 51(2): 1393-1410.
38. HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
39. PAN Shifeng, TAO Jianhua, LI Ya.The CASIA audio emotion recognition method for audio/visual emotion challenge 2011[C]//Proceedings of Affective Computing and Intelligent Interaction.Berlin, Germany: Springer Berlin Heidelberg, 2011: 388-395.
40. JAITLY N, HINTON G.Learning a better representation of speech soundwaves using restricted boltzmann machines[C]//Proceedings of 2011 IEEE International Conference on Acoustics, Speech and Signal Processing.Prague, Czech Republic: IEEE Press, 2011: 5884-5887.

A Design and Implement of an Automatic Intelligent Car

Chengzhi Liu , Yizhun Peng*,Jikai Zhao , Dezhi Yang,
College of Electronic Information and Automation,
Tianjin University of Science and Technology, Tianjin, 300222,China;
*E-mail: * pengyizhun@tust.edu.cn*
www.tust.edu.cn

Abstract

In order to study the situation that some unmanned vehicles need to drive repeatedly on fixed lines, this article introduces an autonomous smart car based on the STM32 platform. Infrared detectors are used to determine whether the vehicle is driving on a prescribed route. Ultrasonic detectors can prevent the vehicle from colliding with other people or vehicles. Inexpensive cars can help express companies to transport packages and letters to reduce labor costs. Large vehicles can be used to carry large cargo such as containers in docks and other places.

Keywords: Automatic driving, Ultrasonic detectors, Infrared detectors, Parcel shipping

1. Introduction

With the development of online shopping and the impact of the epidemic, contactless delivery of goods is now of great importance. In universities and office buildings, cargo delivery robots are becoming more and more common, but high prices often limit their application scenarios. Recently, the China Post Bureau released the May Express Development Index Report. In May 2020, China’s express delivery development index was 344.2, an increase of 73.5% year-on-year. Among them, the development scale index, service quality index, development capability index and development trend index were 291.1. 622.4. 202.4 and 94.0, respectively. Upgrade, showing a development trend of capacity expansion and quality improvement.

2. Hardware design

2.1. Overall structure of the car

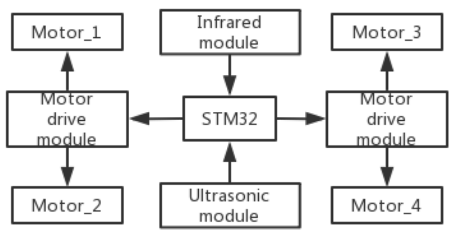


Fig.1.Block diagram of the overall structure

After setting the driving path in the main control of the car, the car will run independently. The ultrasonic module can transmit the signal of whether there are obstacles on the driving path to the car. The infrared module is responsible for the judgment of the driving path of the car, and the driving direction and speed of the car are automatically set by the main control.

2.2. Control section

STM32 microcontroller is used as the main control. The single-chip minimum system includes a display, matrix keyboard, A/D, D/A and other modules, which can significantly reduce the design of peripheral circuits and reduce the difficulty of system design.

The STM32 series microcontroller is a 32-bit microcontroller, based on the ARM Cortex-M3 core specially designed for embedded applications that require high performance, low cost, and low power consumption.

2.3. Drive section

- Step-down module

The step-down module can reduce the power supply voltage to the appropriate voltage required by the module, usually 5v and 3.3v. Therefore LM2596S were chosen DC-DC high-power DC voltage regulator board.

Fig.2.The step-down module

- Motor drive module

The TB6612FNG used in this design is a motor drive IC that can independently control two DC motors in both directions. It has a high degree of integration, and at the same time can provide sufficient output capacity, also has advantages in operating performance and energy consumption. Therefore, it can be used as an ideal motor drive device in an integrated and miniaturized motor control system.

Fig.3.The motor drive module

- Power module.

In terms of power supply, rechargeable batteries, the working principle is the mutual conversion of chemical energy and electrical energy. When the voltage is insufficient, it can be charged to keep the voltage stable, and the voltage can be maintained at a stable value, making the experiment more accurate. Rechargeable

Fig.4.The DC geared motor

batteries can be used multiple times, which is beneficial to protect the environment.In terms of environmental protection and convenience, the power supply module uses rechargeable batteries.

- DC geared motor

In order to enable the car to carry heavier goods and get the speed of the car, a DC geared motor with a reduction ratio of 30 is the better choice.The encoder based on Hall component can return the current speed of the car to the microcontroller to control the car.

2.4. Tracking section

TCRT500 infrared integrated transceiver based on infrared detection method is our choice. The infrared detection method uses the characteristics of different reflection intensities of infrared on the surface of objects of different colors to distinguish the current detection state. The infrared photoelectric tube can greatly reduce external interference. The circuit design is relatively simple, the detection information speed is fast, and the market price is low.

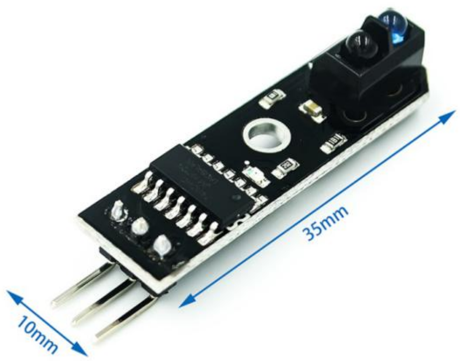


Fig.5.The infrared detection

2.5. Ultrasonic obstacle avoidance sensor

The model of ultrasonic obstacle avoidance sensor used in this car is ULB-1 ultrasonic distance sensor, which has the characteristics of high resolution, high precision and low consumption. Not only in the design but also in the interference noise processing, with anti-noise interference ability. And for the different size of the target and the change of the supply voltage, do the sensitivity compensation. In addition, it also has standard internal temperature compensation, which makes the measured distance data more accurate.

Fig.6.Ultrasonic obstacle avoidance sensor

3. Software design

3.1. Program function description and design ideas

When the car arrives at the loading area, it automatically stops waiting for loading, and the on-board weight detection system will automatically detect whether the weight of the goods exceeds the standard. After the loading is completed, it will drive to the unloading area according to the set path. During the driving process, it can automatically avoid obstacles and drive on large-angle slopes.

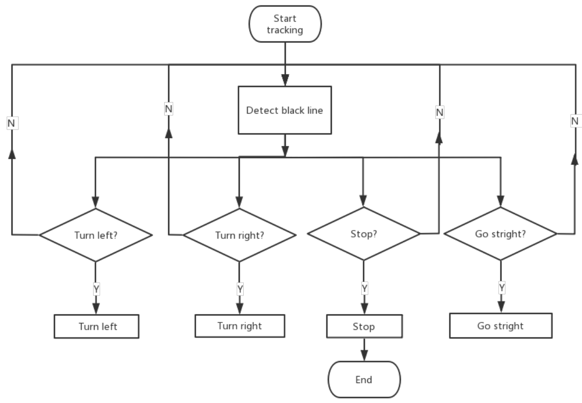


Fig.7.The system flow chart

3.2. infrared sensor

TCRT5000 infrared tracking module is our selected tracking module. The module has three leads: VCC, GND and OUT, and VCC is connected to a +5V power supply. After correct connection, it can be found that when the probe detects the driving track, the indicator light on the rear of the module will remain off, and the output of the OUT terminal will change to low level, and the OUT terminal will change to high level. Normally level.

3.3. DC motor driver

The rotation speed and direction of a motor are controlled by one PWM signal and two control signals. Pin definition of motor drive is shown in Tab.1

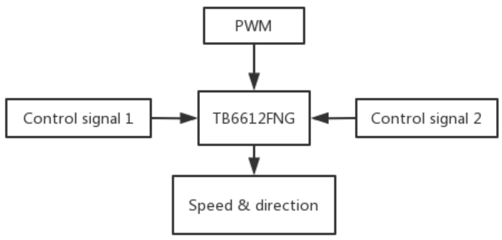


Fig.8.Block diagram of the motor driver

Tab.1.Pin definition of motor drive

Number	Name	Description
1	PWMA	Speed control pin of motor A
3	AIN2	Direction control pin 2 of motor A
4	AIN1	Direction control pin 1 of motor A
5	STBY	Drive enable pin (connect to VCC)
6	BIN1	Direction control pin 1 of motor B
7	BIN2	Direction control pin 2 of motor B
8	PWMB	Speed control pin of motor B
9	VM	Positive motor power supply
10	VCC	Logic power supply positive
11	GND	Logic power supply negative
12	GND	Motor power negative
13	A01	A road motor output 1
14	A02	A road motor output 2
15	B01	B road motor output 1
16	B02	B road motor output 2

4. Testing and conclusion

4.1. Test plan

- Detect the maximum mass that can be carried when the car is given maximum power.
- Lay complex lines on the ground to detect the functions of infrared tracking modules and algorithms.
- Put many cars on the line to run at the same time to detect the ultrasonic obstacle avoidance function.

Fig.9.Line inspection test

4.2. Test results

After multiple tests and repeated experiments, the maximum load of the car is 7.6 kg, and the load is 5 kg when the minimum speed is guaranteed. There is a certain probability of taking the wrong path when driving on a complex route, especially when multiple trajectories intersect in one place.

Fig.10.Weight test

References

1.Research on Application of Automated Warehouse Transportation Path Planning[D]. Teng Zhipeng. Changchun University of Technology 2020

2.Razvan Luca, Fritz Troester, Robert Gall and Carmen Simion,Autonomous parking procedures using ultrasonic sensors,Annals of DAAAM & Proceedings,Jan. 1, 2010 .

3. Wu Guojing, etc. Microcontroller Application Technology China Electric Power Press, 2004.1

4.Design and Research of Mobile Robot Control System Based on Automatic Guidance[J]. Li Peiying, Liu Weiliang. Innovation in Science and Technology. 2020(29)

5.Realization of obstacle avoidance technology for smart car[J]. Li Lei, Cao Shengwei. Micromotor. 2011(06)

6.Design and implementation of ultrasonic anti-collision speed measurement and obstacle avoidance instrument[J]. Su Qiumei, Liu Yarong. Science and Technology Innovation and Application. 2020(29)

Target detection in remote sensing image based on deep learning

Lianchen Zhao [1,2], Yizhun Peng [1,2*], Di Li[1,2], Yuheng Zhang[12]
[1]*College of Electronic Information and Automation,*
Tianjin University of Science and Technology, China

[2]*Advanced Structural Integrity International Joint Research Centre,*
Tianjin University of Science and Technology, China
*E-mail: * pengyizhun@tust.edu.cn*
www.tust.edu.cn

Abstract

For high-resolution optical remote sensing images, there are still many challenges in target detection. In this paper, deep learning algorithm is used to detect the target in remote sensing image. Improve and optimize the deep learning target detection algorithm. When the selected data set is used for target detection, the AP value is improved, which leads to the concept of multi-scale feature fusion feature pyramid and residual network. By improving the selected Yolov3 network model, the detection effect of the two targets of aircraft and ships in remote sensing images has been significantly improved.

Keywords: Deep learning; Target detection; Residual network; Remote sensing image

1. Introduction

For high-resolution optical remote sensing images, there are still many challenges in target detection. This paper introduces the multi-scale target detection in optical remote sensing images by using Yolov3[1], and compares the detection results of multi-scale targets on the selected remote sensing data sets by using the improved Yolov3 model.

Target detection is an important direction of computer vision and digital image processing[2]. Target detection is also widely used in robot navigation, real-time vehicle monitoring, defective product detection in industry, aerospace ship detection and other target detection. Through the algorithm of target detection in deep learning, we can realize the accurate detection of the target, which not only has a great reduction in manpower, but also has a high efficiency improvement, which has important practical significance. Target detection algorithms in deep learning can be divided into two categories: one-stage algorithm and two-stage algorithm[3].

2. One-Stage target detection algorithm

One-Stage, a target detection algorithm based on deep learning, has a faster detection speed when detecting the target[4], because the target detection algorithm discards the process of single region recommendation. This algorithm is first proposed from CVPR 2016, and Yolo (you only look once: unified, real time object detection) is an innovative one stage detection.

The method of generating candidate regions first and then detecting has relatively high accuracy, but it is relatively slow in terms of detection speed.

Yolo did not remove the candidate regions[5], but directly divided the input network image into 49 7 * 7 grids. Any existing grid predicted two boundary boxes, so as to achieve the prediction of 98 boundary boxes. It can be roughly understood as a rough selection of 98 candidate regions on the input image, which cover the entire region of the image. Therefore, regression prediction is used to compare the existing 98 candidate boxes to get the final applicable boundary box.

Yolo network draws lessons from the structure of GoogleNet[6] classification network. Compared with

GoogleNet classification network, the 1x1 convolution layer and 3x3 convolution layers are used to replace the inception module in Yolo. There are 24 convolution network layers and 2 full connection layers in the whole detection network. Compared with the previous version of Yolo, Yolov3 adjusts the network structure, and uses the Darknet-53[7] network structure compared with Yolov2. The specific network structure is shown in Figure 1 below:

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Fig. 1. Darknet-53 network structure

In Figure 1, the main structure of the network structure of darknet-53 mainly consists of DBL module, Upsample module, Shortcut module, Res module and Route module. The main components of DBL are convolution network, BN and Leaky relu.

The size of network input image is 416 * 416 * channels. After 5 times of down sampling and 2 times of up sampling stitching through convolution layer, three kinds of feature maps will be output. There are 53 layers of convolutions in the darknet-53 network. Except for the last FC, there are 52 convolutions as the main network model structure. For the low-level convolution layer, the field of vision is relatively small, which is responsible for the detection of small targets. For the deep convolution layer, the field of vision is relatively large, which is responsible for the detection of larger targets.

The activation function adopted by yolov3 is leaky relu. Compared with relu, this activation function sets all negative values to zero. Leaky relu function only gives a non-zero slope to all negative values. The mathematical expression is shown in (1)：

$$y_i = \begin{cases} x_i & if \quad x_i \geq 0 \\ \dfrac{x_i}{a_i} & if \quad x_i < 0 \end{cases} \tag{1}$$

The leaky relu function image is shown in Figure 2:

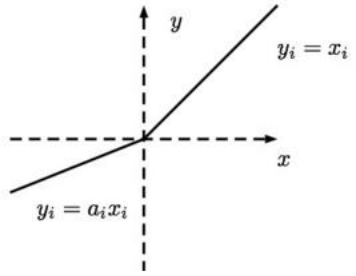


Fig. 2. Leaky Relu function graph

In Yolov3, anchor boxes are needed in bounding box prediction. In Yolov3, K-means clustering algorithm is used to obtain anchors suitable for data sets. The goal of K-means algorithm is to divide n samples into K clusters according to the calculation results[8], so that the similar samples in n samples can be divided into the same cluster. The calculation method used to measure the similarity is to use the size of Euclidean distance (2).

$$d(x,y) = \sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\cdots+(x_n-y_n)^2} = \sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \tag{2}$$

The loss function in Yolov3 is shown in formula (3) as follows:

$$\begin{aligned} Loss = {} & \lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{ij}^{obj}[(x_i-\hat{x}_i^j)^2+(y_i-\hat{y}_i^j)^2]+ \\ & \lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{ij}^{obj}[(\sqrt{w_j^i}-\sqrt{\hat{w}_j^i})^2+(\sqrt{h_j^i}-\sqrt{\hat{h}_j^i})^2]- \\ & \sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{ij}^{obj}[\hat{C}_i^j\log(C_i^j)+(1-\hat{C}_i^j)\log(1-C_i^j)]- \\ & \lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{ij}^{noobj}[\hat{C}_i^j\log(C_i^j)+(1-\hat{C}_i^j)\log(1-C_i^j)]- \\ & \sum_{i=0}^{S^2} I_{ij}^{obj}\sum_{c\in classes}[\hat{P}_i^j\log(P_i^j)+(1-\hat{P}_i^j)\log(1-P_i^j)] \end{aligned} \tag{3}$$

In the above formula, I_{ij}^{obj} is used to determine whether the j anchor box in the i grid is responsible for the current object. If $I_{ij}^{obj}=1$ is responsible, if $I_{ij}^{obj}=0$ is not. $\hat{C}_i^j$ parameter is the confidence degree, $\hat{C}_i^j$ is the real value in the training process. The value of this parameter depends on whether the bounding box of grid cell is responsible for the prediction of an object. If it is responsible for $\hat{C}_i^j=1$, otherwise, $\hat{C}_i^j=0$.

Central coordinate error:

$$\sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{ij}^{obj}[(x_i-\hat{x}_i^j)^2+(y_i-\hat{y}_i^j)^2] \tag{4}$$

The formula of wide coordinate error in calculation:

$$\sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{ij}^{obj}[(\sqrt{w_j^i}-\sqrt{\hat{w}_j^i})^2+(\sqrt{h_j^i}-\sqrt{\hat{h}_j^i})^2] \tag{5}$$

Confidence error:

$$\sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{ij}^{obj}[\hat{C}_i^j \log(C_i^j)+(1-\hat{C}_i^j)\log(1-C_i^j)]- \\ \lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{ij}^{noobj}[\hat{C}_i^j \log(C_i^j)+(1-\hat{C}_i^j)\log(1-C_i^j)] \tag{6}$$

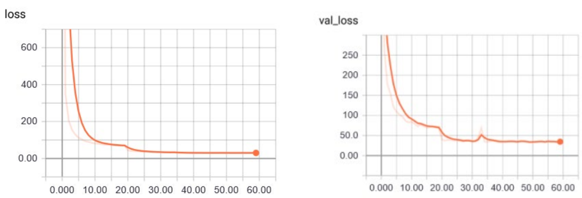
3. Experimental results and analysis before improving the model

3.1. Experimental process

The target detection process is as follows:
(1) The experimental data set is preprocessed, and the large-scale remote sensing image is clipped;
(2) Determine whether the remote sensing image is a single channel image, and select the appropriate defogging processing method;
(3) Remove the objects that are not interested in the remote sensing data set, and keep the whole data set only contain the two kinds of objects of interest in this paper;
(4) The processed remote sensing data are input into the network model for training, and the training model is obtained;
(5) After training, the target is detected and classified.

3.2. Analysis of experimental results

The changes of loss value and val_loss value during training are shown in Figure 3 below:



(a) (b)

Fig. 3. Changes in loss during training

According to the change of loss value in Figure 3 (a), it can be seen that the loss value changes greatly in the first stage, and gradually decreases and tends to be stable in the second stage. According to the change of loss of verification set in Figure 3 (b), it can be seen that the change of loss value tends to decrease during the training process.
After training, you will get trained_weights_final.h5 file is used to test the test set, set the IOU value to 0.5, and then get the final evaluation result of the training result according to the test results. Since the data set used in this experiment is to integrate part of the data in NWPU VHR-10 remote sensing data set and part of the data in DOTA-v1.5 remote sensing data set, and only contains two types of targets, the detection results are shown in Figure 4, and the test evaluation is shown in table 1:

Table 1. Evaluation of test set experiment results

Target category	Precision	Recall	AP
plane	0.64	0.93	83.83%
ship	0.71	0.43	40.02%

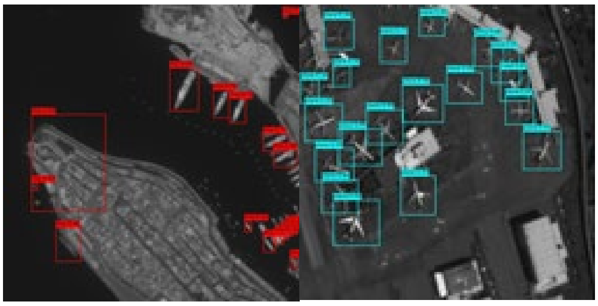


Fig. 4. Test result display diagram

According to the above table 1, the AP value of aircraft is 83.83%, the AP value of ship is 40.02%, and the mAP of two types is 61.93%. From the recall rate, we can see that the recall value for ships is smaller than for aircraft, but the detection accuracy for these two kinds of targets is not high. From the above table, it is not difficult to see that the detection effect is better for large targets such as aircraft, but poor for targets with small ship scale.

4. Experimental results and analysis after improving the model

In this paper, we consider the scale problem to increase the size of the feature map to retain more semantic information on the resolution. By increasing the scale of the input image, more semantic information can be retained. In this paper, the size of the input image is modified to 608 × 608 and 672 × 672.
The prior frame was originally proposed because of Faster R-CNN[9]. According to the size of the ground_truth mark box in the training set, the width and height of the frequently appeared label box are counted, and these mark boxes are taken as the prior box. In order to make the prior frame suitable for the remote sensing data set used in this experiment, therefore, K-means clustering operations are carried out on the dataset for several times.

We use ten times clustering to get ten groups of prior frames, and get the most suitable prior frame by means of average value. As shown in Table 2 below:

Table 2. A priori box improvement

Characteristic map	13*13	26*26	52*52
Perception vision	large	medium	small
Initial prior box	(116x90) (156x198) (373x326)	(30x61) (62x45) (59x119)	(10x13) (16x30) (33x23)
Clustering mean prior frame	(49x37) (65x67) (125x123)	(31x40) (35x26) (43x30)	(14x27) (21x23) (24x23)

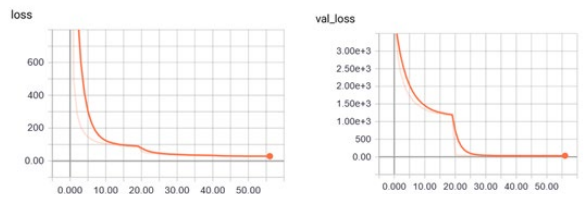
For IOU, IOU is the intersection ratio of detection frame and real frame. In case of non-intersecting border, the gradient will become 0 and cannot be optimized. But for GIOU, this situation can be avoided, so GIOU will be used instead of IOU in this paper. The formula is as follows (7):

$$GIOU = IOU - \frac{|C \setminus (A \cup B)|}{|C|} \tag{7}$$

The loss function of GIOU is shown in formula (8):

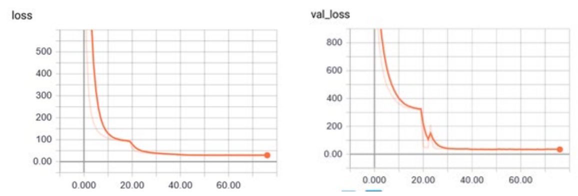
$$L_{GIOU} = 1 - GIOU \tag{8}$$

The evaluation index of the improved model is the same as the evaluation index in the previous section. After the improvement of the model, the loss value of the training set constructed by the partial data of NWPU VHR-10 data set and DOTA-v1.5 data set is constantly decreasing.



(a) (b)

Fig. 5. Change of loss value of input image scale 608×608 after network improvement



(a) (b)

Fig. 6. Change of loss value of input image scale 672×672after network improvement

As shown in Figure 6, the change trend of the loss value in the network training process is basically consistent. Compared with the input image 608 × 608, the change of the loss value of the verification set with the input image scale of 672 × 672 fluctuates, but the overall trend is that the loss value is gradually reduced.

Combined with the change of prior box, the model is changed to input scale of 672 × 672 and 608 × 608. The number of layers is changed from 52 layers to 56 layers due to the addition of 1 to the layers repeated 8 times in convolution layer. With the increase of input scale and the deepening of network layers, not only more semantic information is retained, but also more semantic information can be extracted in feature extraction.

Table 3. Comparison of improved network experiment results mAP

network model	Ship AP	Plane AP	mAP
Yolov3	38.60%	86.77%	61.93%
Yolov3_608	42.85%	86.48%	64.69%
Yolov3_672	40.07%	88.94%	64.84%
Yolov3_608_GIOU	40.02%	83.83%	61.93%
Yolov3	38.60%	86.77%	61.93%

Table 4. The improved network Precision and Recall comparison

network model	Ship Precision	Ship Recall	Plane Precision	Plane Recall
Yolov3	0.814	0.411	0.85	0.92
Yolov3_608	0.82	0.448	0.865	0.918
Yolov3_672	0.806	0.433	0.838	0.939
Yolov3_608_GIOU	0.711	0.438	0.648	0.935

The Yolov3_608_GIOU and Yolov3_672_GIOU versions in the above table refer to the replacement of IOU with GIOU based on the Yolov3_608 and Yolov3_672 versions. Due to the poor detection effect of Yolov3 on small targets, the recall rate is not very ideal. It can be seen from the above table that the recall rate of ships with more small targets has been improved in several improved versions, and the precision will not be greatly reduced. According to the analysis of the mAP value, only the mAP value of the improved network Yolov3_608_GIOU is lower than that of the initial network by 0.75%, and the detection effect of the improved network Yolov3_672_GIOU is the best, and the map value is 3.26% higher than that of the initial network.

5. Conclusion

This chapter first introduces the characteristic pyramid and the loss function of Yolov3, and then optimizes the Yolov3 network. Firstly, K-means clustering method is used to cluster according to the training set, and then the average value is used as the adjustment of prior frame. In order to retain more semantic information, adjust the scale of input data, extract more semantic information and increase the depth of the network, Yolov3 is optimized according to the data set in this paper. Finally, IOU is replaced by GIOU. Compared with several improved networks, the detection effect of the improved network on the test set is improved. Finally, the detection effect of the improved network Yolov3_672_GIOU is the best, and the mAP value has been significantly improved.

References

1. Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection [M]. 2016 IEEE Conference on Computer Vision and Pattern Recognition. 2016:779-88.
2. Felzenszwalb P F, Girshick R B, Mcallester D, et al. Object Detection with DiscriminativelyTrained Part-Based Models[J]. IEEE Transactions on Software Engineering, 2010, 32(9):1627-1645.
3. Chen L-C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolutionfor semantic image segmentation[C]. Proceedings of the European Conference on ComputerVision (ECCV). 2018: 801 – 818.
4. Liu Chang, Wang Pengjun, Zhang Meiling, et al Research on sparse video detection technology based on IOU analysis [J]. High technology letters, 2019(10).
5. Zhang R, Yao J, Zhang K, et al. S-cnn-based ship detection from high-resolution remote sensing images. [J]. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 2016, 41.
6. Walk S, Majer N, Schindler K, et al. New features and insights for pedestrian detection[C].IEEE Conference on Computer Vision and Pattern Recognition, 2010: 1030-1037.
7. Ok A O, Başeski E. Circular oil tank detection from panchromatic satellite images: a new automated approach[J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(6): 1347 –1351.
8. Li Weijun. Overview of K-means clustering algorithm [J]. Modern computer (Professional Edition), 2014 (8): 31-32
9. M. Kang, X. Leng, Z. Lin, et al. A modified faster R-CNN based on CFAR algorithm for SAR ship detection[C]. 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP), Shanghai, China, 2017, 1-4.

A design of home smart nursing robot based on Raspberry Pi

Yuqi Zhao, Yizhun Peng*, Jie Liu, Xiaowei Wu, Jikai Zhao
College of Electronic Information and Automation,
Tianjin University of Science and Technology, Tianjin, 300222, China;
*E-mail: *pengyizhun@tust.edu.cn*
www.tust.edu.cn

Abstract

This product is based on Bluetooth, WIFI and other wireless technologies, with the Raspberry Pi 4 as the core, to meet the health, safety and entertainment needs of the serviced users, mainly the elderly, while assisting the leisure and entertainment of users when traveling outdoors. A multi-functional escort robot designed for safety and other issues. It has many methods such as temperature and humidity monitoring, harmful gas monitoring, and noise monitoring to ensure the safety of users. At the same time, it also has TV projection, audio-visual entertainment touch screen interaction, mobile phone projection and other entertainment functions.
Keywords: Pension; artificial intelligence; entertainment; convenience; accompany; Internet of things; big data

1.Introduction

With the rapid increase of the elderly population, the issue of elderly care has become increasingly prominent, especially for the elderly in empty nests. It has become the norm for the elderly to live alone at home while their children are at work. At the same time, in recent years, the nursing industry such as "monthly sister" has not only experienced frequent problems but also increased industry prices. The home intelligent nursing robot is specially designed for the elderly. It has multiple functions such as health monitoring, smart medicine cabinet, housekeeping services, communication entertainment, abnormal situation alarm, critical situation handling, and action interaction. It is very suitable for the elderly living alone[1].

This home care intelligent robot called the spouse, just like its name, is an intelligent robot specially designed and developed for the elderly living alone. In order to realize the function of accompany the elderly indoors and outdoors, it is based on wireless technologies such as Bluetooth and WIFI. To meet the entertainment needs and health and safety monitoring of the elderly, it can also solve the problems of the elderly in travel, such as loading, rest, health, etc., in all aspects.

2.System overall design ideas

2.1 Overall functional framework

The system includes three components: an indoor body, a physical sign monitoring bracelet, and an outdoor body. It combines Bluetooth communication technology, network communication technology, Internet of Things technology, GPS positioning technology, and intelligent path planning technology. Among the three components, the physical sign monitoring bracelet can monitor the elderly's blood oxygen, pulse, heart rate and other physical sign data and GPS location in real time[2], and realize two-way data exchange through Bluetooth and embedded systems. Through the wireless data network, indoor and outdoor body networking can be realized, and network entertainment resources can be obtained. The outdoor subject can integrate the camera, sensor, GPS coordinates of the bracelet and other information to realize intelligent path planning.

When the user is indoors, the physical sign monitoring bracelet will send the user's physical sign information to the indoor subject via Bluetooth, and calculate the user's health information after processing by the host system algorithm. When the system determines that the user's physical signs are abnormal, it will send the user's

physical signs to the settings through the network The first contact sends a warning carefully and calls 120 when necessary. At the same time, the indoor main body can remind the user to take medicine on time to prevent the elderly from forgetting. The indoor main body can also monitor whether the living environment is safe and issue an alarm in real time while working, so as to avoid the loss of life and property caused by accidents such as fire and gas leakage. The double safety protection of the user's body and living environment is realized; the indoor host is equipped with a large LCD screen specially designed for the elderly, which can be used for entertainment by the elderly without excessively adversely affecting the eyes[3].
When the user is outdoors, the physical sign monitoring and GPS system work as usual, and the outdoor subject will follow the user, and all the user needs to do is act as usual. The outdoor part is box-shaped with storage space inside, and users can use it as a mobile basket. In addition, the box can be used as a seat for rest after fatigue. The outdoor part serves as a self-moving storage space, a movable seat, and can also achieve autonomous climbing of stairs through wheels with a peculiar mechanical structure. It greatly facilitates the lives of elderly people with limited mobility.

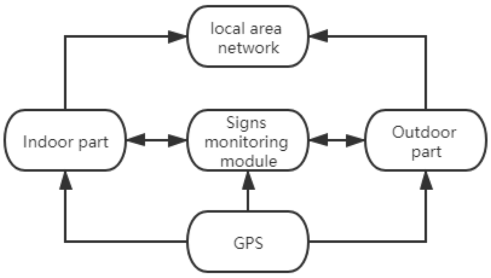


Fig.1 System Block Diagram]

2.2 Vital Sign Monitoring Bracelet Design Scheme

Body temperature detection: Because it is a bracelet to measure temperature, and according to the accuracy requirements, the method of contact temperature measurement is used to measure body temperature. The subsequent influence of the external temperature and the difference between the temperature of the wrist and the underarm temperature are calculated and compensated. So as to get a reliable body temperature. The LMT70 temperature sensor can be used as a low-power, high-precision sensor for temperature acquisition, and feedback to the microprocessor through analog-to-digital conversion. Further calculations are carried out by fitting compensation curves and calculation formulas of outdoor temperature, armpit temperature and wrist temperature in different environments

Blood pressure and heart rate monitoring: The photoelectric sensor can collect the pulse wave at our wrist. By analyzing the pulse wave's rising slope and band time and other characteristic parameters, a specific calculation formula can be obtained to estimate the blood pressure value. For example, when the blood pressure is high, the slope of the rising band of the ejection period will increase. This is also the method adopted by most bracelets on the market, because it is simple and convenient, and will not cause trouble to users[4].
At the same time, the heart rate can be calculated through a certain algorithm. Nowadays, heart rate bracelets basically use light-sensing heart rate, which is tested by the principle of light reflection. Generally, they are affected by ambient light. Some products do not have heart rate data when facing strong light or completely black. Generally, green (OSRAM) is used with a photosensitive photodiode to detect the flow of blood flowing through the wrist at any time. When the user's heart beats, more blood will flow through the user's wrist, and the greater the amount of green light absorbed. Between the beating of the heart, blood flow is reduced, resulting in a reduction in the absorption of green light. The heart rate signal is effectively monitored and extracted through complex and sophisticated algorithms, most of which use the original NXP heart rate.
Based on the secondary development of the ready-made bracelet, integrated GPS module, easy to coordinate with the "wife" to complete positioning and tracking, built-in heartbeat detection, body temperature detection module, based on the Internet connection with the Raspberry Pi, collect the physical signs of the elderly at all times, by the Raspberry Pi Perform data review, and automatically send alarm and first aid information to children's mobile phones when physical signs of dangerous data appear[5].

Fig.2 Bracelet diagram

2.3 Interior design plan

The product integrates the HR202 humidity detection module and the DS18B20 temperature measurement module. By connecting the data to the core processor, it forms an important part of the indoor monitoring and alarm system.

The gas-sensitive material used in the MQ135 gas sensor is tin dioxide (SnO2) with low conductivity in clean air. When there is pollutant gas in the environment where the sensor is located, the conductivity of the sensor increases with the increase in the concentration of pollutant gas in the air. Using a simple circuit, the change in conductivity can be converted into an output signal corresponding to the gas concentration. The MQ135 sensor has high sensitivity to ammonia, sulfide, and benzene vapors, and is ideal for detecting smoke and other harmful gases.

The MAX9814 high-sensitivity microphone sensor module integrated on the product can collect the sound in the environment and send it to the core processor, and then identify the abnormal sound in it, and then can judge the outdoor noise or the abnormal sound in the room (water leak, air leak, people Lock picking, etc.) and further processing as appropriate.

A unique sensor, the thermal imaging module IFD-X, is installed on the product. It monitors the temperature changes in the body of the elderly in real time and transmits the data to the Raspberry Pi for review. If data fluctuations are found to exceed the warning line, warning and detection information can be issued For children, on the other hand, it solves the physical changes that cannot be monitored by the bracelet, detects emergency situations one step in advance, and adds a stronger protection to the lives of the elderly.

IFD-X thermal imaging module, Red Eye Camera (hereinafter referred to as "REC" or "IFD-x" or "equipment") is a non-contact thermal imaging instrument based on infrared array high-precision temperature sensors and advanced software algorithms. Infrared imaging of any object within the field, the imaging resolution is up to 512*384 pixels, the temperature sensitivity is 0.1℃, the absolute accuracy is ±1.5℃, and the refresh frequency is up to 64Hz. Comes with storage and real-time clock, with real-time data output display, photo storage function, digital interface including UART and USB, can be directly connected to the computer and mobile phone, with the upper computer software or mobile phone APP program, it is very convenient to use[6].

This product is connected with a large-screen TV projection module, a touch screen interactive module, a mobile phone projection module, etc., which can meet various entertainment activities and external contacts in the daily life of the elderly. The details will not be repeated.

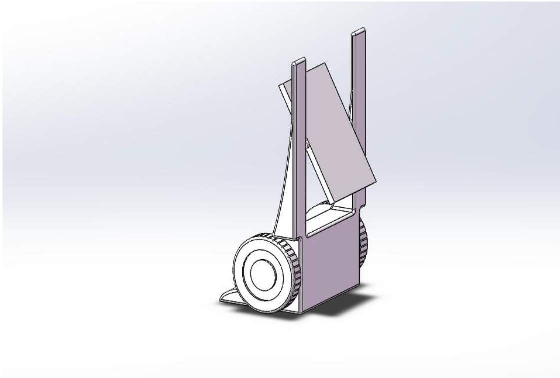


Fig.3 Conceptual drawing of indoor part

2.4 Outdoor part design scheme

This product has the ability to exercise autonomously, and a large-capacity cabinet is reserved in the design, which can help the elderly transport vegetables when the user is out shopping, reducing the burden on the user and freeing hands. The product comes with a buffer module and a shock-proof module. Even if fragile dishes (such as eggs) are placed in the product, it will not break under most roads in the city. At the same time, the product is equipped with an auxiliary device for climbing stairs, which can assist users to climb stairs and reduce the difficulty of users carrying dishes upstairs. The materials are all made of high-quality compressive materials, and the design has a load-bearing structure. The maximum load-bearing capacity is 100kg. The user can sit directly on the product and rest. At the same time, the lid and handle are designed according to ergonomics. Even if sitting for a long time, there will be no back pain . At the same time, it will not exert any pressure on the inner basket. At the same time, the product also has a radio function for users to listen to and relax in their leisure time[7].

The content of intelligent path planning to achieve precise follow is as follows: The positioning method of our core design is based on visual positioning and sensor positioning, each has its advantages and disadvantages. The advantages of sensor positioning are:

1. Able to find the x.y, z coordinates of the target
2. It can be positioned in 360 degrees
3. The positioning target is less affected by obstacles

Based on visual realization, the method is richer, and a variety of technology combinations can be used. For example, based on the depth camera to directly recognize the combination of human bones + face recognition + feature mark recognition. Atlas robots like Boston

Dynamics are based on visual recognition to achieve walking and carrying objects. The advantages of visual recognition targets are: 1. Not only can the x.y.z coordinates be calculated, but also the 3-dimensional deflection angle of the object relative to the camera. Can get richer decision-making information. 2. The visual unit can be used not only to identify targets, but also to identify most obstacles. In summary, we adopt a comprehensive positioning method of vision and sensing to obtain the best positioning effect. The sensor can be an infrared thermal imaging sensor, and the GPS function of the bracelet is used for positioning to determine the location of the owner. At the same time, three different positions on the product are installed with triangular 8266 modules. If the user chooses to use the mobile phone on his body, connect the three modules on the suitcase to this hotspot, and continuously obtain the current rssi value, and use this value to judge the distance between this module and the mobile phone. The three modules are sufficient to judge the mobile phone’s Direction, so as to achieve the purpose of following the user more accurately.

After the human body is recognized, the next step is generally to think of how to recognize the obstacle. Obstacle recognition, commonly used technologies include depth camera recognition, ultrasonic ranging, and infrared ranging. The advantages of depth camera and infrared ranging are that they are cheap and fast, but they cannot identify glass and black objects. Ultrasonic distance measurement can be used as a supplement. If the cost and size are not limited, lidar and millimeter multi-radar can also be considered. We use depth camera recognition, ultrasonic ranging, and infrared ranging to avoid obstacles.

Dynamic path planning, this product uses to build a 3D spatial map and rasterize the map into small grids that can or cannot be passed. With the aid of path planning algorithm, and then follow the target. The difficulty of the whole scheme is that the targets and robots are dynamic and need to be adjusted continuously to achieve optimal path planning. The camera is a more important part of the robot, this product uses the openmv high-definition camera.

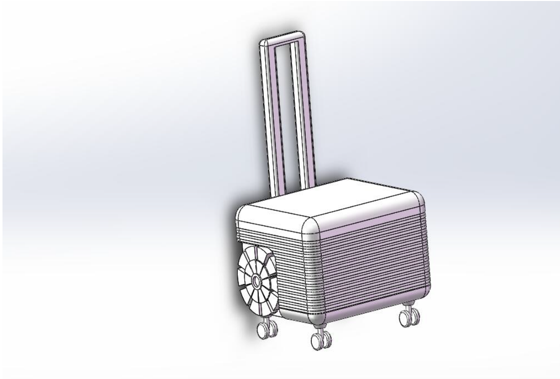


Fig.4 Concept map of outdoor part

3.System software design

3.1Software design of the interior part

The software design of the indoor part includes three parts: the communication between the embedded device and the mobile phone client, the driving reading of the embedded device to various sensors[8], and the communication between the embedded device and the physical sign monitoring bracelet.

1.Drive readings of various sensors by embedded devices:

The core processor of the embedded system adopts the stm32f103 series single-chip microcomputer of STM Company to realize the driving and reading of the sensor. The initialization of each sensor needs to be configured by the microcontroller according to different working principles. Sensor reading generally includes ADC reading and serial communication (including SPI, IIC, etc.).

2.The communication between the embedded device and the physical sign monitoring bracelet: send data via Bluetooth. Bluetooth HC05 is a master-slave Bluetooth serial port module. When the Bluetooth device is successfully paired and connected, the internal communication protocol of the Bluetooth can be ignored and the Bluetooth can be used as a serial port. When a connection is established, two devices share a channel, that is, the same serial port. One device sends data to the channel, and the other device can receive the data in the channel. hc-05 supports standard baud rates from 4800bps to 1382400bps. The baud rate must be matched when using it. Connect vcc to 3.3 or 5V GND to GND TX to RX on the microcontroller and RX to TX.

3.Feedback the user's health information to the mobile phone:

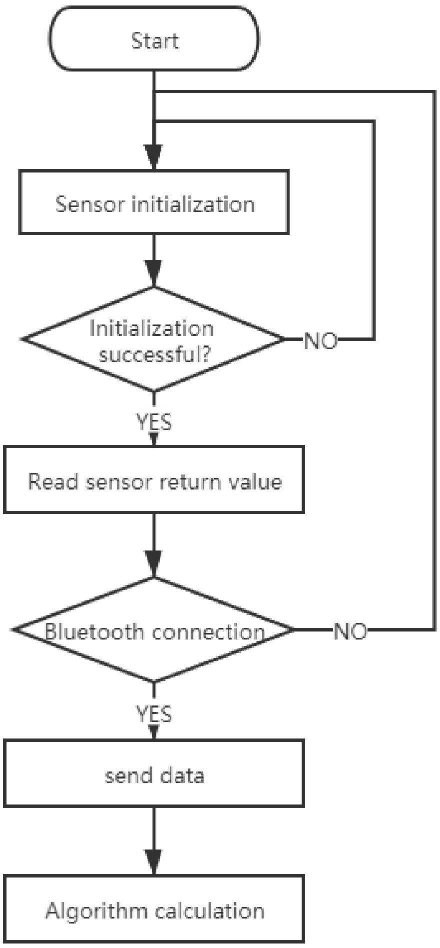


Fig.5 Feedback health information

3.2Sign detection software design

1) Calculate the heart rate through a certain algorithm. The heart rate bracelet basically uses light-sensing heart rate. It is tested by the principle of light reflection. It is generally affected by ambient light. Some products are not exposed to strong light or all black. Heart rate bracelets with heart rate data are generally green (OSRAM) with a photosensitive photodiode to detect the flow of blood flowing through the wrist at any time. When the user's heart beats, more blood will flow through the user's wrist, and the greater the amount of green light absorbed. Between the beating of the heart, blood flow is reduced, resulting in a reduction in the absorption of green light. The heart rate signal is effectively monitored and extracted through complex and sophisticated algorithms, and the original phase heart rate of NXP is used.

2) MAX30100 is a sensor that can read heart rate and blood oxygen. The communication method is through IIC. Its working principle is to obtain the ADC value of the heart rate through infrared LED light irradiation. It is a non-invasive integrated pulse oximetry and heart rate monitoring sensor solution[9].

3.3 Outdoor software design

There are many commonly used computer vision solutions, such as binocular vision, TOF-based depth cameras, and structured light-based depth cameras. The outdoor part of our current intelligent escort robot mainly works in an outdoor environment. The active light source will be greatly affected by conditions such as sunlight. Therefore, the passive vision solution of binocular vision is more suitable. Therefore, the vision solution we adopt is based on dual vision. Visual.

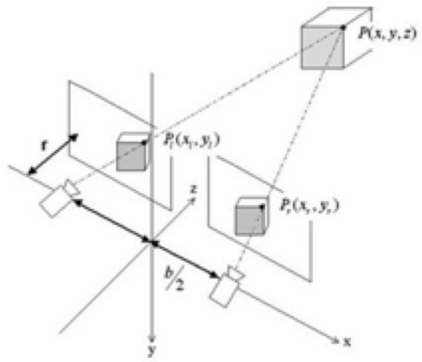


Fig.6 Binocular vision example image

The distance measurement of binocular vision is essentially a triangular distance measurement method. Because the positions of the two cameras are different, just like our two eyes, we see different objects. The same point P seen by the two cameras will have different pixel positions during imaging. At this time, the distance of this point can be measured by triangulation. Different from the structured light method, the points calculated by the structured light are actively emitted and known and determined, while the points calculated by the binocular algorithm are generally image features captured by the algorithm, such as SIFT or SURF features. The feature is calculated as a sparse image. In order to avoid obstacles well, we need to obtain a dense point cloud image and the depth information of the entire scene. This depth information can help us find walkable areas and obstacles in the map scene. The whole output is similar to the 3D point cloud image output by Lidar.

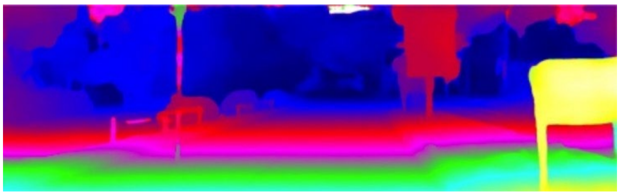


Fig.7 Actual output depth map, different colors represent different distances

At the same time, we use the vector field histogram (VFH) algorithm to achieve accurate obstacle avoidance and follow: during its execution, a local map based on polar coordinates is created for the current surrounding environment of the mobile robot. This local uses the grid map representation method , Will be updated by some recent sensor data. The polar coordinate histogram generated by the VFH algorithm is shown in the figure:

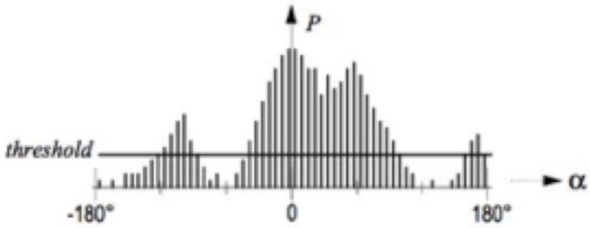


Fig.8 VFH algorithm polar coordinate histogram

In the figure, the x-axis is the angle of the obstacle perceived by the robot as the center, and the y-axis indicates the probability p of the obstacle in this direction. In the actual application process, the histogram will first identify all gaps that are large enough to allow the robot to pass, and then calculate the cost function for all these gaps, and finally select the path with the lowest cost function to pass.

The cost function is affected by three factors: the target direction, the current direction of the robot, and the previously selected direction. The final generated cost is the weighted value of these three factors. The robot's selection preference can be adjusted by adjusting different weights. The VFH algorithm also has other extensions and improvements. For example, in the VFH+ algorithm, the limitations of robot kinematics are considered. Due to the difference in the actual underlying motion structure, the actual motion capability of the machine is limited. For example, the structure of a car cannot be turned in place at will. The VFH+ algorithm considers the blocking effect of obstacles on the trajectory of the robot's actual movement ability, and shields those trajectories that are not occupied by obstacles but cannot be actually reached because of their blocking. Our E-patrol robot uses a two-wheel differential drive motion form, which is very flexible, and its practical application is less affected by these factors. You can look at this icon for details:

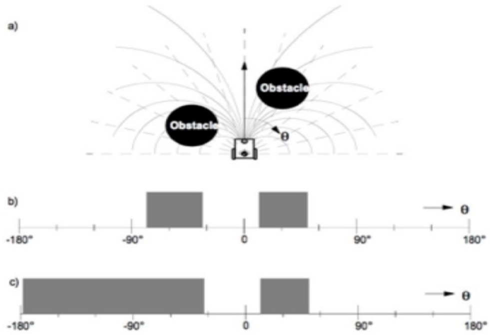


Fig.9 Influencing factors of cost function

4.Concluding remarks

Based on the serious aging problem in our country, the problem of elderly care has become increasingly prominent. The research and development of this robot fills the gap in the market, and to a certain extent alleviates the inconvenience of living alone for elderly people. It is a "wife" invented based on the design concept of serving the society and solving practical social problems. Innovative products of the times that are truly designed for the elderly. It is of great significance to solve social problems such as the difficulty of providing for the elderly in an aging society, the generally low quality of life of the elderly, and the unattended elderly in empty nests. With its comprehensive and humanized functions as its advantages, the home care intelligent escort robot will surely be recognized and supported by the society.

Acknowledgements

This research was supported by Student’s Platform for Innovation and Entrepreneur Training Program, the Ministry of education of China(202010057018).

References

1. Zhang Wenhui, Qi Naiming, Yin Hongliang. Adaptive neural variable structure robot trajectory tracking control [J]. Control and Decision. 2011 (04)
2. Yu Ting. Research on multi-robot queue curve movement [D]. Shanghai Jiaotong University 2009
3. Chen Canfeng. The principle and application of Bluetooth low energy technology [M]. Beijing: Beihang University Press, 2013.
4. Shen Leixian, Pang Jiayi, Zhang Qingyun, Xu He. Design and implementation of positioning service platform based on Beidou navigation[J]. Computer Technology and Development, 2019, (01): 1-4.

5. Chen Kaixuan, Xie Haitao. GPRS principle and its application[J]. Railway Communication Signal, 2003, 39(7): 7-8.
6. Wu Guojing, etc. Microcontroller Application Technology China Electric Power Press, 2004.1
7. Robot path planning Ren Chunmingdi, Zhang Jianxun1 (1. School of Information Technology and Science, Nankai University, Tianjin 300071; 2. Department of Information Science and Technology, Tianjin University of Finance and Economics, Tianjin 300222)
8. Human Health Index-Ling Yisheng [2] Wang Yongyan, Sun Yiwen
9. Wu Xiaofang. New Thoughts on Internet of Things and Big Data[J]. Communication World, 2017, (01): 1-2.[2017-08-31].

Design and Simulation of Indoor Tour Guide Robot Based on ROS

Yuheng Zhang [1,2], Yizhun Peng [1,2*], Lianchen zhao[1,2]

[1]*College of Electronic Information and Automation,*
Tianjin University of Science and Technology,China
[2]*Advanced Structural Integrity International Joint Research Centre,*
Tianjin University of Science and Technology, China
*E-mail: * pengyizhun@tust.edu.cn*
www.tust.edu.cn

Abstract

Aiming at the problems of the small number of lecturers in museums and other venues and the large demand for audience consultation, a ros-based indoor guide robot was designed. The robot consists of a mechanical system, a motion control system, and a sensor system. In order to improve the efficiency of development and debugging, and reduce the cost of experiments, it is necessary to test related algorithms in a virtual simulation environment before the robot enters the actual working state. Experiments have proved that the robot can autonomously guide guests to the destination and explain according to a preset path; the robot has multiple sensors to sense obstacles, and can autonomously avoid obstacles during the explanation and continue to move; the robot can accurately and efficiently recognize faces and provide accurate services Fast.

Keywords: Indoor tour guide robot; Autonomously guide; Voice interaction; Face recognition

1. Introduction

With the rapid development of intelligent technology, the quality of people's lives continues to improve. Since the existing "guide" intelligent robots can only realize the function of navigating the road when in use, the function is relatively single, and the existing "guide" intelligent robots cannot guide tourists and cannot achieve the effect of reminding.[1] Aiming at the problems of museums and other venues such as the lack of interpreters and the large demand for audience consultation, a ros-based indoor tour guide robot was designed and simulated in gazebo software. The robot mainly realizes the following 3 functions:

(i) Welcome explanation: The robot can guide the guests to the destination and give an explanation according to a preset path. After the explanation, the robot will automatically return to the original starting position;

(ii) Audio guide: The robot can interact with people by voice. During the robot's explanation process, the audience can interact with the robot at any time if they have questions or problems.

(iii) Face recognition: The robot takes pictures of the user who is guided for the first time to collect public information, continuously enriching the public's personal portraits, establishing more complete personal information materials, and providing accurate and fast services.

2. The Composition of the Robot

The robot consists of four layers. The first layer is the chassis movement layer, including two motors, two driving wheels, one universal wheel, and three infrared sensors; the second layer is a lidar layer with a lidar; the third layer It is the driving layer, including the lower computer, battery, three ultrasonic sensors, a smoke sensor, and a temperature and humidity sensor; the fourth layer is the decision-making computing layer, including the upper computer, monitor, camera, microphone and speaker. The schematic diagram of the robot is shown in Figure 1.The upper computer is an industrial computer with high antimagnetic, dust-proof and impact-proof capabilities.[2] There is a special power supply in the chassis, and the power supply has strong anti-interference ability. The processor is an inter i7 processor with four usb3.0 interfaces And four usb2.0 interfaces, two network ports. The lower computer controller mainly realizes the bottom-level motor drive, power management and acquisition of various sensors such as ultrasonic and infrared, and at the same time feeds back the collected data to the upper computer in time. The lower computer controller uses STMicroelectronics' STM32F103ZET6 chip as the main control, the chip flash has 512K, the SRAM is 64KB, and the pins are 144.[3] The chip runs the FreeRTOS system. The infrared sensor can realize the function of assisting obstacle avoidance, preventing the robot from falling when it reaches the road pit.[4] Lidar is used to realize the range finding scan and generate the plane point cloud map information of the space. These cloud map information can be used in practical applications such as map surveying, robot positioning and navigation, and object/environment modeling. The block diagram of the robot system is shown in Figure 2.

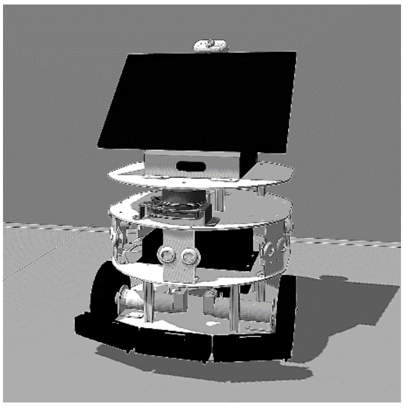


Fig.1. The schematic diagram of the robot

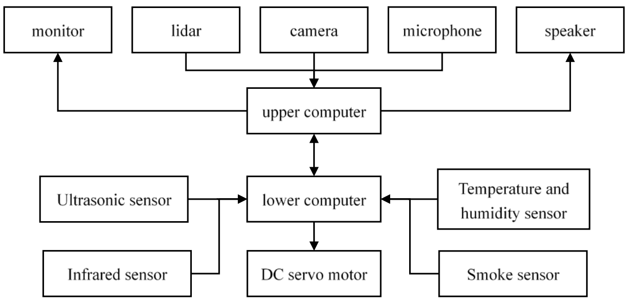


Fig.2. The block diagram of the robot system

3. Robot Simulation Test

3.1. Construction of simulation environment

The simulation environment of the robot is built on the gazebo simulation platform, and the walls are drawn by the building editor tool provided by gazebo. We built the indoor scene of a simulated exhibition hall, including venues in 16 Chinese cities such as Beijing and Tianjin. Each venue is 12 meters long and 4 meters wide. The robot model is built and exported by solidworks software. The initial position of the robot is in the center of the hall of the exhibition hall. The simulation environment of the robot is shown in Figure 3.

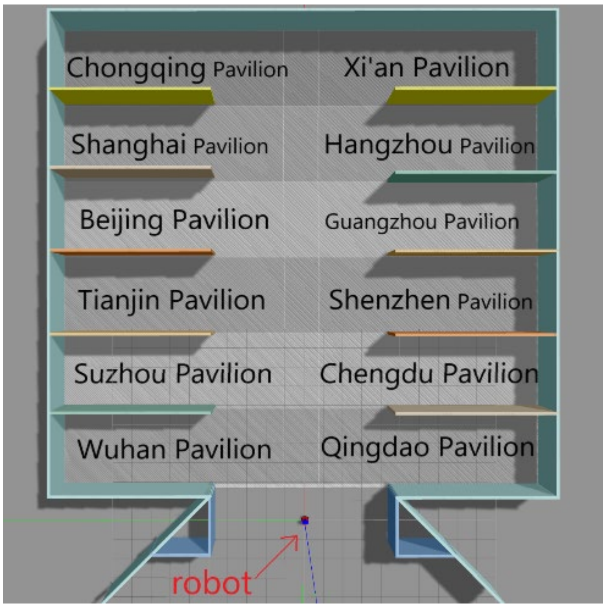


Fig.3. Robot simulation environment

3.2. Mapping and navigation

Only after the establishment of the map is completed and the map is obtained, the robot can be controlled to navigate. SLAM is simultaneous localization and mapping, which can be described as: the robot starts to move from an unknown position in an unknown environment.[5] During the movement, it locates itself according to the position estimation and the map, and at the same time builds an incremental map to realize the robot's autonomous positioning and navigation.[6] This design uses the most commonly used and mature gmapping algorithm in SLAM, gmapping integrates the Rao-Blackwellized particle filter algorithm, the following is the overall framework of gmapping and navigation.

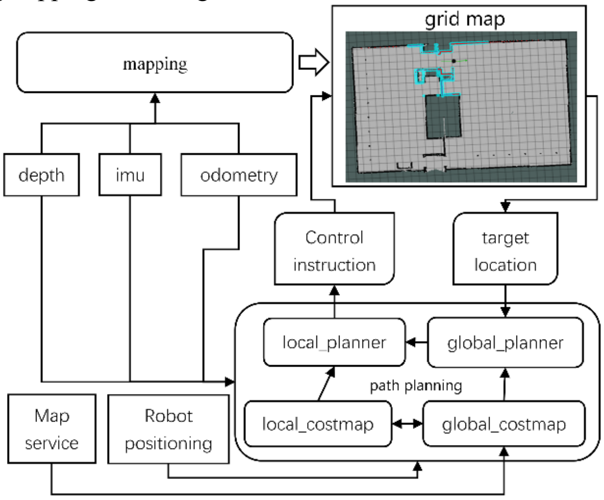


Fig.4. The overall framework of gmapping and navigation

The gmapping function package subscribes to the robot's depth information, IMU information and odometry information, and at the same time completes the configuration of some necessary parameters, you can create and output a probability-based two-dimensional grid map. The key to navigation is robot positioning and path planning. The navigation function package needs to collect the sensor information of the robot to achieve the effect of real-time obstacle avoidance. It also needs the odometry information of the robot and the corresponding TF transformation. Finally, the navigation function package outputs instructions that can control the movement of the robot.

Use the keyboard node to control the robot to move in the Gazebo environment. After the map is created, save it.This map will be called during the navigation process.

Use the move_base function package for global path planning and local real-time planning. According to the given target location and the constructed global map, the A* algorithm is used to plan the global path, and the optimal route from the robot to the target location is calculated as the global route. In practice, obstacles may appear at any time, so local real-time planning must be carried out to make it conform to the global optimal path as much as possible. Use the Dynamic Window Approaches algorithm to search for multiple paths to avoid and travel, and select the optimal path according to the evaluation index. Use the amcl function package to complete the autonomous positioning function, that is, the robot can calculate its position on the map in any state.[7] Perform navigation simulation in Rviz, as shown in the figure below, where the red arrow is the specified target position, and the green line between the red arrow and the robot is the global path planned by the navigation function package.

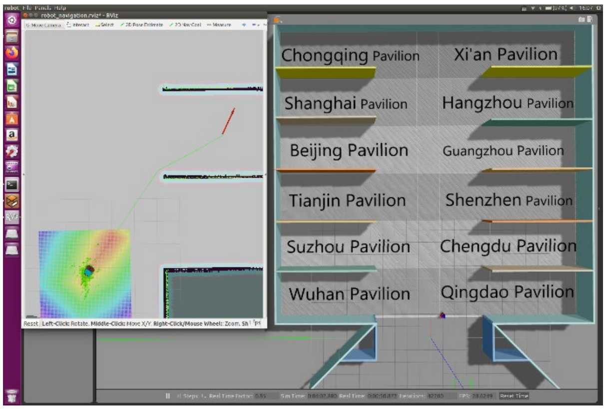


Fig.5. Robot navigation simulation

3.3. Welcome explanation

The robot can guide the guests to the destination and give an explanation according to a preset path. After the explanation, the robot will automatically return to the original starting position. When building a map, record the location of each venue as a navigation target point. Move_base implements path planning through action, and action uses a client/server architecture. Use the send_goal() interface to send to the server of move_base. After receiving the target task, the server will realize the path planning function according to the current pose of the robot, and output the speed command to control the movement of the robot. Go to the target location and return the result to the client. If the result is successful, it will send the introduction of the corresponding venue to the speech synthesis module, and let the robot speak the introduction. This function module is based on the tts_sample program in the iFLYTEK SDK, modify the main code file, add the ros interface, subscribe to the introduction topic, receive the input voice string, and use the SDK interface in the callback function to convert the string into Voice, after the voice broadcast is completed, the robot will navigate to the next target point and continue to broadcast until the last target point is finished.

3.4. Audio guide

This function is based on the speech recognition SDK provided by IFLYTEK and is constructed using the ROS system. The iat_record_sample program in IFLYTEK SDK is modified and the required ROS interface is added. This design adds a publisher, a subscriber and a move_base client. The Subscriber is used to receive the voice wake-up signal. After receiving the wake-up signal, the wakeupFlag variable is set, and then the voice dictation function of the SDK is called in the main loop to recognize Human voice information, set resultFlag variable after successful recognition. Then use dataString.find() to determine the meaning of Chinese voice input, and publish the corresponding target point pose through the send_goal() interface. After receiving the target task, the server of move_base will realize global and local Path planning function according to the current pose of the robot, output speed command to control the robot to move to the target position, and return the result to the client. If the result is successful, it will send the introduction of the corresponding venue to the speech synthesis module, and let the robot speak the introduction.

3.5. Target face recognition

The specific process of face detection is shown in Figure 6. First, load the face detection classifier that comes with opencv. In this article, the classifier based on haar feature values is used. In this article, we mainly use the contour of the face and human eye recognition to realize face detection. The next step is to preprocess the image to be recognized, to grayscale the image, obtain its histogram, and then perform equalization. This step is mainly to make the feature value of the face in the image better extracted come out. Calling the detectMultiscal function to recognize the picture is mainly because the input picture is of different sizes, so it needs to be detected in multiple dimensions.[8] By calling the rectangle function, draw the rectangular frame of the face profile recognized by the program. Finally, the drawn rectangular frame of the face is cut out in the original image.

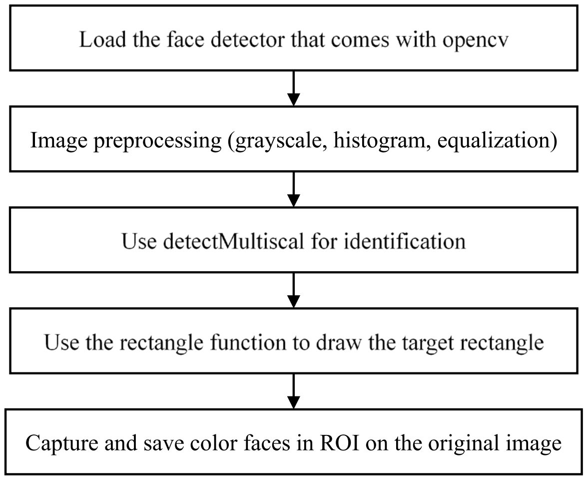


Fig.6. Face detection flowchart

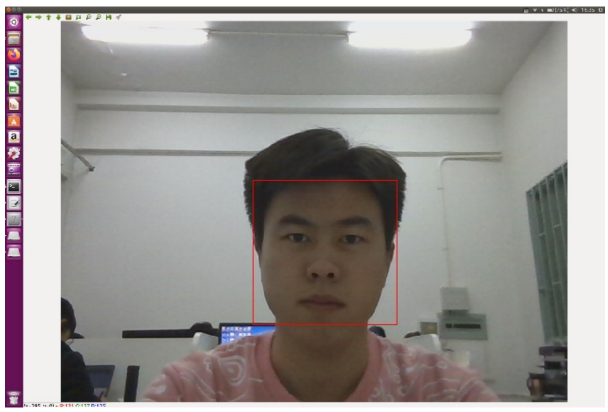


Fig.7. Face detection

To let the program know who to recognize, it is necessary to train the face model. Firstly, the target image is preprocessed, and the sample is normalized, mainly for the same size of the image, and then grayscale. Next, label the sample, this is to tell the training model which number belongs to whose face. Then start training, mainly using three methods, namely EigenFaceRecognizer, FisherFaceRecognizer and LBPHFaceRecognizer. The EigenFaceRecognizer method is also face recognition based on PCA transformation. The principle of PCA transformation is a way to reduce the dimensionality of an image, because in face recognition, the image is generally processed as a vector, but at the same time, the dimensionality of the vector is too large, and the huge dimensionality is quite difficult for subsequent image calculations. Therefore, it is necessary to reduce the image dimension without losing important information as much as possible. The FisherFaceRecognizer method is a face recognition based on Fisher transform.[9] LBPHFaceRecognizer is a face recognition method based on local binary patterns. After training, save the training model locally, and then you can start face recognition.

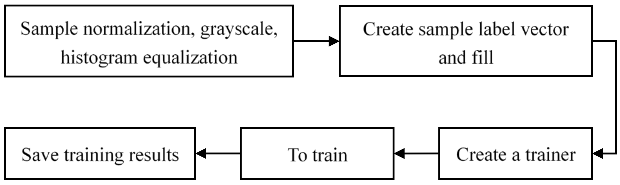


Fig.8. Face training flowchart

After having the above foundation, the final face recognition can be performed. First, image preprocessing is performed, and then the trained classifier is loaded to obtain the label value on the input image, and then the character name is obtained through the label value.

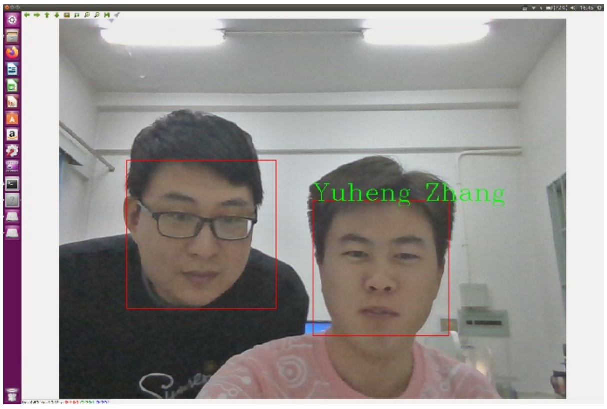


Fig.9. Target face recognition

4. Conclusion

In view of the small number of lecturers in museums and other venues and the large demand for audience consultation, the overall robot design was completed, and experiments were carried out on the gazebo simulation platform. The test proved that the robot can complete the functions of welcome explanation, voice guide, and face recognition. By interacting with robots, visitors can understand the situation of each exhibition area. The robot's way-guide function can also provide more convenience for visitors. This solution can improve the construction effectiveness and quality of museums and other venues, as well as a smart and personalized service system.

References

1. Jianfa Dai, Jingrong Ying, Hongtao Liu. A "Guide" Intelligent Robot Design, *Science Technology and Innovation*, 2020(11): 142-143.
2. Wenbin Ni, Guanghua Lu, Yi Zhang. A guide service robot design , *Machinery Manufacturing and Automation*, 2016, 45(06): 176-179.
3. Jing He. Design and production of campus tour guide robot, *Electronic Testing*, 2018(11): 76-77.
4. Chenxi Guan. Design and research of indoor inspection robot system based on ROS, *Electromechanical Information*, 2020(30): 128-129.
5. Cadena C.,Carlone L.,Carrillo H.,et al.Past,Present,and Future of Simultaneous Localization and Mapping:Toward the Robust-Perception Age, *IEEE Transactions on Robotics*,2016,32(6):1309-1332.
6. Rowekamper J.,Sprunk C.,Tipaldi G D.,et al.On the position accuracy of mobile robot localization based on particle filters combined with scan matching, *Intelligent robots and systems*,2012:3158-3164.
7. Yanbin G.,Shifei L.,Mohamed A.,et al.INS/GPS/LiDAR Integrated Navigation System for Urban and Indoor Environments Using Hybrid Scan Matching Algorithm, *Sensors*,2015,15(9):23286-23302.
8. Honglong Chen, Dongyong Liu, Zhichen Ni,et al.The face recognition experiment platform of quadrotor drone based on YOLO, *Experimental Technology and Management*, 2020,37(10):107-111.
9. REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once:Unified, real-time object detection, in *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, Las Vegas, Nevada, USA, 2016:779-788.

Research on Path Planning of Manipulator

Nana Wang [1], Yizhun Peng [1,2*], Zhou Yang [1], Yuheng Zhang [1]

[1]*College of Electronic Information and Automation,*
Tianjin University of Science and Technology, China;
[2]*Advanced Structural Integrity International Joint Research Centre,*
Tianjin University of Science and Technology, China;
E-mail: [*] *pengyizhun@tust.edu.cn*
www.tust.edu.cn

Abstract

In order to further realize the flexibility, intelligence and human-machine friendliness, the robot must have the ability of flexible motion planning. This paper mainly introduces the general steps of manipulator path planning and common path planning algorithms. First, the path planning of the manipulator is reviewed from three aspects: environment modeling, path search, and path smoothing. In addition, compared various manipulator path planning algorithms, such as RRT algorithm, ant colony algorithm, genetic algorithm, etc.Finally, the development trend of path planning technology is summarized and analyzed.

Keywords: Manipulator, Path planning, RRT algorithm, Ant Colony Algorithm, Genetic Algorithms

1. Introduction

With the progress of the times, the application of robotic arm technology is becoming more and more popular. For the manipulator, reasonably planning its own movement path is the most basic function to complete the task in the real world, which is a key technology in robot research - path planning. Path planning can be divided into two types: global path planning and local path planning. Global path planning refers to path planning based on all environmental information through certain algorithms under the condition that all environmental information is known. Local path planning refers to the situation where part or all of the environmental information is known.Real-time environmental information is collected by the sensor and the collected environmental information is analyzed for real-time path planning[1]. With the rapid development of the manipulator, the requirement for the path planning of the manipulator is getting higher and higher. Therefore, it is necessary to constantly optimize the path planning algorithm to save the movement time of the manipulator and reduce the joint loss of the manipulator. This paper mainly introduces the path planning of 6-DOF manipulator.

2. Path planning of 6-DOF manipulator

Path planning should first establish an environment model that is convenient for computer to carry out path planning. Then, a walking path that enables the predicted performance function to obtain the optimal value is searched through a certain algorithm. Finally, the path searched through the corresponding algorithm is smoothen to make it a practical and feasible path[2].

2.1. Modeling Kinematics of manipulator

2.1.1. Modeling

Aiming at the problem of how to describe the geometric parameters and coordinate system direction between the adjacent links of the mechanical arm, Denavit and Hartenberg came up with the DH method[3], in this method, a coordinate system is fixed on each manipulator link, and the rotation and translation vectors are described by means of homogeneous transformation. The transformation relation from connecting rod i-1 to i is shown in Equation (1). DH method is universally applicable. The joints between the link of the robot can be either rotating joints or prismatic joints, and there are no limits on the length and plane of the joints.

$$
\begin{aligned}
{}^{i-1}_{i}T = Rot(x_{i-1},\alpha_{i-1})Trans(x_{i-1},\alpha_{i-1}) \\
Rot(z,\theta_i)Trans(z,d_i)
\end{aligned} \quad . (1)
$$

2.1.2. Positive solution to the kinematic equation of the manipulator

Forward kinematics is to find out the spatial position and attitude of the end-effector of the mechanical arm relative to the base coordinate system according to the Angle value of each joint of the manipulator and the parameter information of each lever[4]. The 6-DOF manipulator is composed of link connected with six joints, and the position and attitude of the end-effector relative to the base are formula (2)[5].

$$
{}^{0}_{6}T = {}^{0}_{1}T\,{}^{1}_{2}T\,{}^{2}_{3}T\,{}^{3}_{4}T\,{}^{4}_{5}T\,{}^{5}_{6}T \quad .(2)
$$

2.1.3. Inverse solution to the kinematic equation of the manipulator

Inverse kinematics refers to solving the Angle value of each joint of the manipulator based on the spatial position and attitude of the end-effector relative to the base coordinate system. Inverse transformation is often used to solve inverse kinematics, According to the forward kinematics equation of the manipulator, the solution process is formula (3). the Angle of joint 1 is calculated according to the corresponding matrix operation. Move the next joint Angle to the left of the equation and repeat until you have all the solutions.

$$
({}^{0}_{1}T)^{-1}\,{}^{0}_{6}T = {}^{1}_{2}T\,{}^{2}_{3}T\,{}^{3}_{4}T\,{}^{4}_{5}T\,{}^{5}_{6}T \quad .(3)
$$

The inverse kinematics solution of the 6-DOF manipulator with the last three joints perpendicular to each other theoretically has a maximum of 8 sets, and then needs to be selected according to the actual joint Angle range[6].

2.2. Collision detection

Collision detection technology is an important part of the path planning of the manipulator. It is mainly used to judge whether there is a collision between two objects. Accurate and effective collision detection method is the premise of effective obstacle avoidance path planning for the manipulator. Generally speaking, space obstacles are irregular geometric shapes. In order to judge the collision problem of manipulator conveniently, it is necessary to regularize the obstacles. Some scholars have proposed using the regular body envelope of obstacles to approximate modeling. Although this approximation expands the obstacle domain, it greatly simplifies the description of the obstacle domain and effectively improves the efficiency of planning[7]. Common obstacle envelopment models include: Sphere, AABB, OBB, K-DOPS, etc.

Sphere is the smallest Sphere that contains objects[8]. This method is not only simple in structure, but also does not need to be updated in real time as the object rotates and moves. However, its tightness is poor and large redundancy space will be generated.So it's less used.

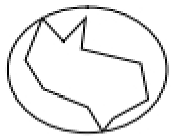


Fig. 1. This is a two-dimensional projection of sphere

AABB[9-10] is the smallest hexahedron that contains an object and its sides are parallel to the axis. The AABB box model was the first to be used.This method is not only simple in construction but also fast in updating and relatively high in efficiency, which can be applied to the collision detection between deformed bodies.However, for some objects with irregular shape, its envelopment is poor, resulting in large redundancy space.

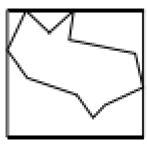


Fig. 2. This is a two-dimensional projection of AABB

OBB[11-13] is the smallest cuboid that contains an object and is arbitrary relative to the direction of the axes. OBB is a more commonly used type of bounding box.The biggest advantage of this method is its arbitrary orientation, which makes it possible to surround the object as closely as possible according to the shape characteristics of the surrounded object, but its construction is very difficult.

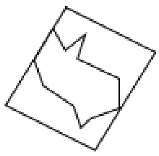


Fig. 3. This is a two-dimensional projection of OBB

K-dop[14-15] is a convex hull containing an object whose normal vector of all faces is taken from a fixed direction. It inherits the good compactness of convex hull and has the characteristics of simplicity[16].

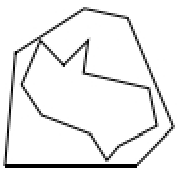


Fig. 4. This is a two-dimensional projection of K-DOP

2.3. Environmental modeling

The actual working environment of mobile robot is a real physical space, while the space handled by path planning algorithm is the abstract space of the environment.Environment modeling is a mapping from physical space to abstract space.This efficient description of the robot's active space is called environment modeling[17]. Environment model is the basis of path planning algorithm, whether it can fully reflect the environment information and establish environment model effectively is related to the performance of the whole path planning algorithm.When the environment information is fully known, the global environment model can be established.In practical applications, the working environment of mobile robots is often uncertain, which makes it impossible for robots to build global environment model in advance, but to build local environment model in real time according to sensor information[18].

Environment modeling is a very important part of robot path planning.The reasonable environment model is beneficial to the reduction of search volume and the saving of space-time cost in the path planning process. Common environment representation methods include raster method, cell tree method and polygon representation method.

2.4. Path search

The path searching stage is to apply the corresponding algorithm to find a walking path based on the environment model so that the predetermined performance function can get the optimal value. Common path search algorithms include RRT algorithm, ant colony algorithm, genetic algorithm, etc.

2.4.1. RRT algorithm

RRT algorithm is a global path planning algorithm based on random sampling proposed by Steven M. aValle. This method does not need to be learned, and is suitable for path planning in dynamic environment. RRT algorithm can be divided into two categories, one of which is Single directional rapidly-exploring random tree (Single-RRT), and the other is Bi-directional rapidly-exploring random tree (BI-RRT). The single-RRT method randomly selects a point in the moving space of the manipulator as a node in the tree, and continuously searches for nodes to expand the tree until a path is found from the starting point to the target point. The idea of BI-RRT algorithm is to generate two fast search random trees from the starting point and the target point until the two trees are connected. The advantage of this algorithm is that the two trees can be expanded in parallel. The complexity of this algorithm lies in how to make the two trees cross connected. The following mainly introduces the single branch search rapid expansion of random number method, the main process is as follows[19].

(1) The starting point Xstart and target point Xgoal of the manipulator are given in the working environment space, and the starting point is taken as the target point of the growth tree.

(2) The search step of the given algorithm is L0, and a random tree Trand in the manipulator workspace with N nodes is generated from the starting point Xstart.

(3) Select any random point Xrand in the manipulator workspace.

(4) Traverse the random tree Trand, calculate the distance from each node of Trand to the random point Xrand, and screen out the point Xnear closest to the random point Xrand among the N nodes.

(5) Take the line of Xnear and random point Xrand as the growth direction, take Xnear as the starting point, and generate a new node Xnew along the growth direction with the step length of the algorithm as the length.

(6) Judge whether Xnew has collision with the obstacle. If there is no collision, add it to the random tree Trand; if there is collision, discard the node Xnew.
(7) Judge whether the distance between Xnew and Xgoal is less than the setting neighborhood (DISfinish) of the end point.If less than, the path search ends, if not less than repeat steps 3-6 until less than.

The advantage of RRT algorithm is that it has powerful ability to search unknown space, and it is often used for path navigation in unknown environment, and it is fast.However, the planned path is not smooth enough. When the sampling domain of RTT algorithm and the original problem are not suitable, the convergence rate of standard RRT algorithm will be slow.

2.4.2. Ant Colony Algorithm

ACA[20] is a heuristic global path planning algorithm proposed by Italian scholar Dorigo Mden.The main idea of this algorithm is that each ant leaves a certain concentration of pheromone on the path it travels while foraging, and the concentration of pheromone is inversely proportional to the length of the path. By adopting the positive feedback mechanism, the shortest path with a high concentration of pheromone will be found soon. The basic steps of the ant colony algorithm are as follows(Take environment modeling as a raster method as an example)[21]:
(1)Set the initial grid of path planning as Sstart and the target grid as Sgoal.The initialization cycle number Nstart is 0, the maximum cycle number is Nmax, the information heuristic factor is α, the expected heuristic factor is β, the pheromone volatility coefficient ρ, constant Q, and the initial pheromone on the path between grids is constant $\tau_{ij}(0)$=const.M ants were placed at the starting point of the path planning, Sstart, and Sstart was added to the tabu.
(2)If the current grid is not the target grid, select the next grid according to the probability calculated by the formula below, and add the selected grid to your own tabu list.The probability that ant K at node I chooses node J as the transfer direction is formula (4).

$$P_{ij}^{k}(t)=\frac{\tau_{ij}^{\alpha}(t)\eta_{ij}^{\beta}(t)}{\sum_{s\in allowed_k}\tau_{ij}^{\alpha}(t)\eta_{is}^{\beta}(t)}, j=allowed_k$$

$$P_{ij}^{k}(t)=0, j=otherwise \quad .(4)$$

(3)Repeat execution 2 until all ants have completed the subsequent grid selection.
(4)Repeat 2,3.Until all ants have moved to the target grid.
(5)By counting the total length of the path taken by each ant, the shortest path and the length of the shortest path in this cycle are obtained.
(6)Update the pheromone concentration on each path according to Formula (5).

$$\tau_{ij}(t+n)=\rho*\tau_{ij}(t)+\Delta\tau_{ij}, \rho\in(0,1)$$

$$\Delta\tau_{ij}=\sum_{k=1}^{m}\Delta\tau_{ij}^{k} \quad .(5)$$

(7)Tabu table cleared, cycle times increased by 1.If the number of cycles is less than the maximum number of cycles, go to Step 2; otherwise, skip the entire cycle and output the optimal path and the optimal path length.

This algorithm is a distributed system, which has strong adaptive ability and good global optimization ability, but it has a large amount of computation, is easy to fall into the local optimal solution, and does not have real-time ability.

2.4.3. Genetic Algorithms

GA[22] is an iterative search method based on the principle of genetics proposed by John Holland in 1962.The algorithm is mainly composed of eight steps, respectively set up maps, initialization of population, calculating the value of individual fitness, choosing the appropriate individuals into the next generation, crossover, mutation, update the population, if appear the optimal path or the number of iterations is output optimal solution, otherwise, continue to iteration, the optimal output of the individual as the optimal solution.

The method requires fewer objective functions and constraints, and the optimization scope is extended to the whole independent variable space. After iteration, the method can approach the global optimal solution with greater probability.At the same time, genetic algorithm can solve nonlinear and complex optimal solutions effectively by using random transformation rules instead of deterministic rules. Its greatest advantage is that it is easy to combine with other algorithms, and it gives full play to its iterative advantages.However, due to the large storage space and long operation time of this algorithm, the operation efficiency is not high and it is easy to converge too early.

2.5. Path smooth

In general, the path planned by path search algorithm is a continuous line segment composed of some discrete points, and the shape of the path is not smooth.This not only reduces the motion stability of the mechanical arm, but also causes the impact and vibration of the motor and accelerates the wear of the mechanical parts.Therefore, path smoothing is very important to ensure the smooth movement of the manipulator.Quadric B-spline curves are often used to smooth obstacle avoidance paths so that the manipulator can work continuously and stably.

3. The conclusion

As the robotic arm is more and more widely used in various fields, the path planning environment of the robotic arm is more complex and changeable.This requires that the path planning algorithm should have the ability to respond quickly to complex environment changes.This paper summarizes the basic methods of path planning.In the future, path planning technology will be applied in more and more fields, and more efficient, flexible and accurate path planning algorithms will be born.

References

1. Chen Gu,Study on modular manipulator design and its path planning method. *jiangsu,Jiangsu University of Science and Technology,2014.*
2. Guanglin Zhang,Xiaomei Hu,Summary of path planning algorithm and its application.*The modern machinery,2011.*
3. Zhaoguang Wang,Rrsearch on obstacle avoidance path planning for 6-DOF manipultor. *Southwest Jiaotong University,2018.*
4. Chengxing Zong,Research on obstacle avoidance motion planning method for spatial multi DOF serial manipulator.*Hefei ,Hefei university of Technology,2017.*
5. Guoxiu Cheng,Research of Obstacle avoidance path planning based on manipulator with six degrees of freedom.*Shenyang, Northeastern University,2012.*
6. WeiSheng,The research of planning technology for obstacle avoidance path based on multi-degree of freedom industrial manipulator.*Jiangsu, Jiangnan university,2012.*
7. Qingxuan Jia,Gang Chen,Path planning for space manipulator to avoid obstacle based on A*algorithun. *journal of mechanical end ineering,2010,46(13).*
8. Guangmiao Jiang,Research and application of collision detection algorithm based on bounding box.*University of Electronic Science and Technology,2012.*
9. Xuewen Chen,Wusheng Chou,Research on collision detection algorithm based on bounding box.*Computer Engineering and applications,2005,41(5):46-50.*
10. Shaowei Fan,Research on real-time collision detection technology. *Zhejiang university,2003.*
11. Zhilei Ding, Fast collision detection based on OBB bounding box. *Lanzhou University of Technology,2006.*
12. Gottschalk S.Collision queries using oriented bounding boxes*:The University of North Carolina at Chapel Hill,2000 .*
13. D.Baraff. Interactive simulation of solid rigid bodies. *IEEE Computer Graphics and Applications, 1995,15(3):63-75*
14. Kim Y J, Lin M, Manocha D. Fast penetration depth estimation using rasterization hardware and hierarchical refinement.*Proc of Symposium on Computational Geometry.2003:386-387*
15. Klosowski J, Held M, Mitchell J S B,etal. Efficient collision detection using bounding volume hierarchies of k2

DOPs. *IEEE Transaction on Visualization and Computer Graphics, 1998,4(1):21-37*
16. Tao Wang,The Study of robotic path planning and related simulation system based on ant algorithm.*Xian,Xi an University of Science and Technology,2009.*
17. Lihui Peng,Research on the kinematics and path planning of manipulator.*Hunan,Hunan University of Technology,2012.*
18. Chao Ma,Research on path Planning of 6R tandem manipulator in complex space environment.*Jinan University,2017.*
19. Jinming Dong,Research on path Planning based on Ant colony algorithm.*Xian,Shaanxi Normal University,2009.*
20. Jincai Ma,Research on path planning based on six-DOF manipulator.*South China University of Technology,Guangzhou.2013*
21. J Tu，S Yang． Genetic Algorithm Based Path Planning for a Mobile Robot.Taiwan: *Proceedings of IEEE Intelligent Conference on Robotics and Automation,3003:P1221-P1226.*
22. Tao Han,Huaiyu Wu,Research on path planning for manipulator to avoid real-time obstacle based on genetic algorithms.*Applicatian Research f Camputers,30(5),2013.*

Research on White-Line-tracking Walking Technique of NAO Robot

Wanlong Peng[1,2], Yizhun Peng[1,2*], Yuheng Zhang[1,2], Tianye Jian[1,2]
[1]College of Electronic Information and Automation,
Tianjin University of Science and Technology, China
[2]Advanced Structural Integrity International Joint Research Centre,
Tianjin University of Science and Technology, China
*E-mail: * pengyizhun@tust.edu.cn*
www.tust.edu.cn

Abstract

Liner-tracking walking technique is an essential part of intelligent robot technology. This paper selected NAO robot as the research platform to study the line-tracking walking[1] problem. Firstly, NAO white shell increased the difficulty of white line recognition, this is different from other robot. A method based on image preconditioning[2] is presented, sorting white edge with cut-point which is non-differential or derivative is zero, then designed the controller based on the improved digital incremental PID(proportion integration differentiation)algorithm and determined the controller parameters through the concise test method on Simulation. Finally, it was realized by programming in Python language.

Keywords: Humanoid robot, Edge classification, Image preconditioning, Simulation

1. Introduction

In recent years, the research and application of intelligent humanoid robots have gradually emerged. Line-tracking robot with certain autonomous decision-making and perception capabilities have quietly become an industry trend. Intelligent robot with line-following walking function can replace or help humans to complete some high-complexity, high-risk and high-workload inspections[3] under complex conditions in some harsh environments, to find and eliminate equipment abnormalities in time. Minimize the loss of personnel and property. Therefore, the research on intelligent robots with line-following walking function has great significance in the industrial field. The NAO robot represent by the humanoid robot is a biped humanoid robot used for scientific research、competition and entertainment, with 25 degrees of freedom and 36 sensors. Possess certain environmental perception and the ability to make corresponding decisions. This paper uses the NAO robot as a platform to study the image perception processing of robot. The main work content of this research includes: Propose a color-segmentation method to increase the speed and reliability of the line recognition algorithm[4];The robot`s built-in camera collects images every 30s,and converts the image information into reference input signals through technical means; Find the difference between the potential and the reference signal as the controller input; Write the program in python and carry out the line-tracking simulation experiment in the software.

2.White line processing

We takes the following steps to processing the white line:

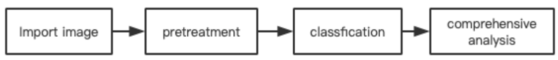


Fig1.white line identification flowchart

2.1 *Image preprocessing*

Through the color segmentation of the image the points outside the white line will generally become the fulcrum,so that the image boundary is easy to find.The original image format provided by the NAO robot is YUV422,but the original image is generally used directly to avoid the overhead of format conversion.

The color table is a three-dimensional array, each dimension represents Y, U, and V respectively. The array value stores the color type, such as 0 for unknown color, 1 for red, etc.Set the dimension to 64,64 and 64 respectively enough to meet the requirements. The color table needs to be established by manually collecting pixels. This is established offline during robot debugging. The established color table is used in the competition.

The image segmented5 by the color table generally has some color points that are different from the actual ones. For example, there are green points on the robot, and sometimes white or unknown colors may appear on the ground that should be green. This phenomenon is difficult to remove, but it can use some noise processing methods to filter out some salt and pepper noise.

3 White edge recognition

The traversal of white points is an edge growth process , and the recognition process6 is as follows: First,scan the image in order, find the first white edge point, and record all adjacent points of this point. If the image is scanned, turn to enter; Second, take one of the neighboring points as the starting point, start from this point, traverse along the edge, and record these edge points,turn to third, turn to traverse, if the neighboring points are traversed, return first to perform the next scan;Third,calculate the relative slope information of the current point and the previous recorded point every few points to determine whether it is a split point, if it is not a split point, continue to traverse, otherwise go to last;Lase, when the split point appears, calculate the slope information of the point recorded in second, determine whether it is a long straight line, a short straight line or an arc, and record the result, and return to second; Next,The result recorded in the comprehensive analysis last Information, output results.

3.1 *Scan and traverse edge points*

Scan the image first to find the white edge points, and click the grid scan here. So as to ensure the effectiveness of scanning and speed up the algorithm at the same time. Scan the image from top to bottom in a column. If white dots and non-white dots appear alternately in the same row or in the same column of the 4 dots during scanning, edge dots appear, and then scan row by row or column by row to locate the edge point.

There is the problem of whether the path is unique during traversal. In addition to the multiple paths that may appear in the outer scan for the first time, there are also two cases where there are multiple path options, as shown in Figure 5 and Figure 6, other points are normal, There is only one way to traverse。

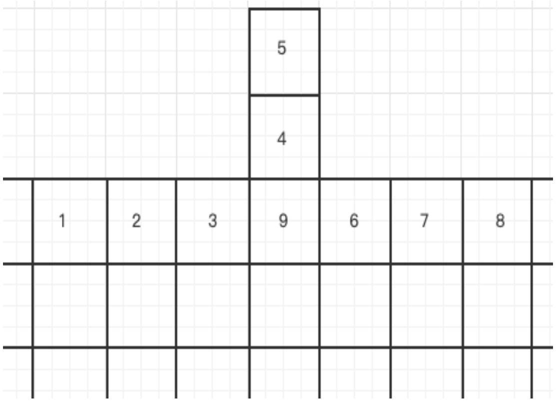


Fig.2 Convex edge

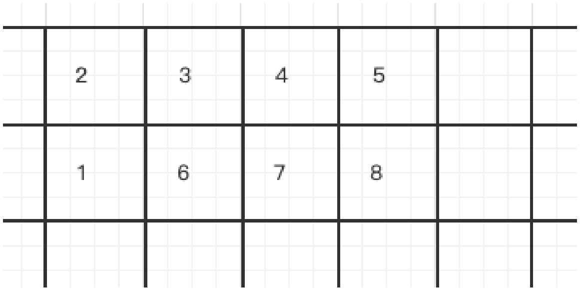


Fig.3 Edge with width 2

The numbers 1 to 8 in Figure 2 are all white edges. When traversing to 4, there are two paths of 5 and 6 in different directions. There are two ways to solve this convex situation: (1) Use the stack data structure to save Such different paths are then traversed by depth-first traversal ; (2) Consider this situation during image preprocessing, and directly filter out the points 4 and 5 that cause this convex edge. Another situation where different paths appear is shown in Figure 3. After research, it is found that when the traversal priority is the same, the traversed paths will not intersect between the edges. This solves the ambiguity path problem that appears when traversing along the white edge.

3.2 *Calculate the split point*

As shown in Figure 4, there are several straight lines and curves on the plane, so how to classify these lines based on the situation of the NAO stadium? This article uses dividing points to divide them. Intuitively, the intersection of the straight lines should be divided Based on this idea, the dividing point in this paper is the non-derivable point and the point where the derivative is 0.

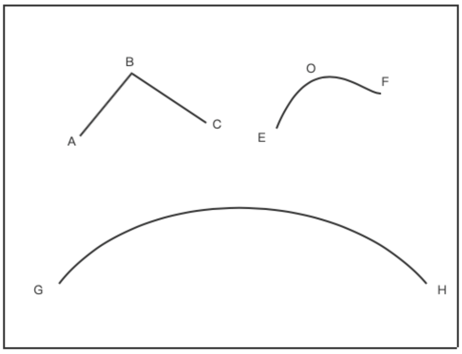


Fig.4 Line classification

There is a polygonal line ABC in 4, and its function is f(x). When accessing from point A to point C along the polygonal line ABC, calculate the left and right limits of each point P on the path. If the left and right limits of P exist and are equal, the formula (1) Is established, then continue to the next point. When visiting point B, it is found that the left limit of point B is equal to the slope of the straight line AB, and its right limit is equal to the slope of the straight line BC, that is, formula (1) is not true, then ABC must not be on a straight line, and point B is the dividing point. Record the path from A to B, then start from point B and continue to visit point C in the same way, thus dividing the polyline ABC into two segments AB and BC. When the straight line and the abscissa are perpendicular, the points on the straight line can be guided but the slope does not exist, and the included angle is 90 degrees. It is worth mentioning that in order to reduce the impact of large changes in the slope of adjacent edge points caused by unsmooth edges[7], you can calculate the angle of the vector every N points along the edge, which may bring about the phenomenon of "turning back" when traversing the edge with a small width. In this way, pay attention to the direction change of the traverse point when using the angle instead of the slope.

$$\lim_{x\to p+} f(x) = \lim_{x\to p-} f(x) \quad (1)$$

The curve EOF is exactly the deformation of ABC, and these can be treated in the same way. The curve GH in Figure 4 is a circular arc with a large curvature. When traversing GH, there is a point on which the derivative is zero, but when calculating the slope, it is found that the slope changes very little and must be within the threshold range. There is no split point.

3.3 *Result classification*

When a segmentation point[8] appears or the image boundary is reached during traversal, a series of points just traversed need to be classified. These points can be classified into long straight lines, short straight lines and arcs. First of all, the number of points can be used to judge whether the current points are short straight lines. The others are long straight lines or circular arcs. Some simple calculations can be used to judge whether they are long straight lines or circular arcs. Obviously, the slope of the start point and the end point of the long straight line should not change much, while the arc should change more. The method used in this experiment is to compare the slopes of the front N points and the back N points of a series of points to be classified to judge, that is, use the difference between slope1 of formula (2) and slope2 of formula (3) to measure , And then judge whether formula (4) is true or not, this method is simple to calculate without losing validity.

$$\mathrm{slope1} = (\textstyle\sum_{i=0}^{N}(\mathrm{PiPi}+1)/N \quad (2)$$

$$\mathrm{Slope2} = (\textstyle\sum_{i=0}^{N}(\mathrm{Pn}-\mathrm{i}-1\mathrm{Pn}-\mathrm{i})/N \quad (3)$$

$$|\mathrm{slope1}\text{-}\mathrm{slope2}|=\mathrm{threshold} \quad (4)$$

Fig.5 import result

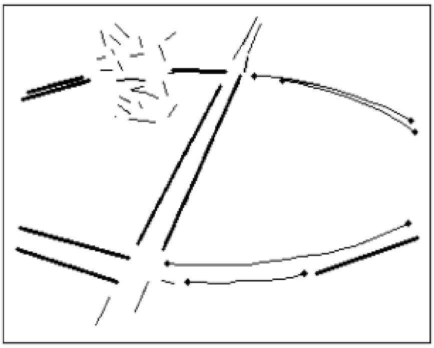


Fig. 6 Classification result

It can be seen from the classification results that the arcs on the court were originally classified as straight lines. It is necessary to consider the arcs and straight lines in combination with the actual situation of the court. In fact, many recognition algorithms directly transform the arc into a short straight line, and then connect, after comprehensive consideration. This can be achieved by the principle of Hough line transformation, but the element considered in Hough line transformation is a point.And here is a straight line,Also pay attention to the parallel features of the lines, and join together short straight lines, long straight lines, and arcs.

4. Experimental results and analysis

The experiment used is 3. The 3 version of the NAO robot, the main hardware parameters9: CPU:AMD GEODE x86 500MHz; Synchronous Dynamic Random Access Memory (SDRAM): 256MB, Flash: 2G; Camera: 2 COMS cameras, providing up to 30 frames/s, resolution of 640 × 480 images; Open CV library comes with the operating system. The criterion of experimental judgment is to identify the main straight lines and arcs. The results are shown in Table 1.

Table 1. Experimental data

distance / light	Light 1	Light 2
Far away (more than 4m)	63%	20%
Medium (0.5m to 3m)	96%	71%
Near (less than 0.3m)	50%	50%

The color table of the experiment in Table 1 is configured under light 1, and the white line that is too far away cannot be recognized basically (when the distance is very far, the recognized court boundary can be used to give an estimated value), and it can be recognized when the distance is gradually reduced Beyond the boundary, there is a good recognition rate around 2m.When the distance is too close, although the image is very clear, the robot's field of view is very small, and it can be easily recognized when there is no interference. When NAO is used to eliminate interference, the robot basically blocks the white line, so the formal recognition rate is 50%. Using the color table configured under light 1 to split the color of the court under light 2, the experimental results are much worse than that under light 1, but the close distance is basically not affected. In fact, most of the recognition results are used by the upper layer in the robot positioning algorithm. Most of the positioning methods used are Monte Carlo filtering or generalized Kalman filtering, so that the program only needs to identify the image The main lines and arcs.his method is used for the calculation of the two-dimensional array of the same price of the three-dimensional image after color segmentation. It also uses grid scanning. The main reason is that the recognition method is simple to calculate. The white line recognition program runs generally in 9ms to 12ms, which satisfies the robot competition. Requirements for straight line recognition. Compared with simple line recognition, the method in this paper pays more attention to the overall efficiency of robot image processing.The image preprocessing in it makes subsequent recognition algorithms more effective.

References

1. Thomas Rofer,Tim laue,.Particle-filter-based Self-localization using Landmaks and Directed Lines Robot Soccer World Cup IX.2005,4020,pp.608-615.
2. Xiangkan,Sun,.Detection and tracking of NAO based on machine learning[J].Journal of Changchun university of Science and Technology(Natural Science Edition),2016,39(2),pp.116-119.
3. Rong-Jong Wai.   Tracking control based on neural network strategy for robot manipulator[J]. Neurocomputing,2002.
4. D.M. Dawson,Z. Qu,F.J. F. Dorsey.Robust control for the tracking of robot motion[J]. International Journal of Control,1990.
5. Shen Zhang,Research on target recognition and localization of NAO robot [D].Anhui University, 2013.pp.20-23.
6. Kun Zhao,Research on the visual navigation method of intelligent inspection robot for substation [D].North China Electric Power University, 2014.pp.43-49.
7. Xuefeng Bai,Bin Yang,Based on NAO robot target recognition and localization algorithm [J]. Journal of Chengdu University of Information Technology, 2014, 29(6),pp.625-629.
8. Yanhong Wang，Research on moving target detection and tracking algorithm based on OpenCV [D].   Hangzhou: Hangzhou Dianzi University, 2013。
9. Xin Wang, Hongliang Yu,Research on diesel engine fault diagnosis based on the fusion of Bayesian and genetic neural network [J]. Ship engineering. 2012,34(1),pp.32-35.

A Design and Implement of Portable Epidemic Detection Device based on STM32

Yuqi Zhao, Yizhun Peng*, Xiaowei Wu, Yusong Zhou

College of Electronic Information and Automation,
Tianjin University of Science and Technology, Tianjin, 300222, China;
*E-mail: * pengyizhun@tust.edu.cn*
www.tust.edu.cn

Abstract

The equipment is mainly designed for epidemic detection. It is an intelligent detection box device. It collects and analyzes data through cameras and sensors. It uses LBP algorithm and LAB color model to enable the detection box to identify the temperature and mask of the entered personnel. Wearing conditions, preliminary comparison of epidemic prevention standards for passers-by, whether to wear masks, whether the body temperature exceeds the standard, and at the same time enter the identity of new personnel and store them in the database, and detect the incoming personnel through face recognition to find out in time Those who are not entered shall take corresponding measures in time to improve the efficiency of epidemic prevention and control. The device does not need to be held in hand, can be fixed anywhere, can effectively avoid cross-infection, and supports self-set temperature alarm thresholds, support voice broadcast reminders, fast and accurate non-contact temperature measurement.

Keywords: Non-contact temperature measurement, face recognition, mask recognition, identity collection

1. Introduction

Since the first case of unidentified pneumonia appeared in Wuhan on December 8, 2019, to December 1, 2020, a total of 93,889 new coronary pneumonia patients have been diagnosed nationwide, of which 87,707 cases have been effectively cured, accounting for approximately the total number of confirmed cases Of 93.41%.

Since the new crown epidemic, domestic measures have been effectively implemented[1]. A large number of detection equipment has been used in epidemic prevention and control, and there has been a relatively large market in China and the world. At the same time, the functional requirements of detection equipment have become higher and higher. Non-contact body temperature detection, mask wearing detection, identity entry, etc. can achieve precise identification and detection. The accuracy of the infrared thermometer is related to four key factors: sensor, ADC accuracy, algorithm, and the structure and production process of the forehead thermometer. So we designed a smart detection box to make detection more convenient and quicker and improve the efficiency of prevention and control.

2. System overview

The device is a temperature detection box with identification. The detection box can realize non-contact temperature recognition, face recognition, identity comparison, mask recognition, voice prompts and other functions, data analysis through[2] camera and sensor recognition, and human-computer interaction through OLED display. The main functions that can be achieved are: 1. Non-contact temperature detection, and at the same time, by setting the alarm threshold, a faster and more accurate identification of epidemic prevention standards can be achieved, and then voice broadcast reminders can effectively improve the prevention and control efficiency. 2. Data collection through the camera, real-time data transmission through the serial port, data analysis through the MCU, and then voice broadcast of the mask wearing situation, mainly used for standard mask recognition in the current epidemic. 3. Through oled display, perform human-computer interaction, enter face data, and store it in the database, so that face information can be effectively recognized, identity recognition, and prevention and control become more efficient. 4. The device does not need to be held in hand and can be fixed anywhere, which can effectively avoid cross-infection.It also supports self-settable temperature alarm thresholds, voice broadcast reminders, and fast and accurate non-contact temperature measurement.。

3. The hardware structure design

3.1. Core controller

The core controller of the device uses the stm32f1 series chip, which has obvious characteristics. ARM's high-performance "Cortex-M3" core: 1.25DMips/MHz, while the ARM7TDMI is only 0.95DMips/MHz, first-class peripherals. 1 μ s dual 12-bit ADC, 4Mbit/s UART, 18Mbit/s SPI, 18MHz I/O flip speed. Low power consumption, consumes 36mA at 72MHz (all peripherals are in working state), drops to 2 μ A in standby, maximum integration, reset circuit, low voltage detection, voltage regulator, precise RC oscillator, etc[3].

3.2. Image acquisition equipment

The device uses an openmv4 camera, which is an open source, low-cost, and powerful machine vision module. With STM32F427CPU as the core, it integrates the OV7725 camera chip. On the small hardware module, the core machine vision algorithm is efficiently implemented in C language, and the Python programming interface is provided. The adopted STM32F427 has abundant hardware resources, leading to UART, I2C, SPI, PWM, ADC, DAC and GPIO interfaces to facilitate the expansion of peripheral functions. The USB interface is used to connect to the integrated development environment OpenMVIDE on the computer to assist in programming, debugging and updating firmware. The TF card slot supports large-capacity TF cards, which can be used to store programs and save photos. The open source environment and rich Python libraries make it very conducive to the development of visual identity.

3.3. Sensor

The MLX90614 infrared temperature sensor is used. MLX90614 is an infrared thermometer for non-contact temperature measurement. Infrared temperature measurement is to determine the temperature of the object according to the infrared radiation energy of the object to be measured. It does not contact the object to be measured, and has no influence on the temperature distribution field of the object to be measured. It has high temperature resolution, fast response speed, and wide temperature measurement range. Restricted by the upper limit of temperature measurement, good stability and other characteristics. The measured target temperature and ambient temperature of MLX90614 are output through the IIC interface. The sensor is directly connected to the core board, and is packaged as a temperature measurement structure through a printed external package. At the same time, the circuit is connected in series with an infrared indicator light to confirm and indicate the object under test.

We need to read the data from the MLX90614 infrared temperature sensor, no matter whether the data is in EEPROM or RAM, we can read it in the same way. We have already said that the command byte for operating

Fig.1 MLX90614 infrared temperature sensor

EEPROM and RAM consists of a 3-bit command and a 5-bit address. We implement the data reading function as follows:

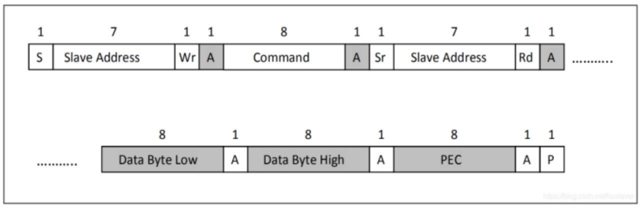


Fig.2 Data reading function

3.4. Display

The display uses an OLED display module, OLED, or Organic Light-Emitting Diode (Organic Light-Emitting Diode), also known as Organic Electroluminesence Display (OELD). OLED has excellent characteristics such as self-luminous, no backlight, high contrast, thin thickness, wide viewing angle, fast response speed, can be used for flexible panels, wide operating temperature range, simple structure and manufacturing process, etc. It is considered to be The next-generation flat panel display emerging application technology.

OLED display technology has self-luminous characteristics. It uses very thin organic material coatings and glass substrates. When current passes through, these organic materials will emit light. The OLED display screen has a large viewing angle and can save power. This kind of display device has been used in MP3 players since 1988.

LCDs require backlighting, but OLEDs do not, because it is self-luminous. For the same display, the OLED effect should be better. With current technology, it is difficult to increase the size of OLEDs, but the resolution can be very high.

The oled display is as follows, which can effectively carry out human-computer interaction

3.5. Voice broadcast

The voice broadcast function of the device is realized by syn6288 module, which broadcasts through the speaker. The SYN6288 Chinese speech synthesis chip used by Speech Synthesizer Bee is a speech synthesis chip with high cost performance, more natural speech synthesis, and for high-end applications[4]. SYN6288 receives the text to be synthesized through the asynchronous serial port, and realizes the conversion of text to voice (TTS). The module is directly connected to the MCU.

4. System process

4.1. System operation process

According to the needs of users, the detection mode is selected. The detection mode includes mask recognition, identity entry and temperature detection.

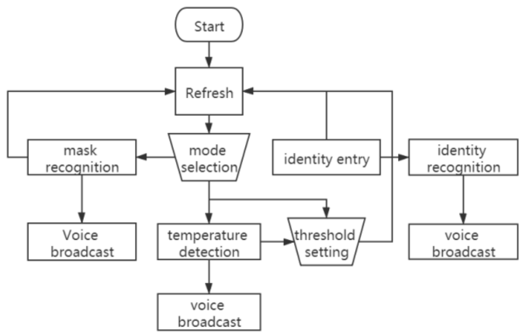


Fig.3 System operation process

When the user enters the mask recognition, the camera module will automatically turn on to take pictures, compare and analyze the photographed image with the mask recognition set that has been trained, and get the result. Mask recognition is based on the neural network model trained by openmv using EDGE IMPULSE's online website. Training the neural network model is to perform machine learning. Through machine learning, you can determine whether there are people in the photo and whether to wear masks and other functions. By establishing two categories, one face and one mask, one hundred images are collected each, and the collected data is sent to the EDGE IMPULSE website for model training. After DSP preprocessing, the following analysis is obtained:

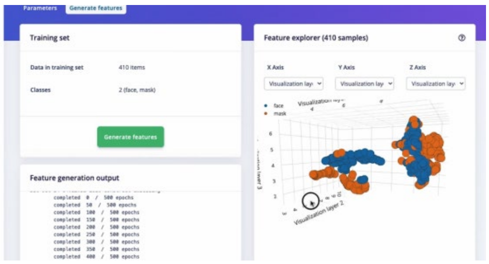


Fig.4 DSP preprocessing results

It can be seen that the two colors of the training model face and mask are clearly distinguished and are not mixed together. After ten rounds of training, the training data obtained is as follows:

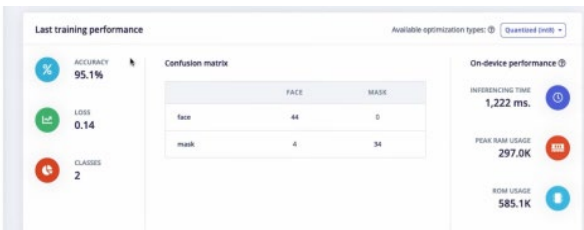


Fig.5 DSP training data

It can be seen that our mask recognition model has an accuracy of 95.1%. Run the training set on openmv and the mask recognition is completed.

4.2. Detection processing

After the user enters the temperature detection, the threshold value of the temperature measurement device can be set by key operation[5], so that the high temperature alarm system remains effective. After the MLX90614 non-contact temperature measurement module collects (DataH:DataL) data, according to the formula:

T=(DataH:DataL)*0.02-273.15

To get the final temperature T, we use the bubble sorting method when collecting, and each temperature measurement will collect ten times of data. After sorting, remove the abnormal temperature value and average it as the final temperature measurement data, so as to avoid abnormal temperature And make the final temperature data more accurate.

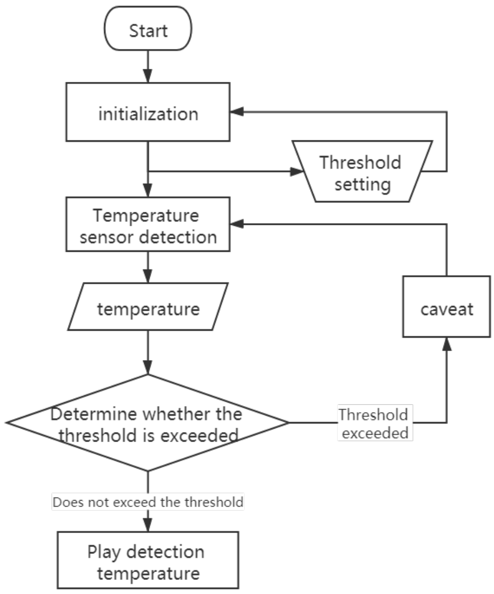


Fig.6 Temperature measurement flow chart

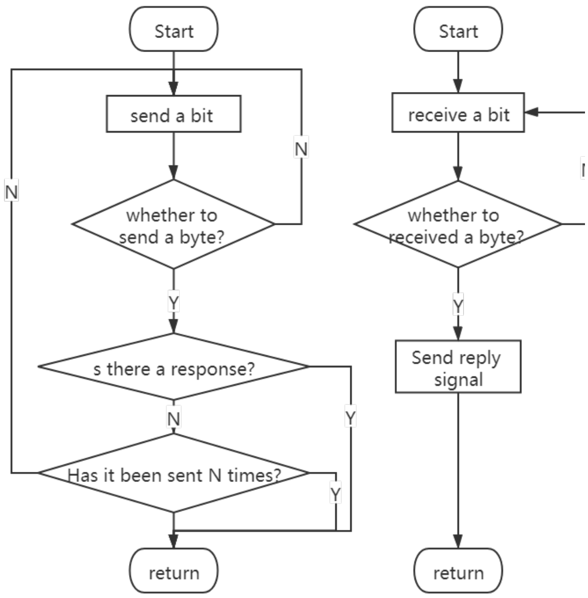


Fig.7 MLX90614 send and receive byte flow chart

4.3. System circuit design

The system is powered by 5v with energy, voltage conversion is connected to the core board through a voltage stabilizing module, the circuit board adopts a homemade pcb circuit board, integrated oled, MLX90614, openmv and power interface, which can make the signal transmission more stable and reduce external signal interference , To ensure the stability and accuracy of the test results.

5. Test and conclusion

5.1. Test plan

- The encapsulated temperature detection gun is used for non-contact detection of different water temperatures, such as 35.7°C, 41.2°C, and 80°C, and the results are tested with the threshold 25-48°C set in advance
- Measure human body temperature, floor temperature, desktop temperature without contact within a certain distance
- Enter the identity in advance, save it in the database, and then perform face recognition and check the face recognition function
- There are two cases of wearing masks and not wearing masks by three people, and testing the three groups wearing masks to test their stability
- Detect the on-site deep learning function of the device, Provide three unfamiliar faces, learn their facial identity within 5 minutes and record them in the database, and then the temperature measuring box identifies whether they are the input personnel, and detects whether they have the function of deep learning

5.2. Test results

After repeated testing, the device has the above functions, and can identify and detect very stably. It can quickly and accurately identify human body temperature and water temperature and broadcast it. It can accurately and stably recognize the wearing of a mask. It can learn on-site, automatically enter the identity and recognize it. . The temperature detection efficiency is greatly improved, and the identification is accurately prevented and controlled.

References

1. Yang Li, Yang Zhen "Principle and Technology of Infrared Thermal Imaging Temperature Measurement" Science Press, 2012.6.
2. Li Min, "Design of Wireless Infrared Temperature Measurement System" Guangdong Vocational College of Agriculture, Industry and Commerce, 2020.3.
3. Zheng Zaisheng "Research on Automatic Fare Collection System Based on Face Recognition" China Computer Room Facilities Engineering Co., Ltd., 2020.9
4. Li Chengyong, Tan Hanzhong, Wang Sha, et al. Intelligent "seeking" car control system based on Open MV[J]. Liquid Crystal Display,2020,08:870-876.
5. Ye Siyan, Zhang Yongning, Li Jin, et al. Design of smart follow luggage based on UWB positioning and Open MV recognition[J]. Journal of Chengdu University of Technology,2020,03:39-33.

Landslide Area Detection from Synthetic Aperture Radar Images Using Convolutional Adversarial Autoencoder and One-class SVM

Shingo Mabu
Graduate School of Sciences and Technology for Innovation, Yamaguchi University
2-16-1 Tokiwadai, Ube, Yamaguchi 755-8611, Japan

Soichiro Hirata
Graduate School of Sciences and Technology for Innovation, Yamaguchi University
2-16-1 Tokiwadai, Ube, Yamaguchi 755-8611, Japan

Takashi Kuremoto
Graduate School of Sciences and Technology for Innovation, Yamaguchi University
2-16-1 Tokiwadai, Ube, Yamaguchi 755-8611, Japan
E-mail: mabu@yamaguchi-u.ac.jp, wu@yamaguchi-u.ac.jp

Abstract

An anomaly detection model using deep learning for detecting disaster-stricken (landslide) areas in synthetic aperture radar images is proposed. Since it is difficult to obtain a large number of training images, especially disaster area images, with annotations, we design an anomaly detection model that only uses normal area images for the training, where the proposed model combines a convolutional adversarial autoencoder and one-class SVM. In the experiments, the ability in detecting normal and abnormal areas is evaluated.

Keywords: anomaly detection, convolutional neural network, adversarial autoencoder, one-class SVM, synthetic aperture radar, landslide

1. Introduction

In Japan, typhoons often approach and pass from July to September every year because of the effects of westerly winds and high atmospheric pressures. Landslide disasters frequently occur due to heavy rains caused by typhoons, which leads to major accidents related to transportation and human life. Until now, observation of the disaster areas, e.g., landslides, has been mainly conducted by aircraft[1]. However, in recent years, remote sensing using satellite images has attracted attention as a method for observing a wide area[2]. In satellite remote sensing, a sensor is mounted on an artificial satellite to observe the ground surface of the earth, and the obtained image is analyzed. Therefore, we can see the damages without going directly to the disaster areas. Typical satellite images include optical images that are obtained from sunlight reflection, and synthetic aperture radar (SAR) images that are obtained by sensors that emit microwaves to the ground surface of the earth. It is easy for a human to interpret optical images, but they cannot be observed at night or in bad weather. On the other hand, SAR can observe the surface of the earth regardless of time and weather; thus, SAR images are useful for rapid rescue activities at night and in bad weather conditions. However, it is difficult for human eyes to interpret SAR images, unlike optical images. Therefore, many methods that analyze SAR images have been proposed to detect disaster areas rapidly[3,4], and machine learning techniques, especially deep learning, have also been applied to landslide area detection[5].

When deep learning is applied to disaster area detection, a large number of training samples are necessary. However, shooting conditions of SAR images are different time by time, e.g., seasons, and it is also difficult to obtain a sufficient number of training samples with annotation (correct class labels) immediately after a disaster occurs. To overcome this problem, we propose an anomaly detection system that detects abnormal images that deviate from the features of normal images. The proposed method uses only normal images for the training without using abnormal (disaster) images. Normal images are relatively easier to obtain compared with abnormal images, thus anomaly detection is useful in a practical situation of disaster area detection. The proposed method consists of a feature extraction part and a detection part. The feature extraction is realized by convolutional adversarial autoencoder (CAAE) that is the extension of adversarial autoencoder (AAE)[6], where the layers of AAE are replaced with convolution and deconvolution layers. CAAE is trained with unsupervised learning, thus the cost of the annotation is not necessary. The detection part is realized by one-class SVM (OCSVM)[7] that detects outliers that deviate from the normal region.

This paper is organized as follows. In section 2, the mechanism and features of SAR is introduced. In Section 3, the proposed method that combines CAAE and OCSVM is explained. In section 4, the experimental conditions and results are described. Finally, section 5 is devoted to conclusions.

2. Synthetic Aperture Radar (SAR)

SAR is an active image radar that synthesizes small antennas mounted on a platform, such as an aircraft or satellite, to realize large virtual antennas and generates high-resolution radar images[8,9,10)]. Because SAR is an active sensor that emits microwaves, it is possible to observe the surface of the earth regardless of the presence or absence of sunlight and clouds. SAR images are applied to the research fields of agriculture, disaster, oceans, earth science, and so on[11,12)]. SAR emits microwaves and receives the reflected microwaves from the surface of the earth (Fig. 1). When a microwave emitted from the SAR antenna enters a conductor or dielectric, a current is induced, and the microwave is reemitted from the induced current. This is called scattering and scattering in the opposite direction of the incident wave is called backscattering. Because backscattering is the diffuse reflection caused by scattering, backscattering is different from specular reflection.

SAR receives the backscattering and executes image reproduction. The scattering intensity of the microwaves strongly depends on the frequency, wavelength and

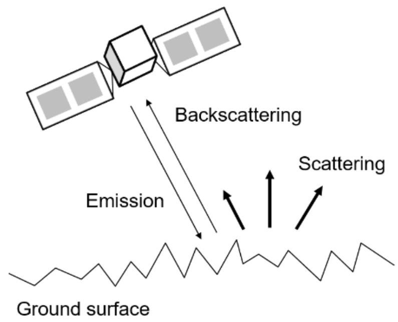


Fig. 1. Microwave emission, Scattering, and backscattering of SAR[10]

electric characteristics (dielectric constant, etc.) of the scatterer. Therefore, for example, seawater or cars made of metal strongly reflect microwaves because a current is easily induced. On the other hand, sand and trees have low reflectivity because hardly any current is induced. In each pixel of SAR images, the intensity of the received microwave is recorded.

3. Materials and methods

3.1. *Dataset*

A SAR image used in this paper is shown in Fig. 2. It shows the northern Kyushu area in Japan on July 7, 2017, taken by Advanced Land Observing Satellite No. 2 (ALOS-2). Large-scale landslides due to the torrential rain from July 5 to 6, 2017, occurred in this area. The yellow areas in Fig. 2 show the locations of the landslide that were annotated by the Geospatial Information Authority of Japan. The size of the image is 6648×4360 [pixels] and the resolution per pixel is about $3m^2$.

In this paper, ROI (region of interest)-based anomaly detection is carried out, that is, we first divided the whole

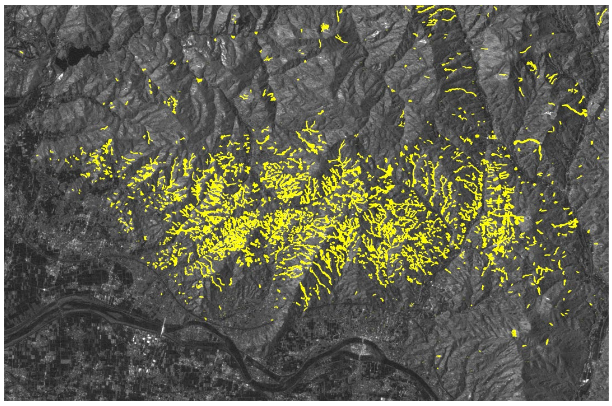


Fig. 2. SAR image and landslide areas (yellow areas)

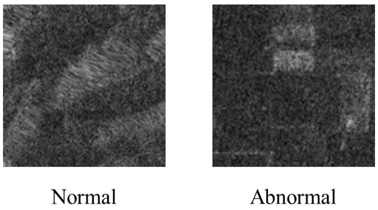


Fig. 3. Examples of 128 × 128 [pixels] ROI images

SAR image into 128 × 128 [pixels] ROI images. Fig. 3 shows examples of normal and abnormal ROI images. Note that the target area for the detection of normal and abnormal is the central 32 × 32 [pixels] area of each ROI image. In other words, the surrounding area of the target area is used as a piece of supporting information (context) for the detection. The ROI images were split into training data including normal only, testing data of normal, and testing data of abnormal.

3.2. *Method*

The overview of the anomaly detection model designed in this paper is shown in Fig. 4, where CAAE generates feature fector $\boldsymbol{z}$, $\boldsymbol{z}$ is transformed by principal component analysis (PCA), and OCSVM classifies $\boldsymbol{z}$ as normal or abnormal.

3.1.1. *Feature extraction using Convolutional Adversarial Autoencoder (CAAE)*

In this paper, the extended model of AAE[6], named CAAE, is used for the feature extraction of SAR images. Fig. 5

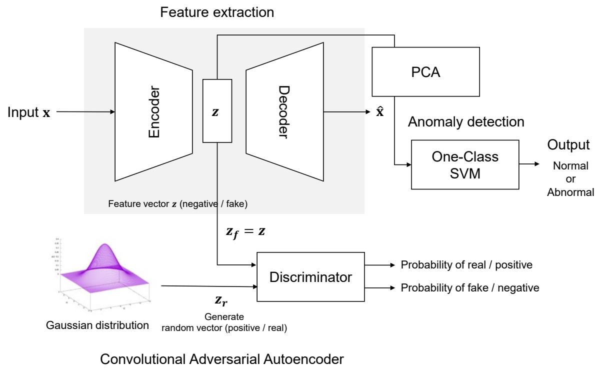


Fig. 4. Examples of 128 × 128 [pixels] ROI images

shows the detailed structure of CAAE that extracts features from ROI images. The upper part of Fig. 5 is a convolutional autoencoder (CAE) that contains an encoder and a decoder. The encoder outputs a 128-dimensional feature vector $\boldsymbol{z}$, and the decoder outputs the reconstructed ROI images. The encoder consists of some Block1 and fully-connected layers, where a Block1 contains convolution and max pooling layers along with batch normalization and rectified linear unit (ReLU). The decoder consists of some Block2, fully-connected and deconvolution layers, where a Block2 contains a deconvolution layer along with batch normalization and ReLU. The loss function (LOSS1 in Fig. 5) is the mean squared error between the input and output images, and the weights of the encoder and decoder are updated by error backpropagation with Adam[13] to minimize LOSS1. The encoder and discriminator at the lower part of Fig. 5 are adversarial networks, where the discriminator

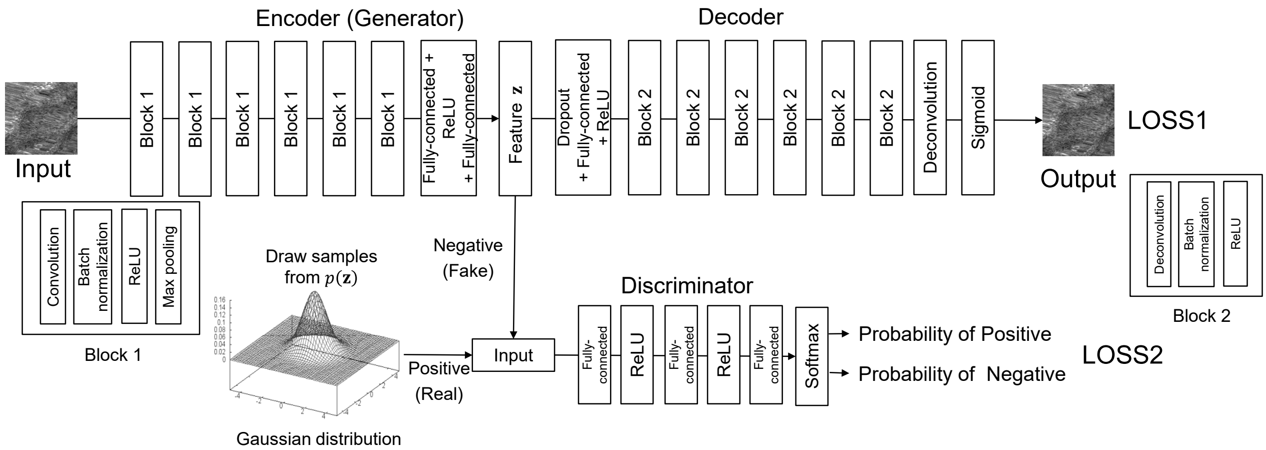


Fig. 5. Detailed structure of CAAE

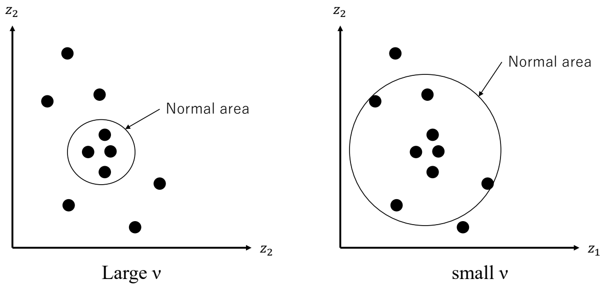


Fig. 6. Effects of ν in OCSVM

distinguishes the vectors generated by Gaussian distribution (regarded as positive or real samples) from the feature vector $\boldsymbol{z}$ inputted from the encoder (regarded as negative or fake samples). The loss function (LOSS2 in Fig. 5) is the softmax cross-entropy based on the positive and negative probabilities outputted by the discriminator. The discriminator updates the weights to minimize LOSS2, while the encoder updates the weights to maximize LOSS2. As a result of the above learning, the feature vector $\boldsymbol{z}$ captures the features of the input images and its distribution follows a Gaussian distribution. The constraint imposed by the adversarial networks increases the robustness of the feature extraction executed by the convolutional autoencoder.

3.3. *Anomaly detection by One-class SVM (OCSVM)*

The feature vector $\boldsymbol{z}$ generated by CAAE is inputted to OCSVM. SVM is originally a supervised-learning-based classification algorithm for two classes, which needs training data with class labels. OCSVM needs only one-class data for the training, and after the training, OCSVM regards the data that deviates from the learned class data as outliers, that is, abnormal. Anomaly detection for the testing images in this paper is carried out by OCSVM. OCSVM has a parameter ν that determines the proportion of normal and abnormal areas. For example, when ν is set at a large value as shown in Fig. 6 (left), the normal area becomes small. In this case, only the data that is very near to the center of the normal area is regarded as normal, which increases the sensitivity of abnormal, but may increase false positives, that is, the true normal (negative) is misclassified as abnormal (positive). Conversely, when ν is small, the normal area becomes large, which increases the specificity, but may increase false negatives, that is, the true abnormal is regarded as normal. Therefore, the value of ν should be determined carefully considering the trade-off between false positive and false negative.

Table 1. The number of ROI images in condition 1

Training/Testing	The number of images
Training (normal only)	13542
Testing (normal)	164
Testing (abnormal)	164

4. Experiments

4.1. *Condition 1*

The class labels of ROIs are given by the following rule. When 60% or more area of the central 32×32 area of each ROI image contains the abnormal area, it is regarded as abnormal, otherwise, normal. The number of training data (normal only), that of testing data (normal), and that of testing data (abnormal) are shown in Table 1. CAAE is trained with the training data and all the training data are encoded as the feature vectors. Then, PCA is applied to the feature vectors to reduce the number of dimensions from 128 to 40. The feature vectors after the dimensional reduction by PCA are used to train OCSVM. Finally, the testing images encoded and transformed by CAAE and PCA are classified as normal and abnormal.

Table 2 shows the results obtained by the combination of CAE, PCA, and OCSVM (called conventional method), and Table 3 shows those obtained by the combination of CAAE, PCA, and OCSVM (called proposed method). The best accuracy obtained by the conventional method is 55.7% (ν=0.5) and that obtained by the proposed method is 57.9% (ν=0.4). Therefore, we can confirm that CAAE is better than CAE as a feature extraction method, however, there is a large room for improvement on the accuracy.

4.2. *Condition 2*

In condition 2, the class labels of ROIs are given by the following rule. The labeling rule of abnormal areas is the same as condition 1, that is, when 60% or more area of each ROI image contains abnormal, it is regarded as abnormal. In the case of the labeling of normal, ROIs without containing any abnormal areas are regarded as normal. In condition 1, the threshold that separates the normal and abnormal ROIs is 60%, which means that even if the ROIs contains 59% abnormal areas, they should be classified as normal. To clearly separate the

Table 2. Results obtained by CAE + PCA + OCSVM (Conventional method) in condition 1

		ν					
		0.2	0.3	0.4	0.5	0.6	0.7
Evaluation metric [%]	Accuracy	53.4	54.9	55.6	**55.7**	49.4	47.4
	Sensitivity	23.4	32.1	41.1	51.1	56.1	63.8
	Specificity	83.4	77.8	70.1	60.3	42.7	31.1

Table 3. Results obtained by CAAE + PCA + OCSVM (Proposed method) in condition 1

		ν					
		0.2	0.3	0.4	0.5	0.6	0.7
Evaluation metric [%]	Accuracy	55.9	56.6	**57.9**	57.4	54.4	52.4
	Sensitivity	24.5	31.5	37.1	39.4	41.9	47.2
	Specificity	87.3	81.6	78.7	75.4	66.9	57.6

features of normal and abnormal, condition 2 adopts the above labeling rule. Table 4 shows the numbers of training and testing data. The procedure of executing CAAE, PCA and OCSVM is the same as condition 1. Table 5 shows the results of the conventional method and Table 6 shows those of the proposed method. The best accuracy obtained by the conventional method is 61.4% (ν=0.2) and that obtained by the proposed method is 66.0% (ν=0.5); thus we can see that the accuracy becomes better than condition 1 and the accuracy obtained by CAAE is better than that by CAE.

5. Conclusions

In this paper, we proposed an anomaly detection method for classifying normal (non-disaster) areas and abnormal (disaster/landslide) areas, where the proposed method consists of CAAE, PCA, and OCSVM. In the experiments, we evaluated the detection performance in the two conditions. From the results, it was clarified that CAAE is better than CAE as a feature extraction method. The remaining problem is the classification of ROI images that contain small abnormal areas. Since it is difficult to make a sharp boundary between normal and abnormal, other method such as segmentaion models would be useful. In addition, the proposed method should be evaluated on other areas of disasters to confirm the generalization ability.

Table 4. The number of ROI images in condition 2

Training/Testing	The number of images
Training (normal only)	7967
Testing (normal)	164
Testing (abnormal)	164

Table 5. Result obtained by CAE + PCA + OCSVM (Conventional method) in condition 2

		ν					
		0.2	0.3	0.4	0.5	0.6	0.7
Evaluation metric [%]	Accuracy	**61.4**	60.4	61.2	59.8	58.7	56.2
	Sensitivity	39.5	47.7	56.6	64.3	68.1	73.3
	Specificity	83.3	73.1	65.8	55.5	49.3	39.1

Table 6. Result obtained by CAAE + PCA + OCSVM (Proposed method) in condition 2

		ν					
		0.2	0.3	0.4	0.5	0.6	0.7
Evaluation metric [%]	Accuracy	64.0	64.8	65.5	**66.0**	62.7	59.6
	Sensitivity	39.6	44.4	50.0	54.8	56.9	60.6
	Specificity	88.3	85.1	80.9	77.2	68.5	58.5

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP20H02417.

References

1. T. Lillesand, R. W. Kiefer, J. Chipman, *Remote sensing and image interpretation*, John Wiley & Sons, 2014

2. Y, Liu and L. Wu, Geological disaster recognition on optical remote sensing images using deep learning, *Procedia Computer Science* **91**, 2016, pp. 566-575
3. H. Shahabi, M. Hashim, Landslide susceptibility mapping using gis-based statistical models and remote sensing data in tropical environment, *Scientific reports* **5**, 9899, 2015
4. N. Casagli, V. Tofani, A. Ciampalini, F. Raspini, P. Lu, S. Morelli, Txt-tool 2.039-3.1: Satellite remote sensing techniques for landslides detection and mapping, *Landslide Dynamics: ISDR-ICL Landslide Interactive Teaching Tools*, Springer, 2018, pp. 235-254
5. P. Liu, Y. Wei, Q. Wang, Y. Chen, J. Xie, Research on post-earthquake landslide extraction algorithm based on improved u-net model, *Remote Sensing* **12**(5), 894, 2020
6. A. Makhzani, J. Shlens, N. Jaitly and I. J. Goodfellow, Adversarial Autoencoders, *International Conference on Learning Representations*, 2016
7. L. M. Manevitz and M. Yousef, One-Class SVMs for Document Classification, *Journal of Machine Learning Research*, 2001, pp. 139-154
8. J. P. Fitch, *Synthetic aperture radar*, Springer Science & Business Media, 2012
9. K. Ouchi, *Principles of synthetic aperture radar for remote sensing*, Tokyo Denki University Press, 2009 (in Japanese)
10. S. Mabu, Y. Nakayama and T. Kuremoto, Landslide Classification from Synthetic Aperture Radar Images Using Convolutional Neural Network with Multichannel Information, Journal of Signal Processing, **24**(2), 2020, pp. 61-73
11. C. Atzberger, Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs, *Remote Sensing*, **5(**2), 2013, pp.949-981
12. E. C. Barret, *Introduction to environmental remote sensing*, Routledge, 2013
13. D. Kingma and J. Ba, Adam: A method for stochastic optimization. *arXiv preprint* arXiv:1412.6980, 2014

Anomaly Detection in Time Series Data Using Support Vector Machines

Umaporn Yokkampon[1], Sakmongkon Chumkamon[1], Abbe Mowshowitz[2] , Eiji Hayashi[1]
[1]Department of Computer Science and Systems Engineering, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan
[2]Department of Computer Science, The City College of New York, 160 Convent Avenue, New York, NY 10031, USA
E-mail: may@mmcs.mse.kyutech.ac.jp, m-san@mmcs.mse.kyutech.ac.jp, amowshowitz@ccny.cuny.edu, haya@mse.kyutech.ac.jp
www.kyutech.ac.jp

Abstract

Analysis of large data sets is increasingly important in business and scientific research. One of the challenges in such analysis stems from uncertainty in data, which can produce anomalous results. In this paper, we propose a method of anomaly detection in time series data using a Support Vector Machine. Three different kernels of the Support Vector Machine are analyzed to predict anomalies in the UCR public data set. Comparison of the three kernels shows that the defined parameter values of the RBF kernel are critical for improving the validity and accuracy in anomaly detection. Our results show that the RBF kernel of the Support Vector Machine can be used to advantage in detecting anomalies.

Keywords: Anomaly detection, Support Vector Machine, Data mining, Factory automation.

1. Introduction

Research on anomaly detection is of great interest in machine learning and data mining. Detecting anomalies or finding outliers involves identifying abnormal or inconsistent patterns in a dataset. Abnormal data often results from unauthorized activity. Credit card fraud offers a well known example. Transactions with a stolen or fake credit card can produce suspicious data. A fake card can be made by copying information from an authorized card and using it to create a new unauthorized one. Data such as personal identifying information may be obtained through phishing or from employees who work in credit card companies [1]. Another source of abnormal data may derive from unauthorized intrusions in networks. Abnormal traffic or user actions are common signs of intrusions, which may occasion breaches of sensitive or confidential data. Intrusions may also cause sensor networks to generate erroneous data. When a sensor malfunctions, it is unable to capture data correctly and thus may produce anomalies. Abnormal changes in data sources may also result in anomalies [2]. Anomaly detection typically uses data mining and machine learning techniques to detect abnormal activities in systems. Over the past decade, many anomaly detection techniques have been developed, including Support Vector Machines (SVM), a supervised machine learning algorithm that can be used for classification or to solve regression problems. In practice, the SVM algorithm is applied with the kernel that transforms an input data space into the required form. Kernel function and kernel parameters affect the performance of SVM. The selection quality of SVM parameters and kernel functions has an effect on learning and generation performance. Appropriate kernel function and associated parameters should be selected to obtain optimal classification performance. When an appropriate kernel function and parameters are selected, the prediction error of SVM can be minimized.

This paper reports on application of the support vector machine method to eight real world time series data sets to detect anomalies using three different kernels for analysis and prediction. In addition, SVM kernels are compared for effectiveness based on AUC, Precision, Recall, and F1-Score criteria.

2. Related Work

Anomaly detection is widely used in many fields, and various methods have been proposed over the years.

In 2003, Ma et al. [3] used one class SVMs for prediction which require a set of vectors as input instead of a time series. They convert the time series into a phase-space using a time-delay embedding process, i.e., by creating overlapping subsequences from a given long sequence. These vectors are projected into an orthogonal subspace that acts as a high pass filter to filter out the low frequency components and allow only high frequency ones (anomalies).

In 2005, Kim and Cha [4] tested the effectiveness of SVM in detecting masquerade activities. Their experiments showed that their model could detect this type of attack with an accuracy of 80.1%., thus empirically demonstrating that SVM offers an effective method for masquerade detection.

In 2008, Sugumaran et al. [5] developed fault diagnostics of roller bearings using a neighborhood score multiclass SVM in EDM machining. The roller bearing is one of the most widely used elements in a rotary machine. RBF is used as a kernel function. This research used a kernel based neighborhood score multiclass SVM for classification and a decision tree for addressing the future selection process. The study of a multiclass SVM showed its effectiveness in diagnosing the fault conditions of the bearing.

In 2012, Caydas and Ekici [6] applied SVM to develop prediction models for surface roughness in the turning process of AISI 304 austenitic stainless steel. The relevant parameters are cutting speed, feed rate and depth of cut using RBF as a kernel function. Three different SVM models were developed, namely, LS-SVM, spider SVM and SVM-KM. Spider SVM offered the best prediction model for surface roughness.

In 2015, Yu et al. [7] proposed a prediction model of bus arrival time based on SVM and a forgetting factor. The actual time of bus arrival at each time point is predicted by taking account of the time, weather and certain historical data as input vectors. A k-Nearest Neighbor Model for Multiple-Time-Step Prediction was introduced to predict short-term traffic conditions. Moreover, the Grubbs' test method was applied to remove outliers from the input data.

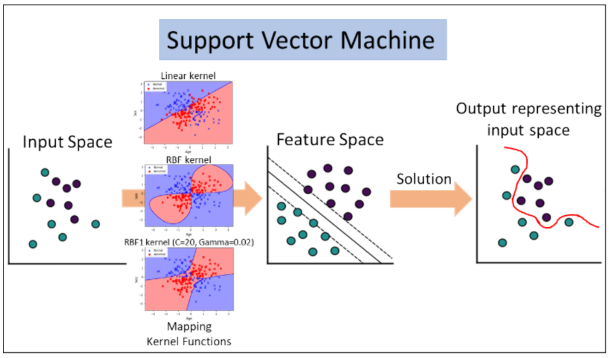


Fig. 1. Structure of anomaly detection in time series data used SVM. We used eight time series data sets processed by SVM, and three different kernels based on AUC, Precision, Recall and F1-Score criteria.

3. Support Vector Machine Algorithm

The Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for classification as well as regression problems. The SVM algorithm's goal is to create the best line or decision boundary that can decompose an n-dimensional space into sets supporting categorization of new data points. Hyperplanes define the boundaries in this space.

For a given dataset x with a number i of training data, SVM finds the maximum margin hyperplane separating different classes of data [8]:

$$x = (x_i, y_i),\ x_i \in \mathbb{R}^p,\ y_i \in \{-1,1\},\ \forall\ i = 1,2,...,N \quad (1)$$

where x_i is the p-dimensional input vector and y_i is the output value (1 or −1). A decision vector separating two classes is given by:

$$w^T \cdot x + b = 0 \quad (2)$$

where w^T is the optimal weighting vector and b is the bias. For linearly separable training data, margins are defined as:

$$w^T \cdot x + b = 1 \text{ and } w^T \cdot x + b = -1 \quad (3)$$

The distance between the margins is given by $2/\|w^T\|$. Hence, the objective function is to minimize $\|w^T\|$. In practice, it is not easy to linearly decompose the training dataset. Let C be the regularization parameter that defines the separation of two classes and the error when using a training dataset. The hyperplane is determined by minimizing:

$$C\sum_{i=1}^{N}\varepsilon_i + \frac{1}{2}\|w\|^2 \quad (4)$$

with constraints $t_i y(x_i) \geq 1-\varepsilon_i,\ i=1,...,N$ where t_i is the target value and ε_i is the set of slack variables.

Instead of employing a minimization model (4), the problem may be formulated using Lagrangian dual multipliers α as:

$$\max \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i, x_j) \quad (5)$$

subject to:

$$0 \leq \alpha_i \leq C \ \ \forall\ i=1,2,...,n \text{ and } \sum_{i=1}^{n}\alpha_i y_i = 0 \quad (6)$$

Using the "kernel trick" reduces the complexity of the optimization problem that now only depends on the input space instead of the feature space:

The objective function for SVM with nonlinear kernel has the form:

$$\max \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j k(x_i, x_j) \quad (7)$$

4. Kernels

Kernel methods form a popular class of machine learning techniques for various tasks. An important feature offered by kernel methods is the ability to model complex data through the use of the kernel trick (Schölkopf & Smola, 2002).

In this paper, two types of kernel functions are chosen and evaluated, namely the linear and radial basis function (RBF). The mathematical formula for the said functions are as follows:

4.1. *Linear kernel*

$$k(x_i, x_j) = (x_i, x_j) \quad (8)$$

4.2. *Radial Basis Function (RBF) kernel*

$$k(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / (2\sigma^2)) \quad (9)$$

The two kernels have their own advantages and limitations. The linear kernel offers ease of performance as well as an ability to deal with small and linearly separable samples. The RBF kernel, on the other hand, is known as a good mapping function since it can be used for all kinds of samples, small or large with both high and low dimensions [9]. The SVM performance for each kernel will be evaluated in this study to determine the optimal kernel.

5. Experiments

This section introduces the data sets and the evaluation metric employed. We have compared the three kernels and evaluated their effectiveness for anomaly detection in Support Vector Machines.

5.1. *Data sets*

Time series data obtained from UCR public data set [10] were used to evaluate effectiveness. The details of the datasets are given in Table I. All datasets are presented in time series form, and every data point is manually labeled. For all datasets, we designated the minority class as an anomaly class. Twenty percent of the data was used for testing.

Table 1. Summary of the datasets

Datasets	Length	Number of instances	Anomaly Ratio
ItalyPowerDemand	24	1096	0.49
Wafer	152	7164	0.11
SonyAIBORobotSurface2	65	980	0.38
ECGFiveDays	136	884	0.50
TwoLeadECG	82	1162	0.50
MoteStrain	84	1272	0.46
Herring	512	128	0.40
Strawberry	235	983	0.36

5.2. *Performance Evaluation*

The accuracy of an anomaly detection method is evaluated using the area under the curve (AUC) of the receiver operating characteristic (ROC), Precision (Pre), Recall (Rec), and F1-Score, defined as follows:

$$\text{Pre} = \frac{TP}{TP+FP} \tag{10}$$

$$\text{Rec} = \frac{TP}{TP+FN} \tag{11}$$

$$\text{F1} = 2\times\frac{\text{Pre}\times\text{Rec}}{\text{Pre}+\text{Rec}} \tag{12}$$

where TP is the correctly detected anomaly, FP is the falsely detected anomaly, TN is the correctly assigned normal, and FN is the falsely assigned normal.

6. Results and Discussion

The efficiency of the following three SVM kernels are compared:

1. Linear Kernel
2. RBF Kernel (Default parameters value)
3. RBF1 Kernel (We define the parameters C = 20, γ = 0.02)

We performed experiments on accuracy of analysis and prediction of anomalies for eight time series data sets using the three different SVM kernels. Accuracy of analysis and prediction can be measured by the AUC as shown in the ROC in Fig. 2 - 9. The blue line is the Linear

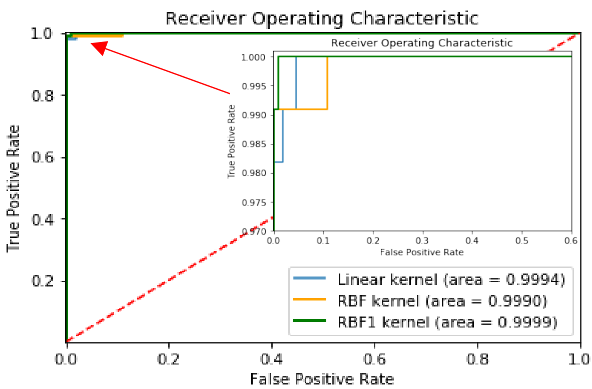


Fig. 2. The kernel performance comparison of Linear, RBF and RBF1 for testing ItalyPowerDemand dataset using ROC.

Kernel of SVM, the orange line is the RBF kernel, and the green line is the RBF1 kernel.

Fig. 2. shows that the RBF1 kernel is slightly more efficient than the Linear and RBF kernels for the ItalyPowerDemand data set. However, all three kernels yield almost 100 percent accuracy.

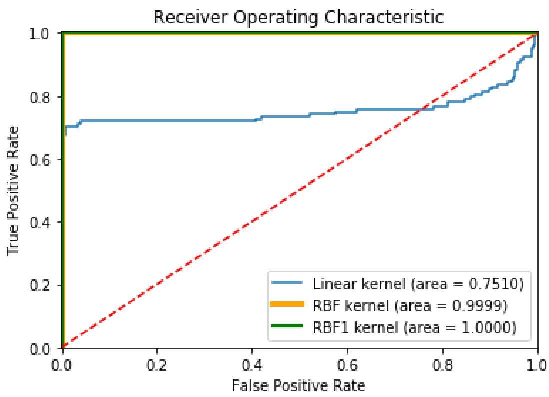


Fig. 3. The kernel performance comparison of Linear, RBF and RBF1 for testing Wafer dataset using ROC.

Fig. 3. shows that the RBF and RBF1 kernels are more efficient than the Linear kernel, and that the RBF

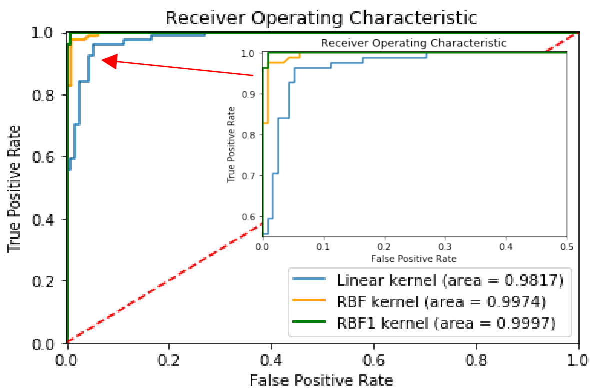


Fig. 4. The kernel performance comparison of Linear, RBF and RBF1 for testing SonyAIBORobotSurface2 dataset using ROC.

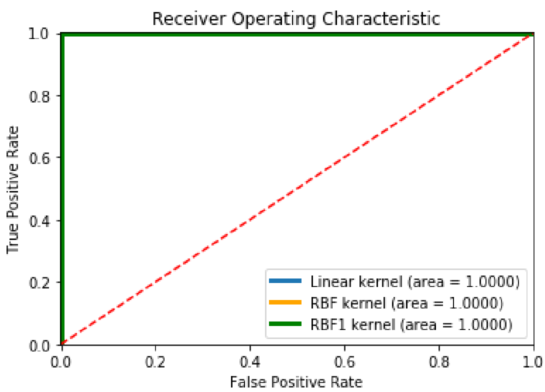


Fig. 5. The kernel performance comparison of Linear, RBF and RBF1 for testing ECGFiveDays dataset using ROC.

kernel is almost 100 percent accurate. In particular, the RBF1 kernel gives a ROC value perfectly for the Wafer data set.

Fig. 4. reveals that RBF1 is slightly more efficient than the linear and RBF kernels for the SonyAIBORobotSurface2 data set. In particular, the RBF and RBF1 kernels provide almost 100 percent accuracy.

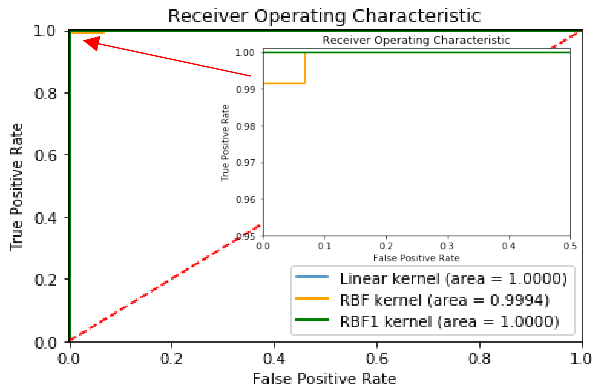


Fig. 6. The kernel performance comparison of Linear, RBF and RBF1 for testing TwoLeadECG dataset using ROC.

Fig. 5. shows that all three kernels give perfect ROC values for the ECGFiveDays data set.

Fig. 6. shows that the RBF kernel is almost 100

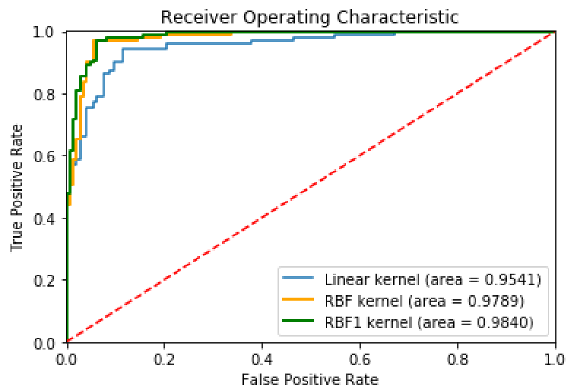


Fig. 7. The kernel performance comparison of Linear, RBF and RBF1 for testing MoteStrain dataset using ROC.

percent. Linear and RBF1 kernels give nearly perfect ROC values for the TwoLeadECG data set.

Fig. 7. shows that the RBF1 kernel to be slightly more efficient than the linear and RBF kernels for the

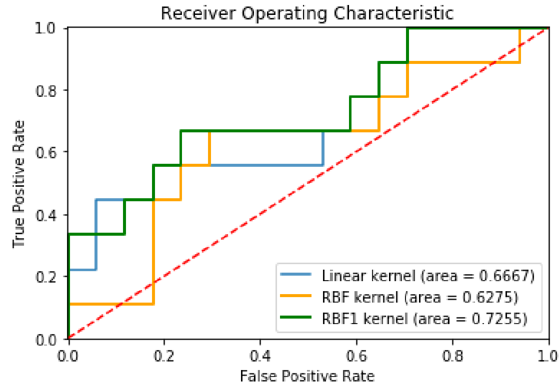


Fig. 8. The kernel performance comparison of Linear, RBF and RBF1 for testing Herring dataset using ROC.

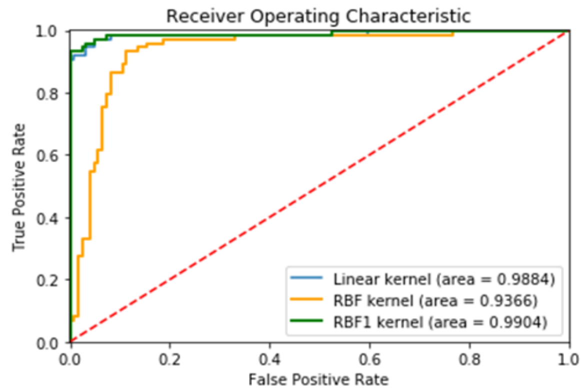


Fig. 9. The kernel performance comparison of Linear, RBF and RBF1 for testing Strawberry dataset using ROC.

MoteStrain data set.

Fig. 8. reveals that the Linear kernel gives slightly more accurate ROC values than does the RBF kernel, but the RBF1 kernel is the most accurate for the Herring data set.

Finally, Fig. 9. shows that the Linear kernel is more accurate than the RBF kernel. In particular, the RBF1 kernel is almost 100 percent for the Strawberry data set.

The anomaly detection results and comparisons are summarized in Table II. The results show that SVM with RBF1 kernel gives the highest accuracy and F1-Score on all aspects and data sets, except for the Herring data set, for which the highest of F1-Score is given by the RBF kernel. All three kernels gave perfect results for AUC, Precision, Recall and F1-Score on the ECGFiveDays data set.

This shows that the RBF kernel with parameter values C = 20, γ = 0.02 exhibits good performance in anomaly detection for time series data.

Table 2. Summary of the kernel performance comparison of Linear, RBF, and RBF1.

Datasets	Kernel : Linear				Kernel : RBF (Default)				Kernel : RBF (C = 20, γ = 0.02)			
	AUC	Precision	Recall	F1-Score	AUC	Precision	Recall	F1-Score	AUC	Precision	Recall	F1-Score
ItalyPowerDemand	0.9994	0.9818	0.9818	0.9818	0.9990	0.9910	1.0000	**0.9955**	**0.9999**	0.9910	1.0000	**0.9955**
Wafer	0.7510	0.9687	0.9946	0.9815	0.9999	1.0000	0.9969	0.9985	**1.0000**	1.0000	0.9985	**0.9992**
SonyAIBORobotSurface2	0.9819	0.9646	0.9478	0.9561	0.9974	0.9910	0.9565	0.9735	**0.9997**	1.0000	0.9826	**0.9912**
ECGFiveDays	1.0000	1.0000	1.0000	1.0000	**1.0000**	1.0000	1.0000	1.0000	**1.0000**	1.0000	1.0000	**1.0000**
TwoLeadECG	1.0000	0.9915	1.0000	0.9957	0.9994	0.9915	1.0000	0.9957	**1.0000**	1.0000	1.0000	**1.0000**
MoteStrain	0.9540	0.9552	0.8767	0.9143	0.9789	0.9783	0.9247	0.9507	**0.9840**	0.9716	0.9384	**0.9547**
Herring	0.6667	0.7647	0.7647	0.7647	0.6275	0.6538	1.0000	**0.7907**	**0.7255**	0.7692	0.5882	0.6667
Strawberry	0.9884	0.9908	0.8710	0.9270	0.9366	0.6327	1.0000	0.7750	**0.9904**	0.9911	0.8952	**0.9407**

7. Conclusion

In this paper, we presented an analysis of anomaly detection in time series data using a Support Vector Machine with three different kernels, namely, Linear, RBF and RBF1. We evaluated the accuracy of anomaly detection methods based on AUC, Precision, Recall and F1-Score criteria. The evaluation results show that the kernel with defined parameters can improve accuracy on all aspects and data sets. This application of the Support Vector Machine method, with the RBF kernel, can be efficient for detecting anomalies in time series data. The results for data set ECGFiveDays show 100 percent accuracy with all three kernels, and the results for the TwoLeadECG show almost 100 percent with all three kernels. Moreover, the results indicate a high degree of accuracy for the three kernels on all the data sets, perhaps because our data was trained in supervised conditions.

In the future, we intend to implement the variational autoencoder method for detecting and predicting anomalies in time series and spectrum data in order to compare it with the autoencoder method.

Acknowledgements

This research was partially supported by a Japanese Government Project in collaboration with the Kyushu Institute of Technology, Yaskawa Electric Corporation, Kitakyushu Foundation for the Advancement of Industry, Science and Technology, and the Hayashi Laboratory.

References

1. E. Hormozi, M. Kazem Akbari, H. Hormozi, Accuracy evaluation of a credit card fraud detection system on Hadoop MapReduce, In *The 5th conference on information and knowledge technology*, 2013, pp. 35-39.
2. Varun Chandola, Arindam Banerjee, and Vipin Kumar, Anomaly detection: A survey, *ACM computing surveys (CSUR)*. **41**(3), 2009.
3. J. Ma and S. Perkins, Time-series novelty detection using one-class support vector machines, In *Proceedings of the International Joint Conference on Neural Networks, 2003*, vol. 3, pp. 1741-1745.
4. H.-S. Kim and S.-D. Cha, Empirical evaluation of SVM-based masquerade detection using UNIX commands, *Computers & Security*, **24**(2), 2005, pp. 160-168.
5. V. Sugumaran, G. R. Sabareesh, K. I. Ramachandran, Fault diagnostics of roller bearing using kernel based neighborhood score multi-class support vector machine, *Expert Systems with Applications*, **34**(4), 2008, pp. 3090-3098.
6. U. Caydas and S. Ekici, Support vector machines models for surface roughness prediction in CNC turning of AISI 304 austenitic stainless steel, *Journal of intelligent Manufacturing*, **23**(3), 2012, pp. 639-650.
7. B. Yu, X. Song, F. Guan, Z. Yang, and B. Yao, k-Nearest Neighbor Model for Multiple-Time-Step Prediction of Short-Term Traffic Condition, *Journal of Transportation Engineering*, **142**(6), 2016.
8. P. Batta, M. Singh, Z. Li, Q. Ding, and L. Trajkovic, Evaluation of Support Vector Machine Kernels for detecting network anomalies, In *2018 IEEE International Symposium on Circuits and Systems*, 2018, pp. 1-4.
9. W. Cao, L. Li and X. LV, Kernel function characteristic analysis based on support vector machine in face recognition, In *2007 International Conference on Machine Learning and Cybernetics*, 2007, vol. 5, pp. 2869-2873.
10. H. A. Dau, E. Keogh, K. Kamgar, C.-C. M. Yeh, Y. Zhu, S. Gharghabi, C. A. Ratanamahatana, Yanping, B. Hu, N. Begum, A. Bagnall, A. Mueen, and G. Batista, "The ucr

time series classification archive," November 2019, https://www.cs.ucr.edu/~eamonn/time series data 2018/.

Relationship Between Tactile Stimuli and Human Body Sway

Masaya Tadocoro, Taro Shibanoki
Ibaraki University, Hitachi, Ibaraki, Japan

Abstract

This paper describes the relationship between tactile stimulation and human body sway. We previously proposed a body sway mitigation system based on tactile stimulation and revealed that simultaneous stimulation behind both auricles were significantly improve human balance function, however, the mechanism of it were not fully discussed. In this paper, some stimulation patterns were applied to participants and COP distribution before/after stimulation was extracted. The results showed that COP values after stimulation could be inclined to the opposite of the stimulation site. It indicates that tactile stimuli can control human balance function.

Keywords: balance function, body sway, tactile stimulation, center of pressure

1. Introduction

According to a survey by the Tokyo Fire Department, falls account for about 80 percent of accidents[1], and prevention of falls is an urgent issue in Japan, where the population is rapidly aging[2].It is said that the decline of motor function with aging increases the risk of falls by about five times[3], and various balance function improvement and fall prevention systems have been developed[4-7].

Studies focusing on improving balance function can be divided into two categories: those that use light touch contact (LTC) [4] and those that use stimulation to the human body to improve vestibular and somatosensory fuctions[5-7]. In LTC, it has been shown that the balance function can be improved by touching a fixed point. It has also been shown that it is possible to decrease and control body sway by applying acoustic[5], electrical[6] and vibratory[7] stimuli. However, because these approaches may exert a physical burden on the use, it is difficult to use in a daily life.

To overcome these problems, our research group has been proposed the body sway mitigation system based on a tactile stimulation and showed that body sway can be reduced via the application of vibratory stimuli around the pinna[8]. Here, acoustic stimulation can induce body sway to the opposite direction of the stimulus[5], however, Previous studies have not fully clarified the relationship between tactile stimulation and body sway.

In this paper, we attempt to quantify the influence of vibration stimulation on body sway by clarifying the relationship between vibratory stimulation near the auricle and body sway deflection. In the experiment, regular on-off stimulation patterns are applied to near the auricle of the subject, and the possibility of inducing center of pressure (COP) sway by the stimulation site is discussed.

2. Method

Figure 1 shows the proposed vibratory stimulation-balance function analysis method based on the body sway mitigation system using tactile stimulation[8]. The system consists of a stabilometer, tactile stimulators, a microcontroller for controlling vibratory stimulation pattern, and a PC. Stimulation patterns can be changed from the PC via TCP/IP communication. Figure 2 shows an example of stimulation pattern. The system can change the on-off interval T_s [s], stimulation interval T_{on}[s], and non-stimulation interval T_{off} [s] of each tactile stimulator via PWM control. The details of the proposed vibratory stimulation-balance function analysis are shown below.

The subjects were asked to maintain an eye-closed tandem limb stance for T [s]. Tactile stimulators were

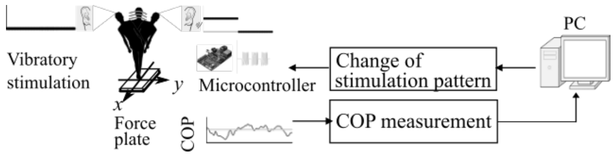


Fig. 1. Overview of the proposed balance function analysis system.

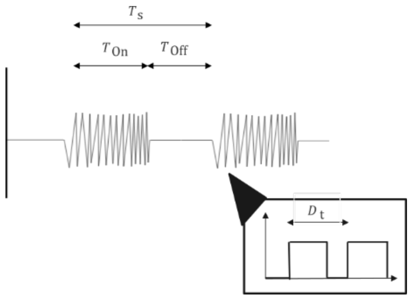


Fig. 2. An example of the stimulation pattern.

applied near the left and right auricles to provide vibration stimulation to the subjects. The stimulus intensity was the maximum stimulus intensity $D_{\max}$ during T_{on} and $D_{\min}$ during T_{off}, which was less than or equal to the minimum stimulus intensity perceived by the subject.

In the experiment, the stimulus was applied to only one of the left and right sides. The time series waveform $\text{COP}_{\{x,y\}}(t)$ of each axis of the measured center of pressure (COP) was smoothed by a second-order Butterworth digital low-pass filter (cut-off frequency: f_c [Hz]), and deviation from the median of each T [s] was calculated. In this method, filtered COP value $\overline{\text{COP}}_{\{x,y\}}(t)$ was divided into M every T_s [s], and the error between median values med_m in each interval $(m = 1, 2, \dots, M)$ and mean values ave_m^{on} for stimulation interval and $\text{ave}_m^{\text{off}}$ for non-stimulation interval were evaluated. This clarifies the bias of body sway with and without vibratory stimulation.

3. Experiments

3.1. *Experimental conditions*

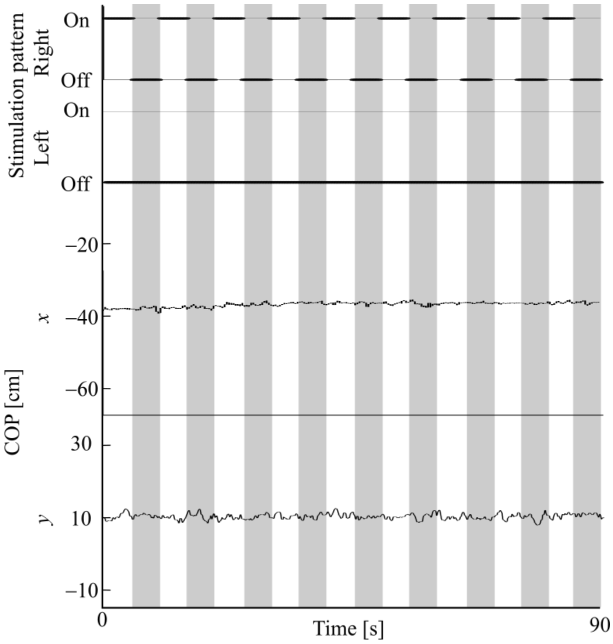
In the experiments performed, COP sway during the application of vibratory stimulation was measured with one healthy male (23 [years]). Two motor oscillators (KD18B1) were connected to a Raspberry Pi 3B and a voltage was applied by pulse width control with a duty cycle of 0.1 [s]. The subject was asked to maintain tandem limb stance with eye-closed and the left leg back for $T = 90$[s] after a preparation time for 10 [s]. The COP was measured at a sampling frequency of 100 [Hz] using a Wii Balance Board (Nintendo Co., Ltd.). The stimulus pattern applied to the subject was steady-state stimulus with $T_s = 10$ [s], stimulus intensity $D_{\max} = 100$ [%] during $T_{\text{on}} = 5$ [s], and stimulus intensity $D_{\min} = 0$ [%] during $T_{\text{off}} = 5$ [s]. This stimulus pattern was applied to left or right side.

The number of trials was twenty per side, and ten trials per side were conducted per day with an interval of one minute. The other ten trials were conducted on different days.

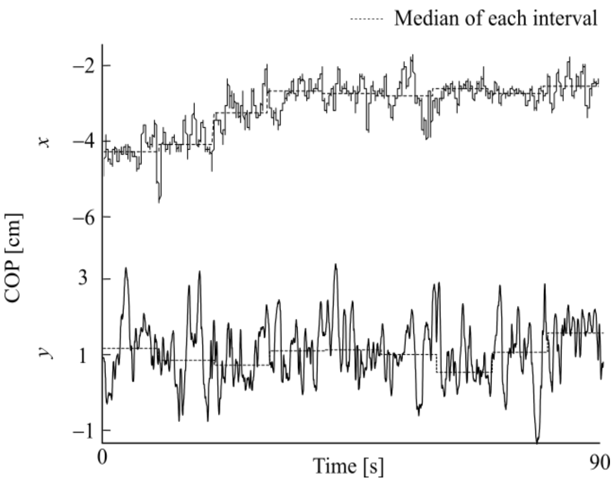
3.2. *Results and discussion*

Figure 3 (a) shows examples of signals measured during the experiments. Figure 3 (a) shows the results from the first trial, indicating stimulation patterns and COPs. The shaded areas represent times during non-stimulations. Figure 3 (b) is an enlargement of COPs from Fig. 3 (a), and it cannot be confirmed that the stimulus affected to the balance function.

Figure 4 (a) and (b) show histograms of COP of y axis (frontal plane) with median and mean values for the left-side and right-side stimulus conditions, respectively. Figure 4(a) of the results for the left-side stimulus condition shows that the median is smaller than the mean value in the stimulus interval, and conversely, the median is larger than the mean value in the non-stimulus interval. In the Fig. 4, positive value shows deviation to the left



(a) COP in each direction

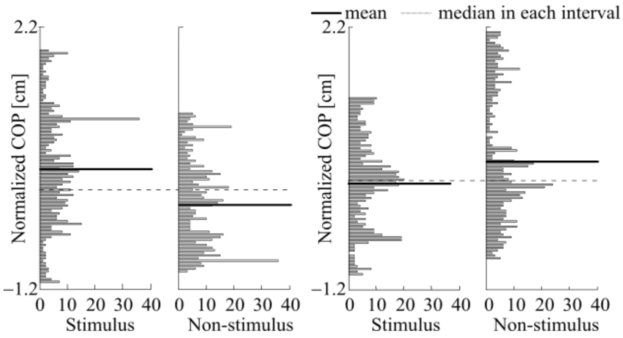


(b) An enlargement of COPs

Fig.3. Examples of experimental results.

side, therefore, the subject may tilt the opposite side of the stimulation.

Figure 5 shows the difference between mean values during stimulation and non-stimulation and the median in the interval of 10 [s] for (a) left-side stimulation condition and (b) right-side stimulation condition, respectively. The positive component represents the bias to the right side of COP distribution, and negative component represents the bias to the left side of COP



(a) Left stimulus (b) Right stimulus

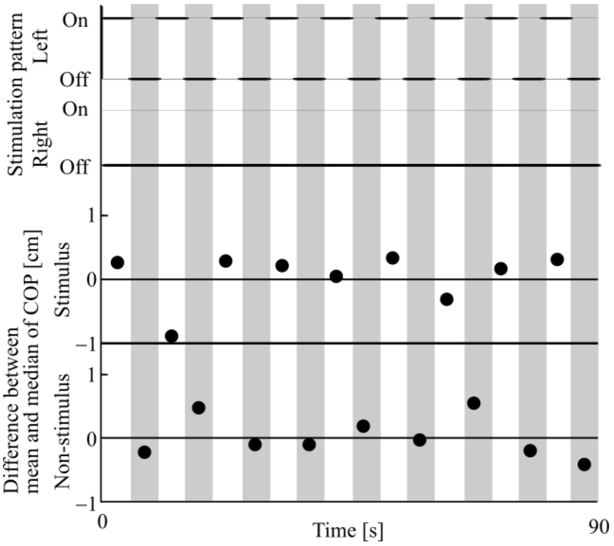
Fig. 4. Examples of COP distributions.

distribution. From the result, it was confirmed that the distribution of COP was deviated to the right during left-side stimulation and to the left during non-stimulation. The results of the heteroskedasticity t-test confirmed a significant difference at the level of 0.1 [%] (see Fig. 6 (a)). On the other hand, when right stimulation was applied, there was a bias to the left side during stimulation ($p < 0.05$), and a bias to the right side during non-stimulation ($p<0.05$) (see Fig. 6 (b). These results suggest that COP bias in the left-right direction may be induced to the opposite side of the stimulated side. It is possible that COP bias is induced to the opposite side of the stimulus, resulting in a larger bias toward the stimulus to maintain body balance equilibrium.

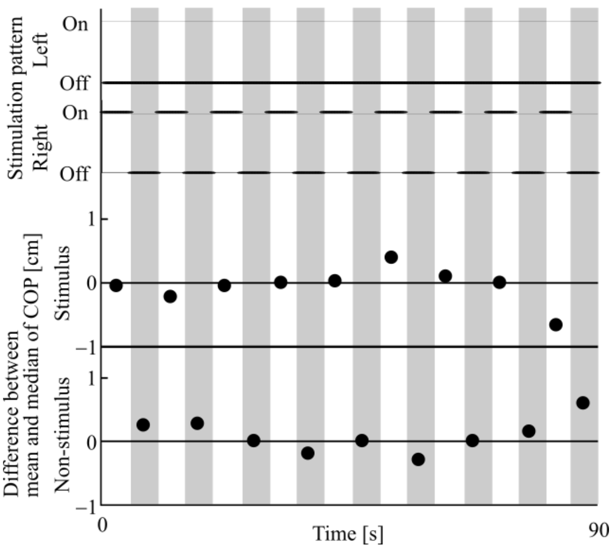
These results indicate that regular unilateral steady stimulation to the auricle induces COP sway on the opposite side of the stimulus, and that the sway may be deflected to the stimulus side to resist it. In the future, we will increase the number of subjects and investigate whether this tendency appears or not, and also whether it is possible to induce body sway arbitrarily by changing the vibration pattern or not.

4. Conclusion

In this paper, we examined the effects of regular unilateral steady-state stimulation patterns on COP bias when they were applied to near pinna. In the experiment, the subject was asked to maintain tandem limb positions with closed eyes, and the change of COP deviation with and without stimulation in each stimulation pattern was compared. It is confirmed that the stimulation may induce the deviation of COP oscillation in the frontal plane to the opposite side of the stimulated side. This



(a) Left stimulus

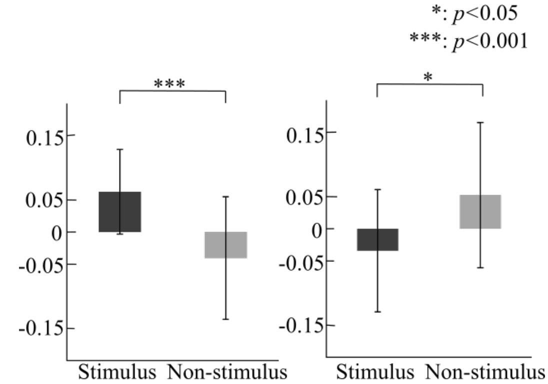


(b) Right stimulus

Fig. 5. The difference between the median and each mean value in the stimulation and non-stimulation intervals.

suggests that the amount of deflection to the stimulus side may increase to maintain the equilibrium against the induced COP oscillation.

In future research, we will increase the number of subjects and deepen the verification of the experiment conducted in this study. In addition, we plan to change the pattern of the vibratory stimulation and investigate the relationship with the body sway.



(a) Left stimulus (b) Right stimulus

Fig. 6. Comparison of mean and median differences in COPs.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 20K20212.

References

1. Tokyo Fire Department: https://www.tfd.metro.tokyo.lg.jp/lfe/topics/202009/kkhansoudeta.html
2. Statistics Bureau, Japan STATISTICAL HANDBOOK OF JAPAN 2020
3. Masdeu J.C., Sudarsky L., Wolfson L., "Gait Disorders of Aging: Falls and Therapeutic Strategies", Philadelphia, Lippincot -Raven, pp. 309–326, 1997
4. M.A. Riley, T.A. Stoffregen, M.J. Grocki, M.T. Turvey: Postural stabilization for the control of touching, Human Movement Science, Vol. 18, No. 6, pp. 795–817, 1999.
5. Miriam S. Welgampola and Brian L. Day, Craniocentric body-sway responses to 500Hz bone-conducted tones in man, J Physiol 577.1, pp 81-95, 2006
6. Chisato Fujimoto, Yoshuharu Yamamoto, Teru Kamogashira, Makoto Kinoshita, Naoya Egami, Yukari Uemura, Fumiharu Togo, Tatsuya Yamasoba and Shinichi Iwasaki, Noisy galvanic vestibular stimulation induces a sustained improvedment in body balance in elderly adults, SCIENTIFIC REPORTS, 2016
7. W. Nanhoe-Mahabier, J.H. Allum, E.P. Pasman, S. Overem, B.R. Bloem, The effects of vibrotactile biofeedback training on trunk sway in Parkinson disease patients, Parkinsonism and Related Disorders Vol. 18, pp.1017-1021, 2012
8. Satowa Watanabe, Taro Shibanoki, Koji Shimatani, *A Body Sway Mitigration Method Based on Tactile Stimulation,* Proceedings of the SICE Annual Conference 2016, pp.518-19, 2016

Analysis of the Consensus Protocol of Heterogeneous Agents with Time-Delays

Jichao Zhao [1], Fengzhi Dai [1,3] *, Yunzhong Song [2]

[1] *College of Electronic Information and Automation, Tianjin University of Science and Technology, China,*

[2] *School of Electrical Engineering and Automation, Henan Polytechnic University, 454003, China,*

[3] *Tianjin Tianke Intelligent and Manufacture Technology CO., LTD, China*

*China E-mail: * daifz@tust.edu.cn*
www.tust.edu.cn

Abstract

In practical engineering applications, the research on the consensus of heterogeneous multi-agents is of great significance. The consensus of multi-agents mainly includes average consensus, maximum consensus, and minimum consensus. The paper studies the average consensus of heterogeneous multi-agents, including continuous-time consensus protocol, discrete-time consensus protocol, consensus with time delay, and consensus of switching topology. The knowledge of graph theory is used to describe the system, and the results are simulated and analyzed for the consensus of time delay and switching topology to verify the correctness and effectiveness of the consensus protocol.

Keywords: Heterogeneous Multi-Agent; Consensus Protocol; Switching Topology; Time-Delays

1. Introduction

A multi-agent system (MAS) is a computing system consisting of multiple robots interacting in an environment. Cooperative robots can be used to perform tasks that are too difficult for a single robot to perform alone. In recent years, as a main branch of distributed system with the development of computer technology and network communication technology, MAS has been widely developed in the fields of distributed sensor network [1], network congestion control [2] and flocking movement [3]. The consensus control of MAS is a very important research topic in distributed system.

At present, the research on MAS is still in the initial stage of development on the whole, and there is still a certain distance from the real application. However, its extensive application indicates great development potential, which will definitely attract more experts and scholars to put into the research work in this field and further explore the theory and application of MAS.

MAS has the characteristics of a large number of subjects, a wide distribution of perceptions, and complex communication. There is often information interaction between subjects. Therefore, it is natural to use graph theory to describe this information interaction channel. Consensus protocol is a rule of interaction between agents, which reflects the process of information exchange between adjacent agents.

In this paper, the validity of agent consensus protocol and the relationship between consensus protocol and graph Laplace matrix are verified by analyzing the state of each node in MAS with inertial link. The relationship between algebraic connectivity (Fiedler eigenvalue) of Laplace matrix and convergence rate of MAS is verified by simulation results.

2. Preliminaries

In this paper, the graph theory is used to represent the MAS. A graph $G = (V, E, A)$ is used to represent the topology of the information exchange network between agents. $V = \{v_1, v_2, \ldots, v_n\}$ indicates that the graph has n nodes and that the system is composed of n multi-agents. $E \subseteq \{(i,j): i, j \in V\}$ is the set of edges about e_{ij}, indicating that there is information exchange between agent i and j, and the information is from i to j. $A = [a_{ij}]$ is the adjacency matrix of the graph, indicating the weight of information exchange between i and j. To simplify calculation, let $a_{ij} \in \{0,1\}$. Note that $a_{ij} = a_{ji}$ in the undirected graph, and $a_{ij} \neq a_{ji}$ in the directed graph.

The set of neighbor node $N_i = \{j | j \in V: e_{ij} \in E\}$, which represents the set composed of all agents that have information exchange with agent i. In the graph, it is represented as the set composed of nodes connected to the wired segment of node i. $deg_{out}(v_i)$, the out degree of node i is the number of edges starting from node i. $deg_{in}(v_i)$, the into degree of node i is the number of edges go into node i.

The matrix $D \in N_{n \times n}$ is the degree matrix, is defined as follows

$$\begin{cases} D_{ii} = deg_{out}(i) \\ D_{ij} = 0 \end{cases}$$

The graph laplacian matrix is defined as

$$L = D - A$$

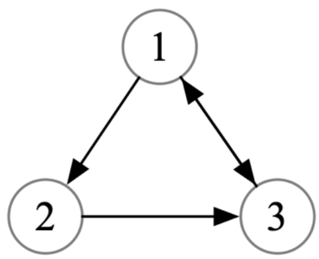


Fig.1. The main global currents

As shown in Fig.1, the system G has three nodes and four edges: $V = \{v_1, v_2, v_3\}, E = \{e_{12}, e_{23}, e_{31}, e_{13}\}$, laplacian matrix L equals

$$L = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & -1 & -1 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix}$$

3. Consensus Protocol

The individual state of the dynamic MAS is described as follows:

$$\dot{x}_i(t) = -\frac{1}{T_i} x_i(t) + \frac{K_i}{T_i} u_i(t), \quad i = 1,\ 2,\ \ldots,\ n. \quad (1)$$

where T_i, K_i denote different coefficients of node i, and $T_i \neq T_j, K_i \neq K_j, \forall i \neq j$;

System to reach consensus, means that the system all individual states are equal, mathematical representation, $x_i = x_j,\ \forall i \neq j,\ i = 1,2, \ldots, n$ or $u_i = 0,\ i = 1,2, \ldots, n$.

When there is a directed spanning tree in communication topology G, information interaction between agents can be realized to reach consensus. Therefore, it is assumed that the system topology G contains a directed spanning tree.

In particular, the consensus problem also includes average consensus, maximum consensus and minimum consensus. But it should be noted that the system consensus is equivalent to the average consistency if the graph is undirected or directed symmetric. But if the graph is asymmetric, the consensus of the system is not equivalent to the consensus of the average [4].

3.1. Consensus Protocol of Continuous Time

For continuous time systems (1), the sonsensus protocol is as follows:

$$u_i(t) = \sum_{j \in N_i} a_{ij}(x_j - x_i) \quad (2)$$

The agreement can be achieved if and only if the undirected graph is connected or the directed graph exists spanning tree. In the case of topological structure switching, the condition of consensus is that the union of switching topological graphs is undirected graph connected or directed graph exists spanning tree. Specific proof can be found in the literature [5].

Refer to the topology of Fig.1, (2) can be described as:

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} -2x_1 + x_2 + x_3 \\ -x_2 + x_3 \\ x_1 - x_3 \end{bmatrix} = -\begin{bmatrix} 2 & -1 & -1 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Combined with the graph laplacian matrix, (2) can be converted into the following form:

$$u(t) = -L * x(t) \quad (3)$$

3.2. Consensus Protocol of Discrete Time

The discrete-time consensus protocol is as follows:

$$x_i(k+1) = x_i(k) + \epsilon u_i(k) \tag{4}$$

$\epsilon > 0$ means step size, the length of time for each step.

3.3. Switching Topology Model

Since the topological network is not fixed, there is an additional switching signal k is a function of time t. The protocol is as follows:

$$u_i(t) = -L_k * x(t), \qquad k = s(t) \tag{5}$$

Fiedler eigenvalue is the second smallest eigenvalue of the Laplace matrix of the graph, which is positively correlated with the convergence rate of the graph. If the larger the Fiedler eigenvalue is, the faster the convergence rate of the system will be. On the contrary, the same relationship exists.

3.4. Time-Delays Model

The consensus protocol with time delay is:

$$u_i(t) = \sum_{j \in N_i} a_{ij}(x_j(t-\tau) - x_i(t-\tau)) \tag{6}$$

Among them, time delay τ has a threshold value of $\frac{\pi}{2\lambda_n}$.

For a network with the same time delay, assume G is undirected and connected. If and only if at the time delay τ meet $\tau < \frac{\pi}{2\lambda_n}$, $\lambda_n = \lambda_{max}(L)$, the system can asymptotically achieve consensus [4].

4. Simulink Results

Given a MAS topology graph, as shown in Fig.2, the system has a total of 6 multi-agents, and the three topologies G_a, G_b and G_c all have the same number of agents, but their topological structures are different.

Let the initial state of the system for each agent is x_i as shown in Tab.1. Notice that the average of the initial state of the system is $Ave(X_0) = 5 \neq 0$. Then, the dynamic equation of each node in the MAS is established. The dynamic equation parameters T_i and K_i are shown in Tab.1, and the parameters obey the random Gaussian distribution. The system state differential equation was established, the system differential equation was solved by software tools, and the system convergence results were drawn as shown in the Fig.3.

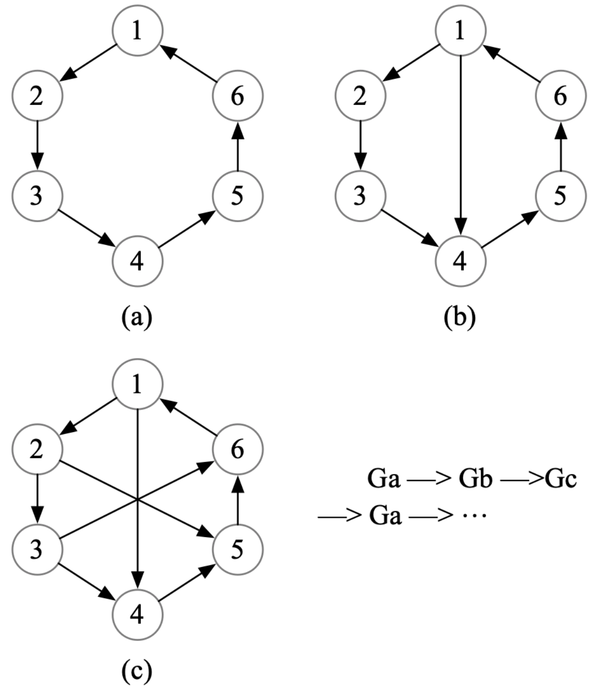


Fig.2. Topology graph of the system

In Fig.2, the order of switching topology is Ga-Gb-Gc, when $t = 0$, the system is topology A, when $t = 1$, the system is topology B, and then c, and the loop is carried out in this way.

Tab.1. MAS Parameters

i	$x_i(0)$	T_i	K_i
1	-20.50	0.81	0.91
2	30.25	0.27	0.55
3	-13.25	0.95	0.49
4	24.75	0.79	0.96
5	-7.25	0.67	0.76
6	16.00	0.70	0.03

Fig.3 shows the input-free convergence graph of the MAS, the fixed topology convergence and the switched topology convergence of the system. When there is no input, the state of the system changes with its inertial link. When input is available, Fiedler eigenvalues of topology G_a, G_b and G_c increase in turn, and the convergence velocity of the three topology graphs in the graph is observed in turn, which can verify the positive correlation between Fiedler eigenvalue and the convergence velocity of the system.

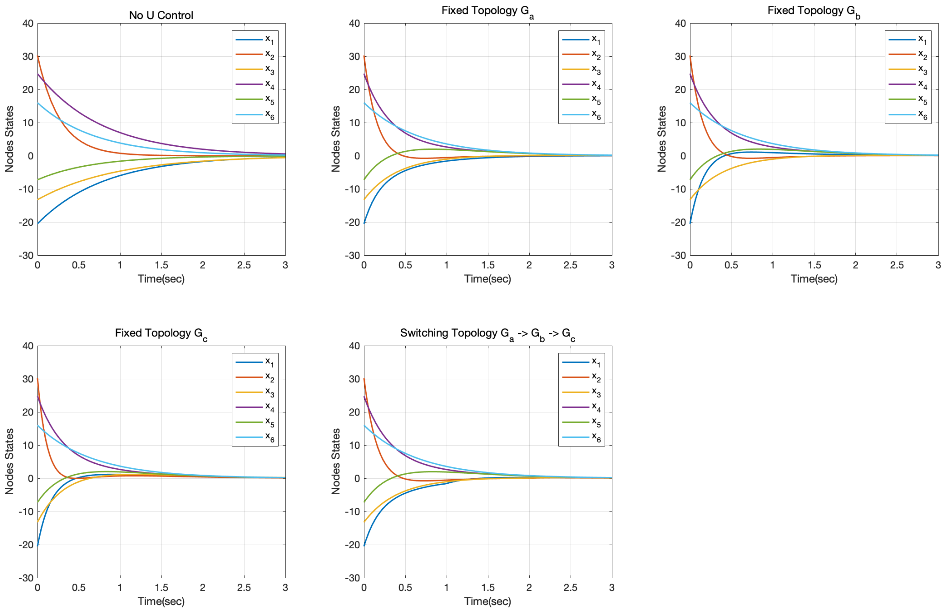


Fig.3. The result of input-free, fixed topology and switched topology

5. Conclusion

In this paper, the consensus protocols of MAS are sorted out and each protocol is simply explained. An example is given to illustrate the relationship between multi-agent system and Laplacian matrix in graph theory. The multi-agent consensus protocol is simulated, and the protocol is analyzed and verified by combining the knowledge of graph theory and system convergence graph.

References

1. M. M. I. Khan et al. Efficient and fast convergent consensus algorithms for faulty nodes tracking in distributed wireless sensor networks. 3rd International Conference on Electrical Engineering and Information Communication Technology, Dhaka, 2016:1-6.
2. Y. Wang, Z. Ma et al. Avoiding Congestion in Cluster Consensus of the Second-Order Nonlinear Multiagent Systems. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(8): 3490-3498.
3. C. Bhowmick, L. Behera, et al. Flocking control of multi-agent system with leader-follower architecture using consensus based estimated flocking center. IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, 2016:166-171.
4. Olfati-Saber R, Murray R M. Consensus problems in networks of agents with switching topology and time-delays. IEEE Transactions on Automatic Control, 2004, 49(9):1520-1533.
5. Ren W, Beard R W. Distributed consensus in multi-vehicle cooperative control. London: Springer London, 2008.

Burrows-Wheeler transform acceleration based on CUDA

Chang Sheng, Fengzhi Dai*
Tianjin University of Science and Technology, Tianjin, China
*E-mail: * daifz@tust.edu.cn*
www.tust.edu.cn

Abstract

Burrows-Wheeler transform (BWT) is a commonly used transform in compression or text comparison. For example, in bzip2, BWT is used to preprocess the original data, then the same characters in the original data are close to each other, which improves the compression rate. Because the prefix tree of the original string can be easily obtained from the result of the BWT, BWT is also applied to the search and comparison of strings. For instance, the comparison of DNA sequences uses the BWT algorithm. However, BWT is not a fast algorithm, only tens of megabytes per second on CPU. This article uses the GPU to sort the original string by the base of the 4-byte key size radix sort. After radix sort, the part with insufficient length is sorted again to complete the BWT algorithm.

Keywords: BWT, acceleration, CUDA, GPU

1. Introduction

Burrows-Wheeler transform (BWT) is a data compression algorithm proposed by Burrows and Wheeler in 1994 [1]. It can be used before other compression algorithms, such as MTF, Huffman, and RLE, so that these zero-order entropy codes can achieve the effect of high-order entropy coding. Therefore, it can replace the LZ77 sliding window search algorithm before the general compression algorithm. Moreover, thanks to the restoration method of BWT transformation, we can obtain the prefix tree of the original string through the result of BWT transformation. Through the prefix tree, the search for the string may be completed faster, and the result of the BWT transformation takes up less space than using the prefix tree. Therefore, BWT transformation is used in bzip2 compression and DNA sequence alignment.

For the implementation of the BWT algorithm, first, add a terminator to the original string, so that the starting position can be found when restoring. For example, for the string "aabcg", add the terminating character '#' to get a string of length 6 "aabcg#", and then perform a circular shift to get 6 strings, then sort, and the last column of the sorted result is the transformed result "g#aabc", as shown in Table 1.

Table 1. BWT Algorithm

Rotate left	Sort	first column	last column
aabcg#	#aabcg	#	g
abcg#a	aabcg#	a	#
bcg#aa	abcg#a	a	a
cg#aab	bcg#aa	b	a
g#aabc	cg#aab	c	b
#aabcg	g#aabc	g	c

The last column is the result. The method of restoring the string is to start with the character '#' in the first column of Table 1, get the first 'g' in the last column, then find the first 'g' in the first column, and repeat this operation to get '#gcbaa', then reverse the arrangement to get the original string.

In the process of performing the BWT algorithm, we can find that the original algorithm needs to occupy

O(m*m) space. Of course, this can be turned into O(m) space through the position index. So choosing the suitable string sorting algorithm becomes the main point of acceleration. If a complete string comparison is performed each time, memory access will be discontinuous, so using 4 bytes as a key, and indexing the corresponding value, sorting the key-value pairs can increase the continuity of memory access. For the sorting algorithm, it is realized by selecting the radix sorting that is convenient to run on the GPU. Experiments show that this method is effective.

2. Main ideas

Fig.1 is the Implementation process.

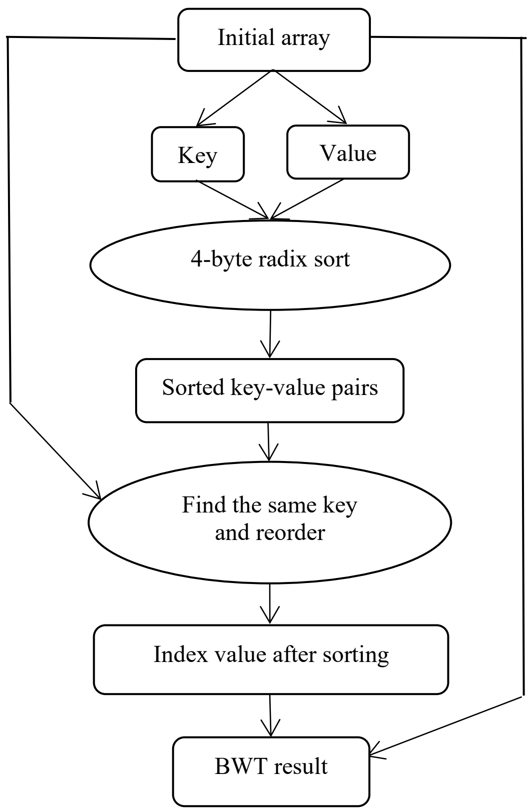


Fig.1 Implementation process

In order to complete the BWT algorithm, first, use 100M data to generate 4-byte keys and values, Then use the radix sort to sort the generated key-value pairs according to the key. For the case where the key values are the same because of the limited length, re-sort them, and finally use the value corresponding to the sorted key to generate the BWT result.

3. Implementation details

For the above process, the specific implementation can be divided into the following 4 parts.

3.1. Generate key-value pairs

This saves space and is a step towards shifting to key-value pair sorting. The key is the first 4 bytes of m*m strings, and the value is the number of rows corresponding to m*m strings. So 3 bytes from the current position, plus the current byte, a total of 4 bytes are keys, and the data from 0 to 100000000-1 is the value.

It is worth noting that the memory access must be performed in a continuous manner as shown in Fig.2, otherwise the performance will be greatly reduced. The subsequent algorithms in this article use this method when this access is optional.

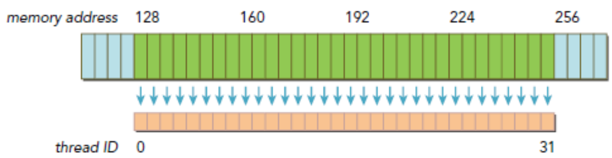


Fig.2 Memory access method

3.2. Key-value pairs radix sort

By sorting the key-value pairs, most of the data can be in the correct order. For the sorting algorithm, radix sorting is selected here. Because the radix sort has a time complexity of O(nlogn) when the radix size is determined, and occupies 2n space, the algorithm is simple and regular, suitable for implementation by GPU, and has been implemented on GPU.

3.3. Same key recognition and reorder

Because the 4-byte fixed-length mode is used, and the actual string length is 100M, it needs to be compared again when the 4-byte keys are equal. This program compares all keys. If they are not equal, the value in the key-value pair is used to index the original data, and the comparison is performed until the result is different. This algorithm sorts multiple data with smaller length at the same time.

Because the comparison of indefinite length is performed, the comparison sorting algorithm is used for the reordering part. Merge sort is considered for parallel sorting of the remaining data. However, because the merge sort has low parallelism in the final stage, and some data has a large length, the bitonic sort algorithm is used for the part with a length greater than 2000. It has high parallelism and is suitable for sorting small-length data.

3.4. Generate BWT results

After the previous step, the sorted keys have been obtained. According to the corresponding value, the sorted result of the original m*m strings can be obtained. By formula (1), the result corresponding to the last column of Table 1 can be obtained.

$$pos2 = (pos1 + L - 1)\%L \quad (1)$$

In formula (1), pos2 is the result of the last column, pos1 is the index value obtained after sorting, L is the length of the original data, % is the same as that represented in the C language for taking the remainder.

In this way, the BWT results can be obtained. Because the data is sorted, the access to the original string becomes random access, so this step consumes time.

4. Experimental results

In order to determine the difference in performance of various graphics cards, the sensitivity to data and the internal performance of the algorithm, the following experiment was carried out.

4.1. Performance on various graphics cards

In order to obtain the performance of this algorithm on different graphics cards and data volumes, the following tests were performed with random data. The result is shown in Fig.3.

It can be seen that when the amount of data is small, the concurrency of the GPU is not enough and the speed is reduced. As the amount of data increases, the speed begins to rise, and for random data, as the amount of data further rises, the speed does not decrease.

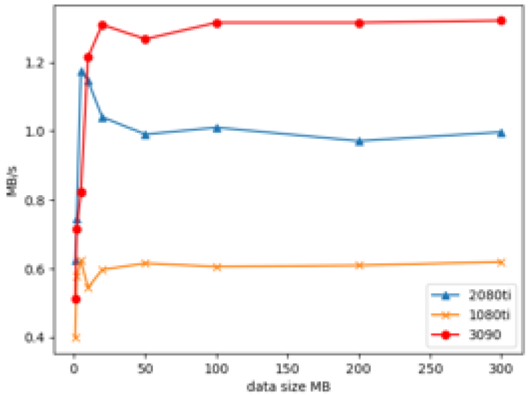


Fig.3 Performance on different graphics cards

4.2. Sensitivity to data

In order to obtain the running performance on different data, tests were carried out. The test data includes random data, win7 installation iso format files, molecular dynamics simulation trajectory trr format files, and human genome fasta format files. With 100MByte data volume, the data of Table 2 was obtained on 2080ti.

Table 2. Running performance

File type	Speed
Random data	**1.01GB/s**
Win7 iso	**0.998GB/s**
Trajectory trr	**0.549GB/s**
Genome fasta	**0.202GB/s**

For genetic data, because there are only 4 kinds of characters per byte, 16 characters are used to generate a 4-byte key, which has an impact on speed.

When there is a certain repetition in the data, this algorithm has appropriate sensitivity to the data. It will increase the amount of reordered data and cause too many character comparisons during comparison. Using Manber-Myers multiplication algorithm [2] can reduce data sensitivity, but will increase the amount of calculation for less repeated data.

4.3. Algorithm internal performance

In order to understand the time consumption of each part of the algorithm, a test was performed on 2080ti using random data with a data volume of 100MByte. The results are shown in Table 3.

Table 3. Time consumption

Stage	Time	Speed
Generate key-value pairs	1.9ms	52.6GB/s
Key-value pairs radix sort	62ms	1.61GB/s
Recognition and reorder	5.67ms	17.6GB/s
Generate BWT results	29ms	3.44GB/s

Time is mainly used to sort the key-value pairs, which is an important step in ordering the data. It also takes a certain amount of time to generate the result, because the access to the original string becomes random after the sort. In the reordering phase, because the amount of reordered data on random data is less, and the length of the reordering comparison is similar, it does not take much time.

5. Conclusion

The algorithm sorts 4-byte key-value pairs by radix sort. It can sort most of the data with less repetitiveness, and then use the reordering method to sort the remaining data. Great performance improvement for data with low repeatability. For data with high repeatability and uneven distribution, the speed is higher than the CPU. Using the multiplication algorithm can reduce data sensitivity [3], and this algorithm has more advantages for data with low repeatability.

References

1. Michael Burrows, David J. Wheeler. A block-sorting lossless data compression algorithm. Technical report, Digital SRC Research Report, 1994
2. Udi Manber, Gene Myers. Suffix Arrays: A New Method for On-Line String Searches. Siam Journal on Computing, 1993, 22(5): pp.935-948.
3. Kartsev, Petr F. High performance OpenCL realization of Burrows-Wheeler transform on GPU. International Workshop on OpenCL, Oxford, ENGLAND, May. 2018, pp. 83-84.

Design of Automatic Water Supply Upper Computer System

Peng Lu[1], Fengzhi Dai[1,2 *], Tianyi Zhang[1]

[1] Tianjin University of Science and Technology, China;

[2]Tianjin Tianke Intelligent and Manufacture Technology CO., LTD, China

*E-mail: * daifz@tust.edu.cn*

www.tust.edu.cn

Abstract

Aiming at the unstable water supply and fluctuating water pressure in the water supply system, an upper computer system of automatic water supply system based on Siemens PLC and KingView software was proposed. The pressure sensor in the water supply pipeline is used to detect the pressure in the pipeline, and the liquid level sensor monitors the liquid level in the tank. The sensor transmits the data to the PLC, and the PLC issues the control instruction after the computation processing. The KingView software can realize real-time monitoring and fault alarm of the system. The system can effectively improve the stability of water supply, and can avoid human entering the dangerous environment to search for unknown faults.

Keywords: KingView, constant pressure water supply, PLC, upper computer system, remote control

1. Introduction

With the continuous development of China's urbanization process, the pressure on the urban water supply system is increasing. In old urban communities, regional water supply is basically achieved by the municipal pipe network through secondary artificial pressurization, water towers and high-level pools.

However, these water supply methods have serious imbalances between water pressure and water consumption. In addition, the monitoring of parameters such as the pressure of the pipe network and the running state of the motor in the old-fashioned water supply system is through manual inspection of measuring instruments. This greatly reduces the operating safety of the operator. And when the water supply pressure changes greatly at different times, it can only be adjusted manually. This adjustment method is difficult to meet the system adjustment requirements in time [1].

Based on the above discussion, this design focuses on the upper computer system for constant pressure water supply. The KingView software and Siemens PLC are used to design the upper computer of the water supply system and connect it with the lower computer of the water supply system. Finally, realize the constant pressure water supply, remote monitoring and remote operation of the water supply system.

2. The Hardware Structure

The core task of this system design is to remotely monitor and operate the water supply system through the host computer. Use PLC to make the frequency converter implement cyclic control of multiple pumps [2]. So the whole hardware system is composed of a liquid level sensor, three water pumps, a pressure sensor, a frequency converter and PLC.

2.1. SIMATIC S7-300 PLC

This design uses SIMATIC S7-300 PLC. SIMATIC S7-300 PLC has good versatility. It has the advantages of easy realization of distributed configuration, modularization, low user difficulty in mastering, and no exhaust fan structure. It is very suitable for the small-scale control [3].

This design controls the frequency converter through PLC, and then controls the start and stop of the water pump through the frequency converter. SIMATIC S7-300 PLC is shown in Fig.1.

Fig.1. SIMATIC S7-300 PLC

The automatic water supply system has a medium amount of program and does not occupy much storage space. And the computing speed, communication resources and programming resources required by the system are not much. So this design chooses the CPU model of 315-2DP in SIMATIC S7-300 PLC. The 315-2DP used in this design is shown in Fig.2.This CPU has the following characteristics:

- 48 KB working memory.
- 0.3 ms/1000 instructions.
- MPI + DP connection.

Fig.2. The 315-2DP CPU

2.2. Frequency converter

This design uses the MicroMaster430 series inverters, which are particularly suitable for driving water pumps and fans. The model used is 6SE6430-2UD27-5CA0. The power of this type of inverter is 7.5~250KW. It is controlled by a microprocessor and has high functional diversity. MicroMaster430 series inverters are particularly suitable for driving water pumps and fans. Has high flexibility and reliability. The following functions can be realized: bypass function, multi-pump switching, energy-saving operation, manual/automatic switching, broken belt and water shortage detection, etc [4]. The inverter terminal interface is shown in Fig.3.

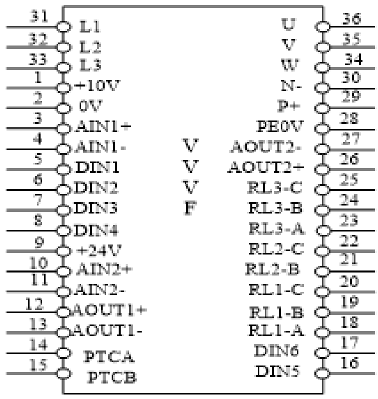


Fig.3. Inverter terminal interface

2.3. Sensor module

In this system design, MC20B economical pressure sensor and MC20C drop-in liquid level sensor are used.

The MC20B economical pressure sensor selected in this design can modify the range and zero point according to the needs on site. And it has high anti-interference and overload performance, and it is not prone to failure.

The MC20C drop-in liquid level sensor uses an isolated diffused silicon sensor to convert the hydrostatic pressure into an electrical signal. Then after temperature compensation and linear correction, it is converted into standard electrical signal output. The liquid level sensor has the following characteristics:

- Measuring range: 0.3M~200M.
- Working temperature: -20 ºC ~ 80 ºC.
- Output signal: 4mA~20mA.

3. System programming

In the system programming, this design consists of PLC programming and configuration simulation design. The combination of these two programs can well reflect the control process of the system.

3.1. PLC programming

This design uses STEP7 programming software for PLC programming. The programming method used is the logic algebra design. It is to edit the PLC ladder diagram program according to the simplified logical expression [5].

In this design, the PLC program needs to realize the following control actions. The first water pump starts and stops. When the frequency of the second frequency converter reaches the upper limit, the water pump is switched between frequency conversion and industrial frequency. Thirdly, when the pressure fluctuation of the pipeline is small, the water pump shall be alternated periodically. The fourth is to detect the liquid level, and the system can be stopped when the water supply level is too low.

PLC programming adopts modular programming. The main program OB1 is composed of seven functions or function blocks. Make the main program simple and visual for easy troubleshooting. The program in function FC1 is used to process the input analog quantity. The program in function FC2 is used for boot initialization. The program in function FC3 is the start-up program of the water pump motor. The program in function FC4 is the switch between frequency conversion and industrial frequency of the water pump motor. The program in function FC5 is an alarm program. The program in function FC6 is an analog output program. The program in function FC7 is a water pump alternate working program.

The main program OB1 part of the program design is shown in Fig.4.

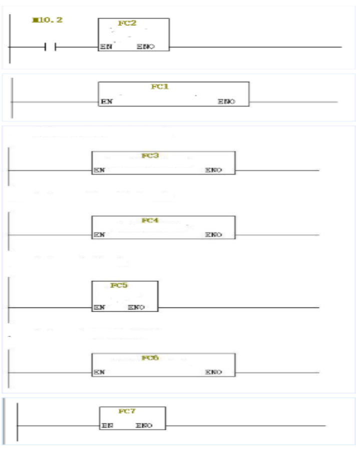


Fig.4. Main program OB1

3.2. Configuration simulation design

In this design, the design of the configuration screen of the upper computer system uses the KingView software produced by Beijing Yakong Company. The software is widely used in the current industrial field, and it has various functions such as process control design, field control, and resource management [6].

The configuration screen designed by this system covers the main screen, real-time curve, alarm screen, historical data report, historical curve, and real-time data report. In the main configuration screen, the operator can start, stop and emergency stop the system. And can adjust the pressure of the pipe network. At the same time, it can monitor the running status of the water pump, the liquid level of the pool and the pressure of the outlet pipe network. On the main page, you can also perform a pump fault test. The main configuration screen is shown in Fig.5.

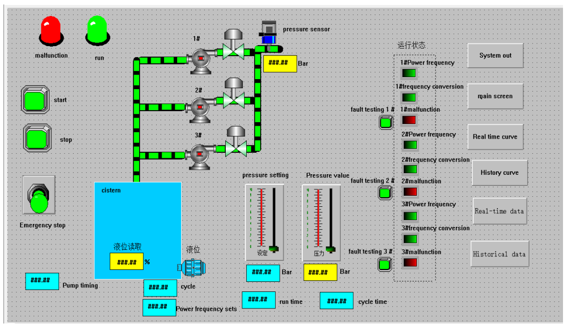


Fig.5. Main screen

4. System Simulation

After the system design is completed, the system can be simulated on the Kingview software.

In the simulation of this design, the main screen of the system can realize real-time monitoring of the operation of the automatic water supply system. In the real-time curve interface, the operator can intuitively monitor the pipe network pressure and pool water level in the system through real-time changes in the curve. It can be seen in the real-time curve simulation image that by changing the system water supply pressure setting value, the system will automatically adjust the actual water supply pressure to match the standard value. The real-time curve image is shown in Fig.6.

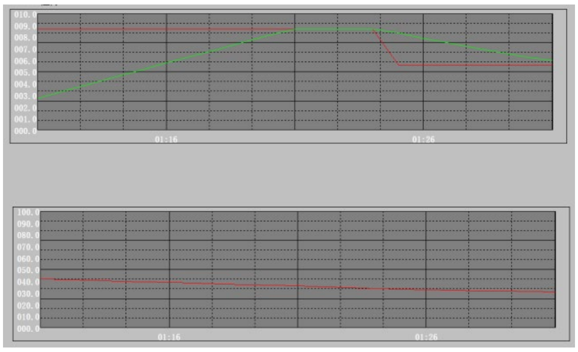


Fig.6. Real-time curve

In the historical curve interface, the operator can intuitively monitor the changes of the pipe network pressure and the water level of the pool in the historical time through the historical changes of the curve. After many simulations, the historical curve interface is shown in Fig.7.

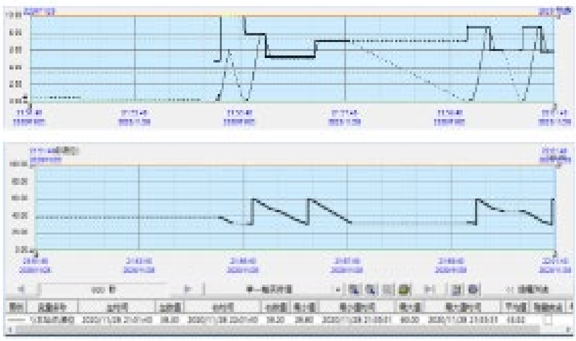


Fig.7. Historical curve

5. Conclusion

This design uses the KingView software combined with PLC and frequency converter to form an automatic water supply control system, replacing the previous valve control and manual water supply system. It is ensured that the pressure of the water supply pipe network will not drop during the time when the residents have more water. It also ensures that when the residential water is low, there will be no impact on the components of the entire system due to excessive pressure on the pipe network. It greatly slows down the aging of the water supply system components and extends the service life of the system.

The upper computer system in this design realizes the monitoring of the running status of the lower computer and the water pump. Can read the real-time data and historical data of the automatic water supply system. Monitor the changes in the operation of the water pump and the occurrence of system failures when the pressure of the water supply system fluctuates. The combination of KingView and PLC in this design makes the system program simple, integrated and highly transplantable.

References

1. Xiaoling Mo, Based on PLC the Automatic Constant Pressure Control Water Supply System Design, *Applied Mechanics and Materials*, 2014, 12: pp.1587-1590.
2. Tiewei Bu, The Overall Design of Frequency Conversion Water Supply Control System Based on PLC, *Advanced Materials Research*, 2014, 4: pp.1465-1468.
3. Yilin Qin, *Siemens S7-300 PLC application technology*, Electronic Industry Press, Beijing, China, 2007.
4. Yong Pan, Jian Wang, Quick commissioning of MICROMASTER440 inverter, *Automation application*, 2011, 1: pp.21-23.
5. Xuiling Lu, Several commonly used programmable controller ladder diagram design methods, *Machine Tool Electric Apparatus*, 2004, 10: pp.43-45.
6. Zhenfang Liu, Based on KingView configuration software and PLC water tower water level control system, *Digital Communication World*, 2019, 9: pp.85-86.

Analysis of boiler water level system Based on the fuzzy control

Tianyi Zhang[1], Fengzhi Dai[1,2 *], Peng Lu[1]

[1]*Tianjin University of Science and Technology, China;*

[2]*Tianjin Tianke Intelligent and Manufacture Technology CO., LTD, China*

*E-mail: * daifz@tust.edu.cn*

www.tust.edu.cn

Abstract

Based on the three-stroke water supply system, this paper analyzes the performance characteristics of the boiler water level control system in reality, so as to reason out the appropriate fuzzy control rules, design fuzzy controller, and applied to the control system, so that the system for self-adjustment of PID parameters, constitute a fuzzy PID control system. On this basis, this paper analyzes the performance, advantages and characteristics of two control systems: the traditional PID control system and the fuzzy PID control system, and simulates the parameters of the input variables for comparison and analysis.

Keywords: Fuzzy PID, three-pulse, MATLAB, PID control

1. Introduction

Boilers are playing a very important role in factory production. The main task of boiler production is to produce and transfer heat energy and converting energy from coal or natural gas to other energy sources such as electricity. It is a large-area capacity facility used in industrial production processes. Therefore, boiler safety is an issue of great concern to us.

Boilers are often subject to a wide variety of objects and system disturbances. It would be unrealistic to rely on the operator to keep it running accurately in the plant for long periods of time. Therefore, the study of automatic control systems for boiler equipment is a general trend [1].

In the boiler drum water level fuzzy control system, the core control method is the fuzzy control. Fuzzy control is a control method based on fuzzy logic theory, fuzzy set theory and fuzzy language variables [2].

The fuzzy PID boiler water level control system is investigated and analyzed in terms of model and characteristics [3,4]. The design, simulation, and analysis were performed in the Fuzzy Logic Toolbox in MATLAB.

The fuzzy control is not necessarily an accurate mathematical analysis of the controlled object, but rather a combination of fuzzy control algorithms and PID control to control non-linearity, make the boiler level system stable and robust.

2. System Control Principle

Generally speaking, the larger the boiler drum capacity, the larger the volume of steam mixture and steam in the drum. Sufficient space is required for separation and circulation, and the moisture content of the generated steam is much more stringent. In turn, the water level of the cartridge is also very strict.

In a double-pulse control system, automatic adjustment of the water supply valve does not occur in time, resulting in a delay. Therefore, the water level in the packet will deviate due to the delay. The tri-pulse control system introduces the control amount of the water supply signal based on the level and steam signals of the double-pulse control system.

2.1. boiler water level system structure

Boiler drum water level control system consists of Economizer, Drum Boiler, Down tube, Superheater, Lifter tube, Water Supply and so on.

The boiler obtains water through a water supply control valve, then passes through an economizer where it absorbs heat from the burning fuel, regulates it to the amount of water supply, dynamically equilibrates the steam mixture and liquid water in the steam package by regulating the water supply, then passes through the water cooling wall around the furnace in the top connection box, and finally returns to the steam package[3]. Boiler drum water level control system is shown in the Fig.1.

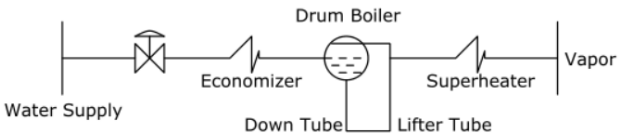


Fig.1. Boiler drum water level control system

2.2. principles of PID control

Start with the classical PID control, which consists of proportional, integral, and differential components. The proportional part of this is to maintain system stability . As the scale factor increases, the adjustment time becomes shorter and the resulting system deviation becomes smaller. However, if the proportionality factor is too large, oscillations can occur. Therefore, the scale factor K should be appropriate to achieve the effect of small transition time, small static difference, and high stability.

The function of the integral part is to eliminate systematic deviations. The larger the integration time, the weaker the system's ability to eliminate deviations and the longer the transition time. The smaller the integration time, the greater the ability to eliminate deviations that may cause the system to oscillate. The role of the microcomponent is to improve the dynamic performance of the system. Classical PID Control Schematic Block Diagram is shown in Fig.2.

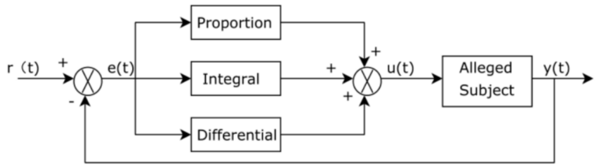


Fig.2. Classical PID Control Schematic Block Diagram

The control rule is

$$U(t)=K_P\left[e(t)+\frac{1}{T_i}\int_0^t e(t)\,dt+T_D\frac{d_{e(t)}}{d_t}\right] \quad (1)$$

When selecting a PID controller, choose a simpler and more accurate incremental PID controller.

$$\Delta U(k)=u(k)-u(k-1) \quad (2)$$

Only the first three deviations are needed to calculate Δ U(k), and the incremental PID control algorithm is now widely used.

2.3. cascade control systems

Cascade systems form an inner and outer closed loop in the system configuration. The execution of the control sequence of the system is the outer ring first, followed by the inner ring. Cascade Control Systems is shown in Fig.3.

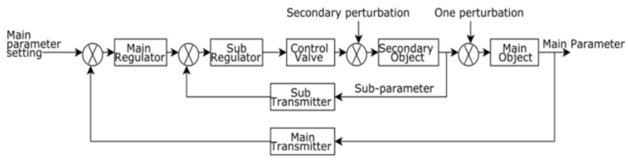


Fig.3. Cascade Control Systems

3. Design of Fuzzy Controller

The following introduces the basic structure of the fuzzy controller, and the fuzzy controller designed in this paper.

3.1. the basic structure

It consists of four parts: Knowledge Base, Fuzzy Reasoning, Fuzzification of Inputs, and Accuracy of Outputs.

First, the input quantity is fuzzified and reasoned about by the knowledge base, and finally the output quantity is decoupled and refined. Basic structure of the fuzzy controller is shown in Fig.4.

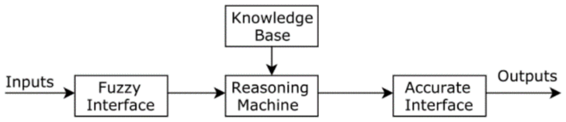


Fig.4. Basic structure of the fuzzy controller

3.2. design process for fuzzy controller

There are the following steps.

- Structural design with two-dimensional fuzzy controller for maximum efficiency. Controller has two inputs and three outputs.
- Fuzzing of Inputs. The set of input and output words is denoted by {NB,NM,NS,ZO,PS,PM,PB} and the subordinate function is chosen to be an isosceles triangle. Isosceles triangle affiliation function image is shown in Fig.5.

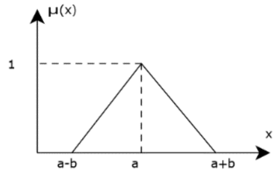


Fig.5. Isosceles Triangle Affiliation Function Image

- Define fuzzy rules. Make separate tables of the three output variables and analyze how they change on their own as E and Ec change. The 49 control rules are deduced from practical experience. The Kp is shown in the Table 1.

Table 1. Kp Fuzzy Rules Table

e\ec	NB	NM	NS	ZO	PS	PM	PB
NB	PB	PB	PM	PM	PS	PS	0
NM	PB	PB	PM	PM	PS	0	0
NS	PM	PM	PM	PS	0	NS	NM
ZO	PM	PS	PS	0	NS	NM	NM
PS	PS	PS	0	NS	NS	NM	NM
PM	0	0	NS	NM	NM	NM	NB
PB	0	NS	NS	NM	NM	NB	NB

The rule table reasoning for Ki and Kd is the same as above.

- Clarifying Variables. In order to obtain accurate output variables of the control system, the weighted average method (center of gravity method) is used to solve the fuzzy.

4. Control System Modeling and Simulation

Design of automatic adjustment of steam packet level according to boiler water level dynamics.

4.1. classical PID control modeling

From empirical data, under unit step perturbation of feedwater flow rate, floating velocity of boiler water level V=0.05 mm/s.

The time constant T2 = 15s, so the transfer function is determined as G(W) = 0.05/(15S^2+S). Under steam flow perturbation, the amplification factor K2 = 5, T2 = 15s, yielding the transmission G(D) = -0.05/S+S/(15S+1).

The boiler water level model is built using the SIMULINK toolbox in the MATLAB environment, and the three parameters Kp=10, Ki=0.0045, and Kd=0 are obtained by the improvised test method [4]. Classical PID control model is shown in Fig.6.

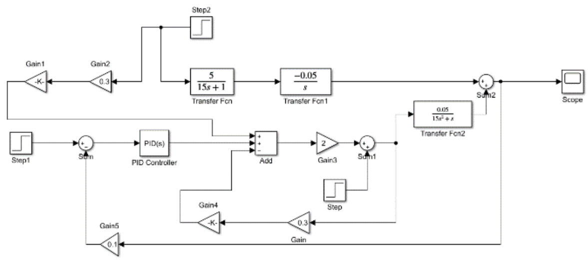


Fig.6. Classical PID control model

4.2. fuzzy PID control modeling

Add a fuzzy logical controller with associated fuzzy rules to the classical PID control simulation model. Fuzzy PID control model is shown in Fig.7.

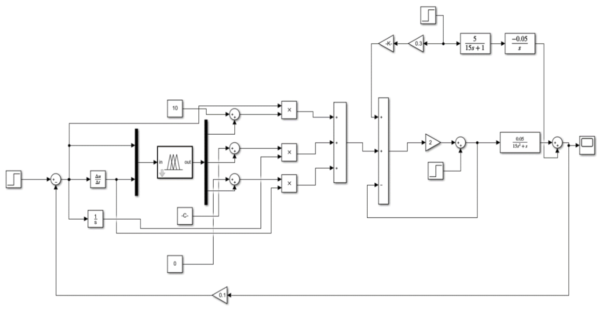


Fig.7. Fuzzy PID control model

4.3. simulation results analysis

- Simulation graphics with only level perturbation

Simulation diagrams with only level disturbance is shown in Fig.8. The left figure shows the simulation results of the classical PID control model and the right figure shows the simulation results of the fuzzy PID control model.

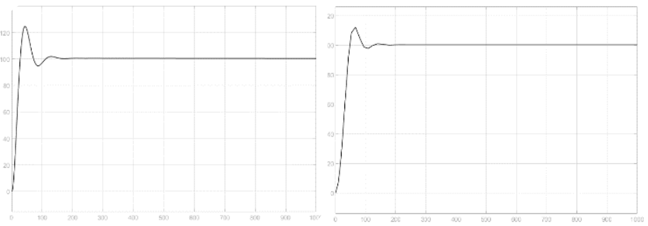


Fig.8. Simulation diagrams with only level disturbance

- Add steam disturbance at 500s in addition to level disturbance. Simulation diagrams for liquid level disturbance and vapor disturbance is shown in Fig.9.

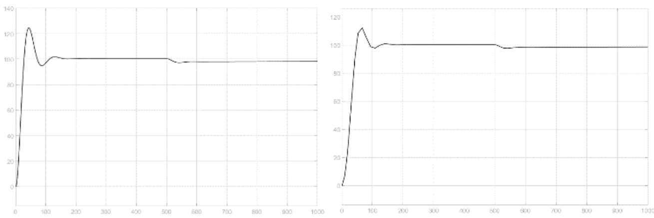


Fig.9. Simulation diagrams for liquid level disturbance and vapor disturbance

From the two graphs, the fuzzy PID control system (right) has better overshoot and setting time than the classical PID control system (left). When multiple perturbations are added, the fuzzy PID control system is more stable.

5.Conclusion

By analyzing the dynamic and static characteristics of the automatic drum water level control system, the paper constructed a conventional PID simulation model and a fuzzy PID simulation model to validate the results. It can be concluded that the fuzzy PID control system has a small amount of overshoot, short oscillation period and short setting time.

Fuzzy PID control system has greater flexibility and robustness, and better fault tolerance.

References

1. Zhongning Li. Design and analysis of boiler water level fuzzy control system based on MATLAB, 2008, pp. 5-7.
2. Jian Yang. Water level modeling and control strategy for large marine fuel boilers. Shanghai Jiao Tong University, 2007, pp.42-43.
3. Ping Chen. Study of fuzzy PID control of boiler drum water level, 2006, pp. 5-10.
4. Na Feng, Yao Wang. MatLab simulation study of boiler drum water level fuzzy control system. Automation and Instrumentation, 2019, 34(04): 80-84+99.

Development of the Circuit System for Greenhouse Environment Regulation

Yuhui Cheng, Fengzhi Dai *, Chengxu Ji, Peng Lu

Tianjin University of Science and Technology, China;

*E-mail: * daifz@tust.edu.cn*

www.tust.edu.cn

Abstract

Aiming at the demand of agricultural modernization, this paper proposes a circuit system for greenhouse environment regulation. For the greenhouse, due to the requirements of the crops for the stability of the growth environment, the cheap single-chip control system can be used to complete this work well. The system can obtain the current environmental data in the greenhouse through the DHT11 sensor and the light sensor, regulate and control the environment through external equipment to keep the environment stable.

Keywords: STC89C52 microcontroller, DHT11 temperature and humidity sensor, light sensor

1. Introduction

With the development of automation technology and the gradual reduction of the number of people engaged in agricultural work, the method of agricultural production has begun to turn to automation and unmanned. Many new terms are beginning to be known by people, such as automatic harvesting, automatic cultivation and so on.

On greenhouse, the temperature, humidity and light intensity control three environmental factors, for crops grown in greenhouses of great importance. Closely related to the development and environmental changes in greenhouse crops, from sowing a seed to germinate and then pull back strain results, optimal plant growth environment in different growth period the demand is not the same. Most direct impact is the temperature, humidity, and in light of these factors [1].

Based on the above discussion, this circuit system mainly focuses on the collection and automatic adjustment of relevant data in the greenhouse. The control and calculation module of this system is a single chip microcomputer. The system to achieve data collection work by DHT11 temperature and humidity sensor and an illumination sensor can be real-time monitoring of temperature, humidity, light intensity in the greenhouse. The system will be on the sensor data collected in real time comparison, if the preset threshold is reached, then the corresponding module will start work in order to achieve the purpose of the environment within the greenhouse to be regulated.

2. The Hardware Structure

This system is mainly responsible for the collection of environmental data in the greenhouse and the control of the corresponding work modules. The system is mainly divided into three parts. The first part is the calculation and control part composed of single-chip microcomputers, the second part is the sensing part composed of sensors, and the third part is the external equipment controlled by the relay.

2.1. DHT11 temperature and humidity sensor

DHT11 digital temperature and humidity sensor is a temperature and humidity composite sensor with calibrated digital signal output. It uses dedicated digital module acquisition technology and temperature and humidity sensing technology to ensure that the product has extremely high reliability and excellent long-term stability [2]. The DHT11 module is shown in Fig.1.

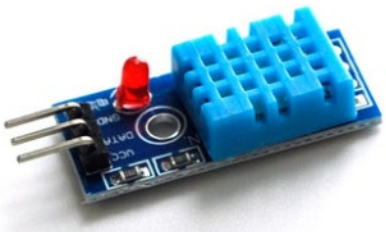


Fig.1. DHT11 module

2.2. BH1750FVI module

The BH1750FVI module (Fig.2) is a digital light intensity sensor for a two-wire serial bus interface, with a built-in 16-bit ADC. The module has the following characteristics: Wide brightness range: 0LX ~ 65535LX; Operating Voltage 3.0V ~ 5.0V.

Fig.2. BH1750FVI module

2.3. LCD1602 liquid crystal display

LCD1602 liquid crystal display is a dot matrix liquid crystal module used to display letters, numbers, symbols and other characters. The liquid crystal display consists of multiple 5X7 or 5X11 dot matrix character bits to form a dot matrix character display module, which is shown in Fig.3.

Fig.3. STM32F103ZET6 chip

2.4. Main control chip

In this design, the external MCU is the STC89C52RC.The chip has 8KB flash memory and 512B RAM. And it has more than 30 general-purpose IO ports, multiple timers, which fully meet the design of the required pin and memory requirements. The main control chip is shown in Fig.4, and it has the following characteristics:

- Operating Voltage 3.3V ~ 5.5V.
- Operating temperature range: 0 ºC ~ 75ºC.

Fig.4. STC89C52RC chip

3. System circuit module design

In the circuit design, this system uses temperature and humidity data acquisition module, light data acquisition module and relay control module. These modules play a vital role in this system.

3.1. Design of connection circuit for temperature and humidity module

The IO pin of the DHT11 module for data exchange is connected with a 4.7K pull-up resistor in parallel with VCC pin to increase the pin level to transmit data. In addition, a capacitor is connected to the power supply for decoupling and filtering to improve voltage stability. Because the module's single-bus design greatly reduces the occupancy rate of circuit resources in actual use, the overall circuit looks very simple and clear.

3.2. Design of connection circuit for Illuminance module

The VCC pin and GND pin of the module can be connected to the positive and negative poles of the microcontroller. ADD is the address pin to zero. SCL and

SDA are the clock and data lines in the I2C bus respectively. SCL is connected to the P10 pin of the microcontroller. The SDA pin is connected to the P11 pin of the microcontroller.

3.3. *Relay module control circuit*

Relay is a circuit control element, which is triggered at high level. The input terminal is generally provided with three pins, GND, VCC, and IN. What we use in this system is two dynamic close type relay modules, controlled by the single-chip P7 pin and P8 pin.

4. Introduction of functional module

Through the work of each functional module, the greenhouse environment control circuit system can well complete the set work.

4.1. *Data collection module*

The data acquisition module of this system is composed of DHT11 temperature and humidity sensor and light sensor. When the system is powered on, each sensor will read the environmental data in the greenhouse in real time and send it to the processing unit.

4.2. *Airkiss distribution network*

The system can adjust the threshold of environmental parameters by pressing the buttons. The functions of each button are shown in Table 1.

Table 1. Introduction to key functions

Button	Function
Key1	reset
Key2	selection
Key3	value up
Key4	value down

The adjustable parameters of the system are shown in Table2.

Table 2. System adjustable parameters

Name	Nun1	Num2
Temperature	upper limit	lower limit
Humidity	upper limit	lower limit
Light	upper limit	lower limit

4.3. *Environmental Control Module*

The control circuit system of the greenhouse environment regulation system controls the external equipment through the relay.

When the environmental data collected by the system is not in the preset range, it will control the relay to turn on the corresponding external equipment such as water pump, exhaust fan, fill light, etc. according to the type of value. This design method can replace different types of external equipment according to different environmental requirements, which has better practicability.

5. Conclusion

This design integrates the environmental information in the greenhouse and manages it uniformly.

From the perspective of operation, it can save a lot of human resources and avoid some mistakes caused by manual operation. From the perspective of production, the system regulates various factors affecting the growth of crops in the greenhouses to keep them in an appropriate range, and the yield and crop quality of greenhouses have been improved to a certain extent. From the point of view of cost, the hardware cost of the system is very low, which is conducive to large-scale laying.

References

1. Yin Liu, Huijuan Lv, *The influence of external conditions on vegetable growth and its control methods*, *Modern Agriculture Research*, 2011, 2: pp.10-11.
2. Xinxiu Xu, ShiYuan Zhao, *The influence of external conditions on vegetable growth and its control methods*, *Modern Information Technology*, 2020, 14: pp.57-59.

Design of An Intelligent Car Based on MSP430

Ruming Kang, Fengzhi Dai *

Tianjin University of Science and Technology, Tianjin, China

E-mail: * *daifz@tust.edu.cn*

www.tust.edu.cn

Abstract

The system takes the MSP430 single chip microcomputer as the control core, uses the reflective photoelectric sensor TCRT5000 module to track the line, and realizes the automatic tracking. In the experiment, the black-and-white line is used as the route. The system is driven by the L298N module and controlled by the PWM DC motor. This paper introduces the principle of the reflective photoelectric sensor and the circuit diagram of the tracking module. It also explains how to realize the automatic tracking based on the MSP430 single chip microcomputer. The technology can be used in warehouse, unmanned production line, intelligent service robot and other fields.

Keywords: MSP430, Sensor, Tracking module, Motor driven

1. Introduction

This smart car material comes from the 2020 National University Student Electronic Design Competition. It used TI's MSP430 microcontroller as the control center of the four-wheeled car, and transmits the signals of each sensor to the microcontroller for analysis and processing, thereby controlling the L298N motor drive, controlling the car and buzzer the device tweets.

The car is required to be able to follow the designated route and automatically follow the line on the ramp. The car must run independently, and no equipment (including power supply) can be used outside the car. The weight of the trolley (including battery) is less than 1.5kg, and the outline dimension on the ground is not more than 25cm× 25cm. The ramp is made of blockboard with a length and width of about 1m, allowing the natural color and natural wood grain on the board.

The surface of the wood board is laid with 1cm×1cm black and white paper strips (hereinafter referred to as the marking line) as route instructions; the starting section of the marking line is a straight line, parallel to both sides of the board; the marking line turns 90° at the top of the slope, and the turning radius is 20cm; The distance between the marking line parallel to the top of the slope is ⩾30cm, and the distance from the top of the slope is ⩽20cm; the total length of the marking line is 1m. The parking mark is a black line with a width of 1cm and a length of 5cm, perpendicular to the marking line on the top of the slope. The slope angle of the trolley and the top view of the driving route are shown in Fig.1.

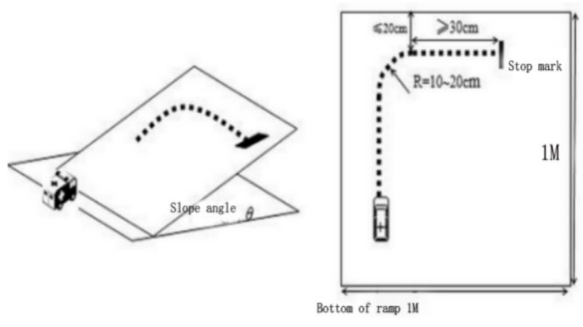


Fig.1 Schematic diagram of car ramp angle and top view of driving route

2. Modularization of the car hardware

The smart car can be divided into: power module, sensor detection module, microcontroller module and DC motor drive module. The smart car can realize functions such as tracking and parking. It uses the MSP430 single-chip microcomputer as the main control chip, and uses a DC motor as the power. The hardware system structure diagram is shown as in Fig.2.

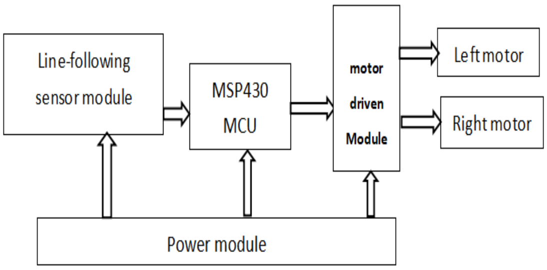


Fig.2 Hardware system structure diagram

2.1 Motor drive module

The motor adopts a PWM DC motor, which can be used as an ordinary motor (no coding function is used to make it have better mechanical speed regulation performance, and the DC-DC step-down voltage stabilized power supply module is used to realize the mechanical speed regulation of the motor), and it is driven by L298N.

The single-chip microcomputer controls the level of the L298N enable terminal to realize the rotation of the motor and realize the forward or steering of the trolley. The principle of the direction change of the DC motor is: if turning left, the left wheel will not move, and the right wheel will turn; if turning right, the opposite; if going straight, both left and right wheels will turn. But the problem that 298 chips are prone to heat is difficult to solve. The physical picture of L298N is shown in Fig.3.

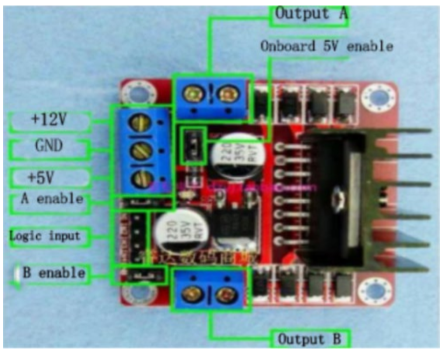


Fig.3 Physical Picture of L298N Module

2.2 Tracking module

Component: CRT5000 photoelectric sensor

Principle: The infrared emitting diode of the TCRT5000 sensor continuously emits infrared rays. When the emitted infrared rays are not reflected back or reflected back but the intensity is not strong enough, the infrared receiving tube is always in the off state, and the output terminal of the module is high at this time. Indicating that the diode is always off; when the detected object appears in the detection range, the infrared is reflected back and the intensity is large enough, the infrared receiving tube is saturated, and the output of the module is low at this time, indicating that the diode is lit.

The application in the actual program is: two TCRT5000 sensors are installed on the left and right sides of the car. (1) When the left and right sensors are in the white area, that is, the emitted infrared rays are emitted back, both the left and right motors rotate. (2) When only the left sensor detects the black line, the infrared is not reflected back, indicating that the car is driving to the right. At this time, the motor on the right rotates and the motor on the left stops, and the car shows a left turn correction. (3) When the right sensor detects a black line, the emitted infrared is not reflected back, indicating that the car is to the left. The motor on the left rotates and the motor on the right stops, and the car turns right to correct.

The schematic diagram of the tracking process and the circuit diagram of the tracking sensor module are shown in Fig.4 and Fig.5.

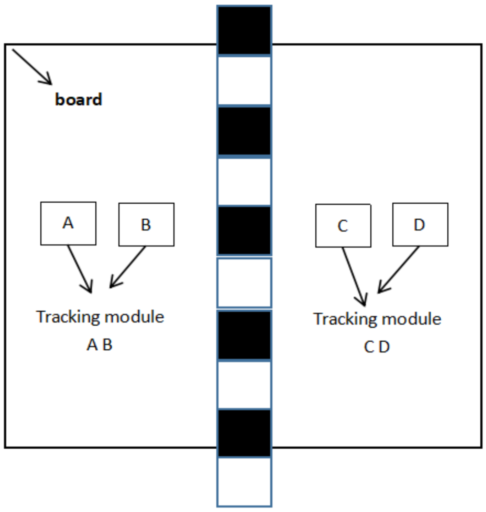


Fig4 Schematic diagram of tracking process

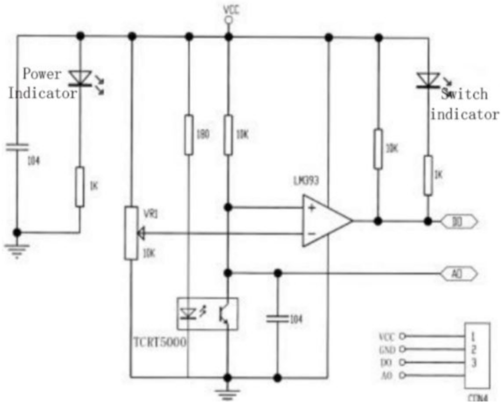


Fig.5 Schematic diagram of car ramp angle and top view of driving route

2.3 Power Module

The power module is divided into two parts: one is the chip power module, which uses a 4.2V lithium battery to drive the chip; the other is the motor drive module, which provides higher power to drive the motor through L298N.

Motor drive power supply: Use two 4.2V lithium batteries connected in series to form about 8.4V to provide power to the L298N chip. Generally, the battery has high power, is more stable and durable, can be charged more economically, and has better stability, and is more convenient to use.

Chip drive power supply: Since the power requirement is not large, but the stability is high, so use the 8.4 voltage regulator circuit to drive (the voltage drop has been considered, of course, the voltage drop of each electronic device is different, and you need to debug according to your own.

3. System software design

The smart car is based on the modular design concept, and the software design corresponds to the hardware system, mainly for tracking subroutine modules and parking subroutine modules. The main program calls these subroutine modules in logical order.

The single-chip microcomputer collects the signals collected by the sensors of each module, after processing, and then controls the actions of the trolley through the motor driver to complete the predetermined target. When the single-side sensor detects the black line, it judges the status according to the tracking subroutine and makes steering adjustment. When both the line-following sensors detect the black line, it judges the turning state according to the parking subroutine and makes a parking action.

If it reflects that the obstacle avoidance and parking actions are inaccurate or there are loopholes, you can appropriately add line-following sensors on both sides, and make corresponding modifications in the subroutine to improve the accuracy of obstacle avoidance and parking. The obstacle avoidance flowchart is shown in the Fig.6.

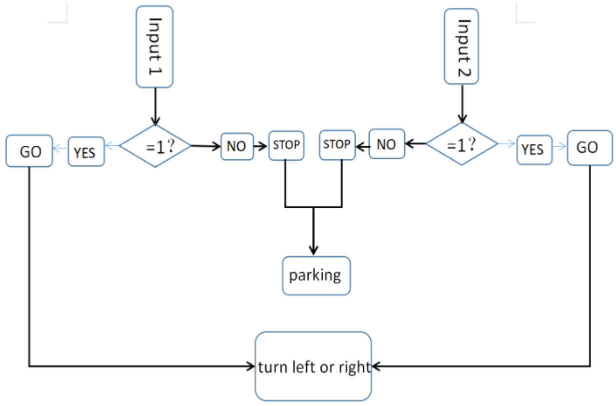


Fig.6 System software design flow chart

4. Conclusion

This article explains the intelligent car control system based on MSP430 single chip microcomputer chip. Through the infrared tracking sensor, the car tracking and parking functions are realized, and the expected design effect is achieved. In the design process, light has an occasional impact on the tracking. LED lights can be installed adjacent to the line-following sensor to enhance the light and strengthen the recognition of black lines. At the same time, the tracking algorithm should be optimized to make the tracking more accurate.

References

1. Ying Ceng, He Ming Zhao, Zhi Ling Peng, Design of Multifunctional Intelligent Car Based on MSP430, Electronic Measurement Technology, 2017 (6): pp. 10-14.
2. Meng ting Chen, Bai yan Hu, Can Huang, Design of Intelligent Tracking and Obstacle Avoidance Car Based on Single Chip Computer, Smart Robot, 2016(4): pp.47.
3. Zheng Bing Wan, Design of automatic tracking car based on LDC1000, Automation technology, 2016(10): pp. 111-157.
4. Qiu Yan Zhang, Jun Xia Zhang,Yang Yang, Design of Multifunctional Intelligent Car Based on MSP430, Information Technology, 2018(11): pp.54-58.

Design of a WIFI Video Car

Qianqian Zhang[1], Fengzhi Dai [1, 2] *, Jichao Zhao [1], Haokang Wen [1], Hongbo Hao [1]

[1] Tianjin University of Science and Technology, China;

[2] Tianjin Tianke Intelligent and Manufacture Technology CO., LTD, China

E-mail: daifz@tust.edu.cn

www.tust.edu.cn

Abstract

This subject is mainly composed of motor, frame, STC89C51 single-chip microcomputer and other auxiliary modules. Through the android client broadcast video car cameras, wireless video signal acquisition by the client send commands to the car at the same time, wireless router transmission instruction by wireless router, single-chip microcomputer is run for processing, then the motors are driven by single chip microcomputer control to rotation and steering, so as to realize the control of motor sports and video acquisition.

Keywords: The router OpenWRT, The Socket, WIFI video car, STC89C51, The steering gear

1. Introduction

WIFI video car can achieve real-time monitoring, through the computer, mobile phone, tablet etc to connect wireless router, and through the client sends commands to the WIFI router serial port on a router by transmitted to MCU. MCU program parses the command, then operating LN298 driver module so as to achieve the effect of motor control. It has 128 RAM and 4K of ROM inside.

2. Overall scheme design

In the design and production of circuits, we try to reduce other interference introduced in the measurement to achieve a higher level of performance indicators and functions of electronic products. Usually the circuit design is to take into full account the various factors and the circuit physical.

The circuit mainly involves the connection of router, MCU minimum system, LN298 drive module, steering gear, and camera. Make clear the production task of the system, make specific analysis of the production task, understand the performance of the system, indicators, content and requirements. The important task of the scheme selection is to complete the function of the system according to the knowledge and information, according to the tasks and requirements [1].

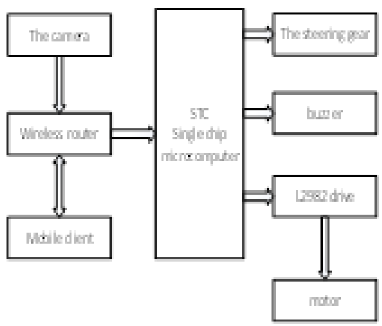


Fig.1. General layout structure

3. Subsystem design

3.1Circuit design steps

Economical and detailed circuit design is half the success of our project. First of all, we need to design a minimum system circuit. We need to know which components are used in a minimum system circuit. In addition, we need to accurately determine the parameters of various components, such as clock frequency, resistance value, capacitance parameters and so on [2].

Draw the circuit diagram of the minimum system and simulate it. After the simulation is successful, the circuit welding is carried out. The circuit diagrams are shown in Fig.2 to Fig.4.

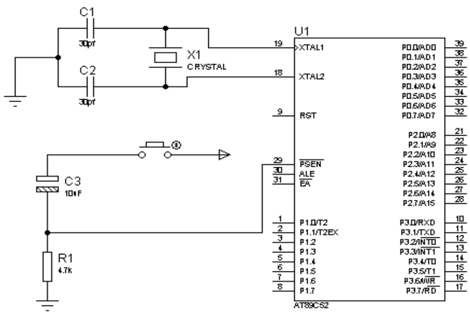


Fig.2. Minimum system simulation circuit

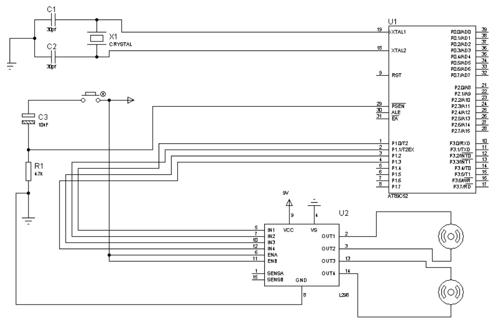


Fig.3. L298N drive module connection circuit

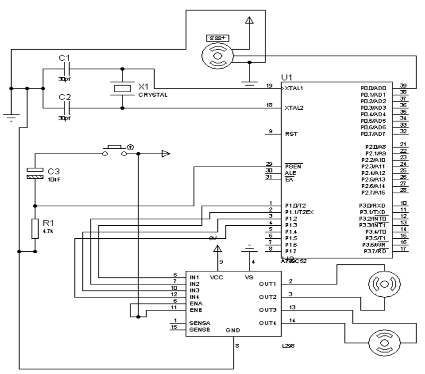


Fig.4. Steering gear connection circuit

Actually, the whole hardware system is not complicated, the main route as the core, openWRT as the main control system. For the next bit machine, 51 of the procedures will be responsible for driving motor and steering gear, we mainly use the two timers, the first is for steering gear, adjust the duty ratio of the steering gear, so that we can through the command code to control the rotation Angle of steering gear [3].

The second is to use a serial port interrupt, timer is used to set the baud rate (refers to the speed of data transmission) to 9600, and the timer 2 should set automatic loading mode (refers to give it an initial value). The timer need to be set up only once, after the timer overflow, automatically return to the initial value.

The 298 drive is based on the high and low level connected to the controller pin to determine the positive motor inversion, of course, the drive board is to be in common with the minimum system, which is used to run the program to be effective control driver board, so as to control the motor rotation and steering. When the anode also points to a branch of the steering gear connected to the power supply, in most cases the steering gear is needed, the power converter.

3.2 Wireless Router

The wireless router here is actually the same as the one we use in our homes. WIFI is a network protocol, in fact, it is equivalent to a computer, can run the operating system, also need to install the driver. Router connection is shown in Fig.5.

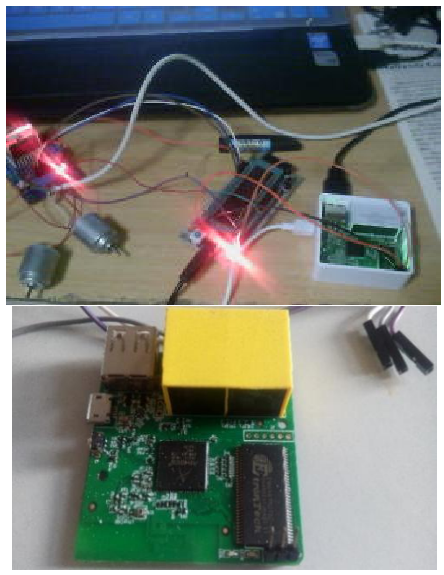


Fig.5. Logic control circuit

3.3 Principle of steering gear

The operating principle of steering gear is: in the same period, there are high and low level, high level is the reference signal, generally 1.5ms, the period is generally 20ms, which is the duty ratio we often mentioned. The difference in voltage between the high and low levels allows the small motor inside to rotate.

Control mode of steering gear is: The control of steering gear generally requires a 20ms time-base pulse, and the high level part of the pulse is generally the angle control pulse part within the range of 0.5ms~2.5ms [4]. Take the steering gear with an angle of 180 degrees as an example, then the corresponding control relationship is as follows: 0.5ms-- 0 degrees; 1.0 ms-- 45 degrees; 1.5 ms -- 90 degrees; 2.0 ms-- - 135 degrees; 2.5 ms -- 180 degrees; Duty ratio, cycle and signal line of the steering gear are shown in Fig.6.

Fig.6. Coplanar camera

4. The OpenWRT system and construction of development environment

OpenWRT is a small Linux system, which was originally researched by Apple. However, due to the use of Linux source code, it was forced to open the OpenWRT source code to comply with the GNU Project. That allowed us to use the prototype, which has evolved over time into an operating system commonly used in the embedded space [5].

Environment setup: OpenWRT is an open source system based on Linux. We can download the source code from its official website. Of course, before developing OpenWRT, we should first build Linux development environment Ubuntu.

4.1 Function implementation

This time, OpenWRT system was used to mainly complete the data acquisition of the camera, and the image was formed after MJPG video streaming software was installed. Through this system to mount the USB interface form of the camera, we use the Onio 12 megapixel camera. This system can receive the command code sent by the mobile phone client, and send it to the next computer through Ser2net software in the way of serial port. So as to achieve the control video car effect.

4.2 Socket programming

Socket is just an interface, it's just at the transport level, and of course in order to do network communication, we have to explore the next level, which is the IP level, so there's a lot of connection between network programming and IP addresses, port Numbers.

An IP address is the identity of a host on the Internet. The port number is used to send packets received by the host to that process. *Inet_aton*( ) is the function used to convert a string referred to by STRPTR into a 32-bit network byte order binary value. The header file is #include< ARPA /inet.h>, inet_addr( ). Same as above, and returns the converted address. The TCP server and client process is shown in Fig.7.

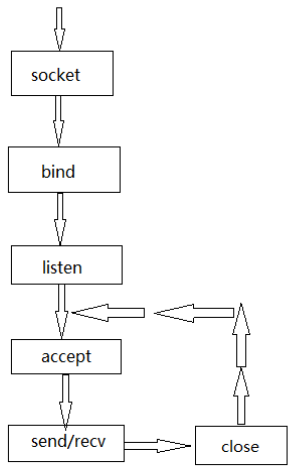


Fig.7. TCP processes

The function *socket*( ) creates a socket, *bind*( ) binds the address and port, *connect*( ) establishes a connection, and *listen*( ) listens for a socket. The function *accept*( ) accepts connections, *recv*( ) receives data, and *send*( ) sends data, *close*( ) closes the socket.

4.3 Description of WIFI video car

WIFI video car is mainly loaded with router, steering gear, two independent 5V power supply, minimal system and camera. The steering gear is used to control the up, down, left and right swing of the camera. This car has been successfully realized [5].

The mobile client commands are parsed, and the video collection data is sent to the mobile phone, and the mobile real-time monitoring system is successfully completed. WIFI video car is shown in Fig.8.

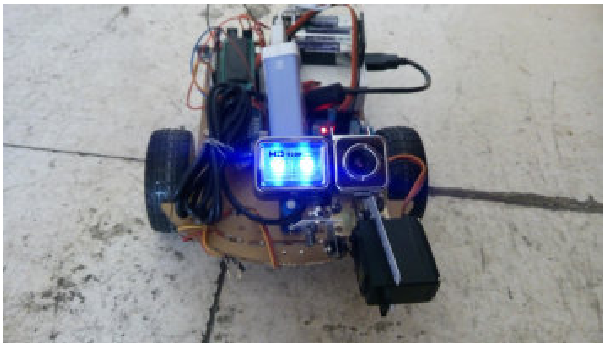


Fig.8. WiFi video car object

5. Conclusion

Embedded technology is now regarded as an indispensable high technology in the field of electronics. All the experiments in the design and coding are all belong to the category of embedded. In science and technology developed today, embedded single no longer refers to the single chip microcomputer, it has to grow, or even embedded into many pieces, such as embedded hardware engineer, embedded application engineer, embedded systems engineers, embedded driver engineer.

Our design involves the application layer and hardware, also used in the system, embedded development. Further, we should add a voice transmission function, thus realize the audio/video synchronization.

References

1. Li Chaoqing. Principle and Interface Technology of Single Chip Microcomputer. Beijing University of Aeronautics and Astronautics Press, 2013.
2. li shengli. Fundamentals of analog electronics technology. School of electrical and electronic engineering, Wuhan institute of engineering, 2013, 13.
3. Yuan Zengmin. Basic Course of Analog Electronics technology.Beijing: Electronics Teaching and Research Group, Tsinghua University, 2014.
4. Song jingbin.Linux network programming (2nd edition). Beijing: Tsinghua University Press, 2014, 2.
5. Liu gang, zhao jianchuan. Linux system transplantation (second edition). Beijing: Tsinghua University press, 2014, 2.

Design of WebGIS transportation and distribution system based on the genetic algorithm

Hongbo Hao[1] , Fengzhi Dai [1, 2 *]

[1] Tianjin University of Science and Technology, China;

[2] Tianjin Tianke Intelligent and Manufacture Technology CO., LTD, China

*E-mail: * daifz@tust.edu.cn*

www.tust.edu.cn

Abstract

With the concept of "smart logistics" put forward, the transformation of the logistics industry facing information technology is facing great challenges. At present, China's logistics industry is facing problems such as extensive resource allocation, low management level and high cost. Aiming at the transportation problem in logistics industry, this paper designs a solution method of balanced/unbalanced transportation problem based on the Monte Carlo similarity and genetic algorithm. Two coding methods, Prufer number and matrix are adopted. On this basis, dynamic mutation rate and random mutation strategy are designed, and Monte Carlo similarity receiving method is introduced. Finally, from the perspective of system requirements, the WebGIS transportation and distribution system based on genetic algorithm is designed and developed.

Keywords: smart logistics, balanced/unbalanced, Prufer, WebGIS

1. Introduction

Modern logistics industry is based on information technology and transportation technology. Logistics industry is composed of goods packaging, goods handling, goods allocation, goods management and goods storage and transportation. These elements are developed around transportation.

Therefore, the study of transportation is very important and it has a strong practical significance. The transportation problem is mainly to solve the problem of allocation between producing area and selling place, and its mathematical model belongs to the category of linear programming.

The traditional algorithm is not suitable for high-dimensional transportation problems. Therefore, the intelligent algorithms such as ant colony algorithm and neural network are usually used to solve the high-dimensional transportation problems.

In this paper, a genetic algorithm based on Monte Carlo similarity is proposed for transportation problem, and Monte Carlo similarity operator is constructed to avoid the algorithm falling into local optimal solution. The research and development system has a certain practical significance for the promotion and development of intelligent logistics [1].

This paper describes some basic theories involved in the algorithm, such as genetic algorithm, Monte Carlo, similarity theory and the related technical framework of WebGIS transportation and distribution system. It also includes the mathematical model and transformation relationship of the balanced/unbalanced transportation problem. An intelligent optimization algorithm based on the combination of Monte Carlo similarity and genetic algorithm is designed. The Monte Carlo similarity receiving method is introduced, and the conclusion that the genetic algorithm based on Monte Carlo and adopting matrix coding is better [2].

With the continuous enhancement of desktop processor performance, such as the WebGIS framework represented by ArcGIS API for JavaScript launched by Esri, the pressure on back-end services has been greatly reduced.

2. Overall scheme design

This paper uses the basic theory of genetic algorithm, Monte Carlo thought and common similarity algorithm. Then it introduces the key technology of system design and development. The front end uses bootstrap, Vue and ArcGIS API for JavaScript framework. The back end is based on the springboot framework, combined with spring, spring MVC and mybatis framework. MySQL is used as the system storage database.

2.1. Front end architecture

This paper uses bootstrap that is very popular nowadays. Vue.js.Bootstrap is an open source CSS framework developed by twitter engineers Jacob Thornton and mark Otto to improve code reusability. Bootstrap is a flexible, elegant, easy to use and extensible front-end framework.Vue. It is a progressive framework for building the front-end interface. Different from other frameworks, it can be applied layer by layer from bottom to top and developed based on MVVM mode Vue.JS Framework [3]. MVVM pattern is Vue.JS, one of the biggest features of the pattern architecture that is shown in Fig.1.

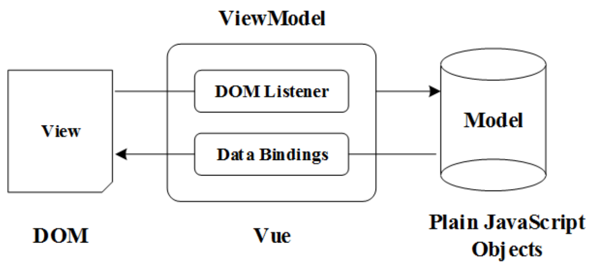


Fig.1. Vue.JS MVVM mode architecture diagram

2.2. Back end architecture

The Springboot solves the characteristics of various configuration, low development efficiency and complex deployment of back-end development. The Maven dependency is simplified by using start POM, and the configuration is completed by annotation. The single instance servlet improves the running efficiency. It is adopted, combined with spring, spring MVC and mybatis framework to complete the development. Therefore, spring takes IOC and AOP design patterns as the core to construct a bean container to store singleton objects. The architecture of spring is shown in Fig.2.

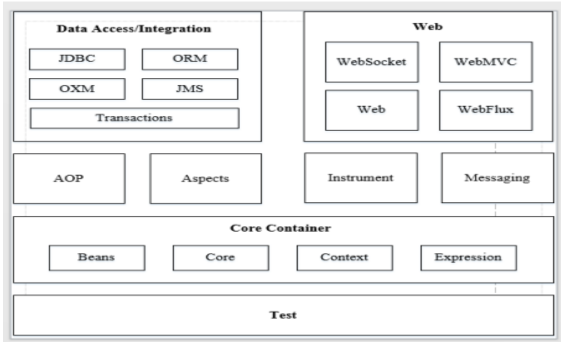


Fig.2. Spring architecture diagram

3. Design and solution of transportation model

3.1. Balanced transportation model

The equilibrium transportation problem is described as follows: suppose a product has m producing areas W_1, W_2...Wm and N outlets Q_1, Q_2...Qn. Among the output of producing area is W_i, and the sales volume is Qj. The cost of transporting a unit of goods from W_i to Q_j with a certain route s in a certain period of time is $C_{ij}(t, s)$, which is represented by C_{ij} below. Schematic diagram of origin and sale of place [4] is shown in Fig.3.

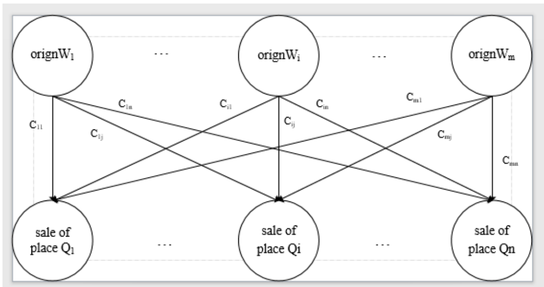


Fig.3. Schematic diagram of origin and sale of place

3.2. Fitness function design

The fitness function of genetic algorithm is used as an index to evaluate the quality of individuals in a population. The larger the fitness value is, the better the individual is. The fitness function is shown below.

$$f(X) = e^{-\lambda\left[tr(C^T X)-\omega\right]} \quad (1)$$

3.3. Solving transportation problems

The algorithm process steps are as follows:

- Set the operation parameters;
- Initial population;
- The fitness of each chromosome was calculated;
- The chromosomes were sorted according to the fitness;
- The first 10% individuals with high fitness were reserved;
- Monte Carlo similarity method is used to receive the difference solution between the mutated individual and the mutated parent [5].

3.4. Algorithm results and analysis

The performance of GA algorithm based on PRüfer number coding and matrix coding are compared. It can be seen that the convergence speed and optimal value of TGA and algorithms are not as good as the improved ITGA and IMGA algorithms without using strategies.

In the process of crossover and mutation of IMGA algorithm, the generated offspring meets the constraint conditions, while the ITGA algorithm is easy to generate infeasible solutions in the crossover process.

The experimental results show that the final convergence result of IMGA algorithm is better than that of ITGA algorithm. In Fig.4, the horizontal axis is the number of iterations and the vertical axis is the average of the optimal value of each iteration in every 30 experiments. The average value of the optiomal value of each algorithm's iteration is shown in the Fig.4.

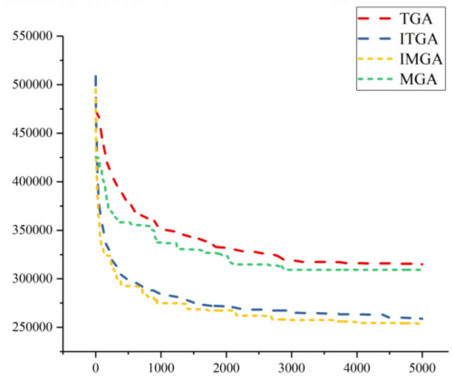


Fig.4. Average of the optimal value of each algorithm's iteration

4. Design of WebGIS distribution system

4.1. System Requirements Analysis

The WebGIS transportation and distribution system abandons the previous purely digital representation. It combines the GIS system to introduce spatial relationships into the transportation and distribution system, and selects production and sales locations in an intuitive and optional way. The system is mainly based on the task creation module. It is supplemented by task list module, OD cost module, distribution plan module, task analysis module and personal center module.

4.2. System technical route

The WebGIS transportation and distribution system adopts the B/S architecture model, which reduces the burden of the client on the server and the cost of system maintenance and upgrades. The system server level uses the GIS server in ArcGIS Enterprise and the Tomcat-based Web server. The system uses the SSM framework, which is divided into a presentation layer, a business processing layer and a data persistence layer [6].

4.3. System database design

The design of the system is mainly based on the task creation module. The table structure and relationship are designed in a task-driven way, including system user table, task list, expense table, expense detail table, distribution table, distribution details table, and algorithm analysis table , sales surface, algorithm iteration record table, etc.

4.4. System function module design

According to the actual system requirements, the WebGIS transportation and distribution system is equipped with various module functions of WebGIS, which are mainly divided into 6 modules: task list module, task creation module, OD cost module, distribution plan module, task analysis module and personal center module.

4.5. System function module design

The distribution plan module plays the role of query of calculation results in the WebGIS transportation and distribution system. This module undertakes the task creation module and the OD expense module, obtains the preprocessing data necessary for the algorithm. It calls the intelligent algorithm according to the task. The distribution plan module visually allows users to view the distribution of production and sales. With WebGIS technology, the production and sale area are marked in the form of a map. The number and shapes are combined to achieve the purpose of abstract digital visualization.

5. Conclusion

This article aims at the completeness and ease of use. The specific operation process of the WebGIS transportation and distribution system is shown in Fig5.

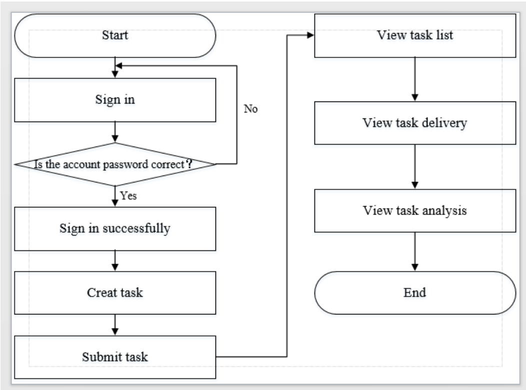


Fig.5. WebGIS distribution system flow chart

This paper takes the transportation and distribution system of a logistics company as the background. It combines the Monte Carlo similarity GA transportation problem algorithm based on matrix coding on the basis of considering the actual transportation problem requirements. The front-end framework uses Bootstrap, Vue.js and ArcGIS API for JavaScript [7]. The back-end framework adopts SSM and other Web system development technologies, and with the help of ArcGIS platform.

The WebGIS system that meets transportation and distribution is finally designed and realized. The system uses Intellij idea as the integrated development environment. The core algorithms and back-end services are built in Java language. The front-end development framework uses ArcGIS API for JavaScript, Bootstrap and Vue.js, etc., and data storage and management uses MySQL and ArcGIS DataStore [8].

Acknowledgements

The research is partly supported by the Project of Tianjin Enterprise Science and Technology Commissioner to Tianjin Tianke Intelligent and Manufacture Technology Co., Ltd (19JCTPJC53700). It is also supported by the Industry-University Cooperation and Education Project (201802286009) from Ministry of Education, China.

References

1. Elena Rudskaia, Rudskaia Elena, Eremenko Igor. Conceptual framework for solving problems of commercial transportation of goods in the city logistics system. *IOP Conference Series: Materials Science and Engineering*, 2020, 918(1): pp. 012038-.
2. Liu Lang. No-core Monte Carlo shell model calculations with unitary correlation operator method and similarity renormalization group. *Chinese Physics C*, 2015, 39(5).
3. Matevž Pustišek, Andrej Kos. Approaches to Front-End IoT Application Development for the Ethereum Blockchain. *Procedia Computer Science*, 2018, 129: pp. 410-419.
4. Meijie Ji. Study on Location Selection and Optimization of Logistics Center based on Particle Swarm Optimization. *International Core Journal of Engineering*, 2020, 6(9).
5. Qi Li, Rui Miao, Jie Zhang, et al. An Intelligent Recommendation Method for Service Personalized Customization. *IFAC-PapersOnLine*, 2019, 52(13): pp. 1543-1548.
6. YONGGANG ZUO, WEIPING ZHU. Aviation Logistics Information System Based on SSM Framework. *2017 2nd International Conference on Computer Engineering, Information Science and Internet Technology*, 2017, pp. 533-538.
7. A. Taivalsaari, T. Mikkonen, D. Ingalls and K. Palacz. Web Browser as an Application Platform, *2008 34th Euromicro Conference Software Engineering and Advanced Applications*, Parma, 2008, pp. 293-302.
8. Wiart Jaspar, Wong Elaine. Walsh functions, scrambled-nets, and negative covariance: Applying symbolic computation to quasi-Monte Carlo integration. *Mathematics and Computers in Simulation*, 2021, 182: pp. 277-295.

A Study of YOLO Algorithm for Target Detection

Haokang Wen [1], Fengzhi Dai [1, 2 *], Yasheng Yuan [1]

[1] Tianjin University of Science and Technology, China;

[2] Tianjin Tianke Intelligent and Manufacture Technology CO., LTD, China

*E-mail: * daifz@tust.edu.cn*

www.tust.edu.cn

Abstract

With the development of deep learning, target detection has become one of the research directions of many scholars. As one of the more mature algorithms, the YOLO series of algorithms have been widely used in real life. Combining the development history of the YOLO algorithm, this article focuses on the main framework and main content of the current latest YOLOv5 algorithm, and uses the YOLOv5 model to identify and detect footballs. This article evaluates its detection effect. The test results show that YOLOv5 has a wider application meaning in real life.

Keywords: target detection, YOLOv5, deep learning, computer vision technology

1. Introduction

Computer vision includes target detection, target segmentation, target tracking, image description, event detection, and activity recognition. Target detection is the cornerstone of other more complex vision tasks. Its main task is to use computers to predict: a given image and video Object, what is it or where it is.

Currently, target recognition technology is widely used in the following fields:

- Security field: fingerprint recognition, face recognition, etc.
- Military field: terrain survey, flying object recognition, etc.
- Traffic field: license plate number recognition, unmanned driving, traffic sign recognition, etc.
- Medical field: electrocardiogram, B-ultrasound, health management, etc.
- Life field: smart home, shopping, etc.

2. Object detection algorithm

With the rapid development of deep learning technology, since 2012 target detection algorithms have shifted from traditional target recognition algorithms based on manual features to target recognition technologies based on deep neural networks. A road map of object detection [1] is shown in the Fig.1.

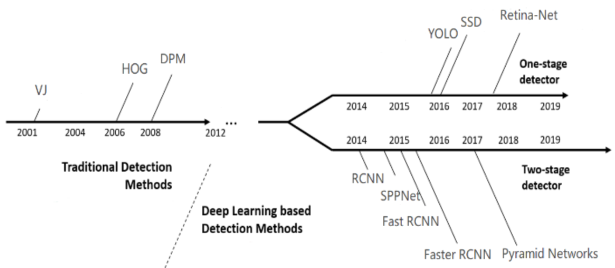


Fig.1 A road map of object detection

There are currently two mainstream target detection algorithms based on deep neural networks: one is a single-stage algorithm represented by YOLO series algorithms and SSD series algorithms, and the other is a two-stage algorithm represented by R-CNN series algorithms.

The single-stage algorithm treats the target detection process as a regression problem, and uses a unified deep neural network for feature extraction, target classification and bounding box regression, achieving end-to-end reasoning. It has a faster detection speed, but its detection accuracy is lower than the two-stage algorithm.

The two-stage target detection algorithm uses the region proposal network to extract the region of interest, that is, the region containing the target, and then uses a deep neural network to classify the region of interest and return the bounding box, which has higher detection accuracy. However, the detection speed of the two-stage target detection algorithm is too slow to meet the real-time requirements.

Comparing these two methods, the single-stage target detection algorithm is usually simpler, faster than the two-stage target detection algorithm, and has more advantages in real-time detection. This article introduces the detection algorithm based on YOLO algorithm.

3. YOLO algorithm

The YOLO series algorithm is one of the representatives of the one-stage target detection algorithm. Since the first generation YOLOv1 algorithm was proposed in 2016, after recent years of development, there are now five main versions proposed. The main features of each version are introduced below.

3.1. YOLO v1

YOLOv1 proposes to detect the target through grid division, and detect the target through the position of the target center point on the grid, which significantly improves the detection speed.

Before the YOLO algorithm was proposed, object detection methods were based on the method of first generating candidate regions and then detecting. Although there is a relatively high detection accuracy rate, the running speed is slow.

YOLO creatively treats the object detection task directly as a regression problem, combining the candidate area and the detection phase into one. "You Only Look Once", YOLO really can let you know what objects are in each image and where the objects are at a glance [2].

3.2. YOLO v2

YOLOv2 uses multi-scale feature maps to detect objects based on SSD, and proposes pass through layer to link high-resolution feature maps with low-resolution feature maps to achieve multi-scale detection.

YOLOv2 applies the anchor mechanism on the basis of grid constraints. By presetting a priori boxes of different scales, the detector focuses on detecting objects that are similar in shape to the a priori box. At the same time, the batch normalization method is used to accelerate convergence and avoid overfitting.

3.3. YOLO v3

YOLOv3 has made two improvements on the basis of YOLOv2. One is to use the residual model to further deepen the network structure; the other is to use the FPN architecture to achieve multi-scale detection.

YOLOv3 uses the residual network idea in Res Net to design Darkent-53 as the backbone network for feature extraction [3]. On the basis of YOLOv2, it draws on the multi-scale idea of FPN and designs 3 different scales. Three a priori boxes are allocated for each detection layer.

YOLOv3 predicts the position of the object through the method of frame regression prediction, which solves the problem of instability of the linear regression of the prior frame mechanism.

In YOLOv3, each box uses multiple label classifications to predict a bounding box might contain which classes [4]. YOLOv3 architecture [5] is shown in the Fig.2.

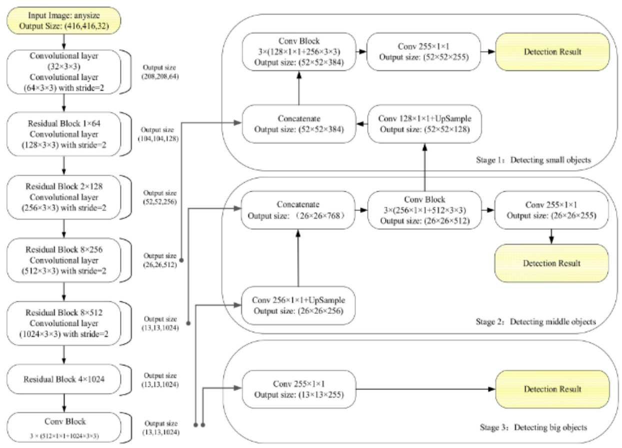


Fig.2 YOLOv3 architecture

3.4. YOLO v4

The backbone network CSPDarknet53 of YOLOv4 is the core of the algorithm and is used to extract target features. Drawing lessons from the CSPNet's experience in maintaining accuracy, reducing computational bottlenecks and memory costs, YOLOv4 adds CSP to each large residual block of Darknet53, divides the feature map of the base layer into two parts, and merges them through a cross-stage hierarchy. This reduces the amount of calculation and ensures accuracy. The activation function of CSPDarknet53 uses the Mish activation function, and the following network uses the leaky_relu function, so that the setting is more accurate in target detection.

Unlike the YOLOv3 algorithm that uses FPN for upsampling, YOLOv4 draws on the idea of information circulation in the PANet network. First, the semantic information of high-level features is propagated to the low-level network through up-sampling, and then it is fused with the high-resolution information of the underlying features to improve the detection effect of small targets. Then increase the information transmission path from the bottom to the top, and enhance the feature pyramid through downsampling. Finally, the feature maps of different layers are fused to make predictions.

3.5. YOLO v5

Compared with YOLOv4, YOLOv5 has a higher accuracy rate and better ability to recognize small objects [6]. YOLOv5 is more flexible and faster than YOLOv4, and has great advantages in the rapid deployment of models.

The YOLOv5 detection network uses CSPDarknet as the feature extraction network to extract rich information features from the input image. CSPNet solves the gradient information duplication problem of network optimization in other large-scale convolutional neural network frameworks, and integrates the gradient changes from the beginning to the end into the feature map, thus reducing the amount of model parameters and FLOPS values. This not only ensures the speed and accuracy of inference, but also reduces the scale of the model.

YOLOv5 proposes to use the Fcos algorithm to participate in the calculation of the frame selection area, which greatly improves the detection efficiency. And through image enhancement, new training samples are generated from the existing training data. Various advanced data enhancement techniques are used to maximize the use of data sets to achieve a breakthrough in the performance of the target detection framework. Through a series of image enhancement technology steps, the performance of the model can be improved without increasing the reasoning delay.

YOLOv5 uses the CSPDarknet feature extraction network to effectively extract image features, and uses BottleneckCSP instead of shortcut residual connection to strengthen the description of image features. The Neck module is mainly used to generate feature pyramids. The feature pyramid can enhance the model's detection of objects of different scales, thereby being able to identify the same object of different sizes and scales.

4. The test of YOLOv5

There are four network models in YOLOv5: YOLOv5s, YOLOv5m, YOLOv5l and YOLOV5x. Based on the environment that is shown in Table 1, the YOLOv5s model was built.

Table 1. Computer environment

GPU	NVIDIA GeForce MX150
Video memory	2 GB
operating system	Windows 10
CUDA architecture	CUDA 10.2

After training the neural network using YOLOv5s algorithm, actual tests were performed on football pictures and videos. As shown in the Fig.3, it is the result of detecting football.

Fig.3 The result of detecting football

Experiments show that the algorithm can detect football in real time even in complex scenes.

Due to the impact of insufficient training times and insufficient data sets, there are still some errors. But overall, this method has good results and can accurately identify footballs in many complicated scenes.

5. Conclusion

YOLOv5 is an excellent target detection algorithm. After years of development, the algorithm has been continuously improving the network structure to maintain the advantage of faster detection speed while maintaining high accuracy. YOLOv5 algorithm has very considerable prospects in future detection work. In theory, this method has a wide range of application value in real life.

References

1. Zou Z, Shi Z, Guo Y, et al. Object detection in 20 years: A survey, *arXiv preprint arXiv*: 1905.05055, 2019.
2. Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection, *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: pp.779-788.
3. Jing, Junfeng, et al. Fabric defect detection using the improved YOLOv3 model, *Journal of Engineered Fibers and Fabrics*, 2020(15): 1558925020908268.
4. Yuan Y, Dai F, Song Y, et al. On Fatigue Driving Detection System Based on Deep Learning, *Chinese Intelligent Systems Conference.* Springer, Singapore, 2020: pp.734-741.
5. Li Y, Zhao Z, Luo Y, et al. Real-Time Pattern-Recognition of GPR Images with YOLO v3 Implemented by Tensorflow, *Sensors*, 2020, 20(22): pp.6476.
6. Yifan Liu, BingHang Lu, Jingyu Peng, et al. Research on the Use of YOLOv5 Object Detection Algorithm in Mask Wearing Recognition, *World Scientific Research Journal*, 2020, 6(11).

Research on Recognition and Application of EEG Signal Based on SSVEP-BCI

Di Yin [1], Fengzhi Dai [1, 3*], Mengqi Yin[2], Yasheng Yuan[1], Yuxuan Zhu[1]

[1] *Tianjin University of Science and Technology, China;*

[2] *Hebei University of Chinese Medicine, China;*

[3] *Tianjin Tianke Intelligent and Manufacture Technology CO., LTD, China,*

*E-mail: * daifz@tust.edu.cn*

www.tust.edu.cn

Abstract

In recent years, brain-computer interface (BCI) systems based on steady-state visual evoked potentials (SSVEP) have attracted attention due to their high information transfer rate (ITR) and more and more targets. The current mainstream algorithms for SSVEP recognition have greatly improved the accuracy and target detection time. This paper designs a robotic arm application system based on the eCCA-Y method for multi-target recognition. The phase characteristics of CCA's sine and cosine signals are added to the EEG signal. Compared with mainstream algorithms, research shows that this method can improve the SSVEP-based BCI performance. And choose a six-degree-of-freedom manipulator as the actuator of the brain-computer interface, and use a phase-encoded stimulation paradigm for multi-target recognition to conduct experiments on the application of the proposed method.

Keywords: Brain-computer interface, Feature extraction, Multi-target recognition, EEG signal, Steady state visual evoked potential

1. Introduction

Patients suffering from diseases such as muscular dystrophy and spinal injury have physical disabilities but normal brain functions, but they cannot interact with the outside world like normal people. Therefore, an output channel can be created directly by the brain to the external environment, without communicating through peripheral nerves and muscle tissues [1,2]. Brain-computer interface technology is booming and can meet this demand. The brain signals issue commands to directly control external devices, so that patients can restore their body functions and live like normal people.

In the research of brain-computer interface, the feature extraction algorithm of EEG signal is to find a suitable spatial filter, and get the weight vector through projection and linear combination. Zhang et.al proposed the multi-channel CCA [3]. Wong et al. proposed eCCA [4], Masaki et al. introduced the task-related component analysis (TRCA) method to SSVEP identification [5].

We found that no phase information is seen in the sine and cosine reference signal, that is, there is no trace of phase information in the constructed fitting signal, so an algorithm based on eCCA (eCCA-Y) [6], in the SSVEP stimulation paradigm is proposed. The phase is added to the reference signal, and the phase of the reference signal is obtained by optimizing the parameters. Compared with the existing two mainstream algorithms, the recognition rate and ITR have a certain improvement.

In terms of application research, the EEG signal is used to control the robotic arm equipment to verify the effectiveness of the algorithm, complete the communication establishment of the two subsystems, and the EEG signal instructions are converted into motion control instructions for controlling the end effector of the robotic arm.

2. Application of EEG Signal Recognition Based on SSVEP-BCI System

According to the experimental intent of the stimulus paradigm and the current position of the robotic arm, the participants feedback and select the control instructions of the experimental paradigm. At the same time, the EEG acquisition equipment performs a series of processing (preprocessing, feature extraction and feature recognition). Map the recognized instructions to the actual robotic arm commands, and the robotic arm executes actions according to the received instructions.

As shown in Fig.1, it mainly consists of four parts: EEG signal acquisition module, signal decoding module, human-computer interaction module and command output module.

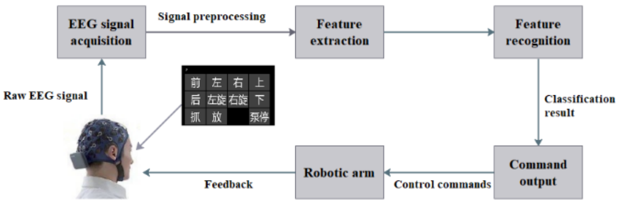


Fig.1 Application flow chart based on SSVEP-BCI system

3. Experiment

3.1. Experimental paradigm

The EEG acquisition system based on SSVEP-BCI compiles some inherent actions of the robotic arm through MATLAB's PSYCHTOOLBOX (PTB) toolbox to form a specific stimulation paradigm. The stimulus paradigm formed at the end of this experiment is shown in Fig.2.

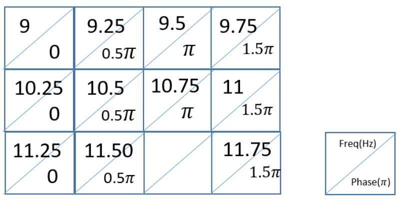


Fig.2 Manipulator stimulation paradigm based on SSVEP-BCI control

3.2. The experimental data

The data set collected SSVEP-BCI records of 5 healthy subjects (4 males and 1 female), the average age of the subjects was 24 years old. Before starting, the two subjects who have been tested before will be trained with data. For each subject, a total of 5 blocks were performed, each with 11 targets. The stimulation paradigm has been shown in Fig.2.

The multi-target recognition sinusoidal coding method was used to accurately generate 11 stimulation frequencies on the LCD display. Just as shown in Fig.2, the location of the eleven targets on the screen stimulation frequency corresponds to 9Hz, 9.25Hz, 9.5Hz, 9.75Hz, 10.25Hz, 10.5Hz, 10.75Hz, 11Hz, 11.25Hz, 11.5Hz and 11.75Hz.

Traverse 11 targets in turn, each target stimulus is 5s (including stimulus prompt 0.5s, stimulus flashing 4.5s). Participants try to avoid blinking each time the stimulus target blinks. Each time the stimulus target blinks, the subjects try their best to avoid blinking, so after the end of each group of experiments rest for 2-3 minutes in between. The experimental arrangement of the test data is shown in Fig.3.

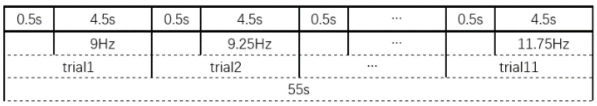


Fig.3 Experimental arrangement of test data

3.3. Data preprocessing

Preprocess the collected EEG data in the EEGLAB toolbox of MATLAB software, extract 8 channels (Pz, PO3, POz, PO4, PO6, O1, Oz, and O2), and get the brain electrical three-dimensional data (channel*points*trial).

In order to make the collected EEG data more convincing in the subsequent feature extraction and analysis process, the number of the experimental groups (block) is merged into the preprocessed three-dimensional EEG data and thus obtain the four-dimensional data (channel*points*trial*block). Down sampling is performed from 1000Hz to 250Hz, and each trial contains 3000 sampling points, which forms 8*750*11*5 EEG data.

4. Feature extraction

When the subjects are training data, they can obtain three kinds of multi-channel signals:

1. Test data $X(t) \in R^{N_C \times N_S \times N_t}$
2. The average signal of the subject's training data $\hat{X}_k{}^{N_C \times N_S}$
3. Constructed sine and cosine reference signal Y_{fk}

Among them, N_C, N_S, N_t represent channel, points, and test set trial respectively. Any two signals can be used to calculate a spatial filter based on CCA.

Chen et al.[7] combined the CCA system with the Pearson correlation coefficient of the test training data of the subjects and proposed a method to extend CCA. Only three spatial filter forms are taken, namely:

1. Test data and average training data $W_X(X\hat{X}_k), W_{\hat{X}_k}(X\hat{X}_k)$
2. Test data and reference signal $W_X(XY_{fk})$
3. Average training data and reference signal $W_{\hat{X}_k}(\hat{X}_k Y_{fk})$

According to the above three spatial filters, four correlation coefficient combinations as in formula (1) are selected:

$$r_k = \begin{bmatrix} r_k(1) \\ r_k(2) \\ r_k(3) \\ r_k(4) \end{bmatrix} = \begin{pmatrix} \rho(X^T W_X(XY_{fk}), Y_{fk}^T W_{Y_{fk}}(XY_{fk})) \\ \rho(X^T W_X(X\hat{X}_k), \hat{X}_k^T W_X(X\hat{X}_k)) \\ \rho(X^T W_X(XY_{fk}), \hat{X}_k^T W_X(XY_{fk})) \\ \rho(X^T W_{\hat{X}_k}(\hat{X}_k Y_{fk}), \hat{X}_k^T W_{\hat{X}_k}(\hat{X}_k Y_{fk})) \end{pmatrix} \tag{1}$$

The correlation coefficient is fused to obtain the correlation coefficient at the k-th stimulus frequency, and then the classification can be completed by selecting the largest correlation coefficient. The principle analysis diagram is shown in Fig.4.

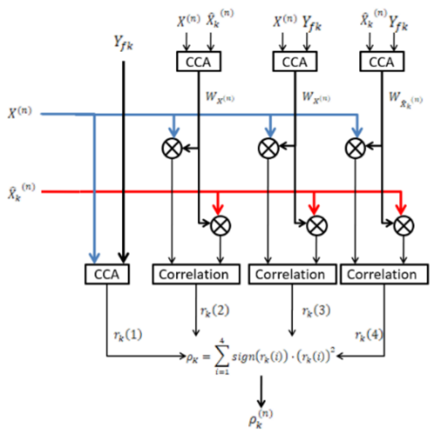


Fig.4 eCCA-Y schematic diagram

The reference signal is:

$$Y_{fk,\theta_k} = \begin{bmatrix} \sin(2\pi f_k t + \theta_k) \\ \cos(2\pi f_k t + \theta_k) \\ \vdots \\ \sin(2\pi N_h f_k t + N_h \theta_k) \\ \cos(2\pi N_h f_k t + N_h \theta_k) \end{bmatrix}^T \tag{2}$$

According to the experimental intent, the stimulus paradigm designed by the researcher determines the size of θ_k, as expressed by formula (3):

$$\theta_k = \theta_0 + \Delta\theta \times \left[(k_y - 1) \times 5 + (k_x - 1)\right] \tag{3}$$

Among them, k_x and k_y represent the row and column indices of the visual stimulation matrix, respectively. θ_0 represents the initial phase, and $\Delta\theta$ represents the phase interval.

5. Research on Application of Manipulator Based on SSVEP-BCI

The robotic arm application based on SSVEP-BCI is composed of two systems, namely the BCI EEG acquisition system and the robotic system. The BCI subsystem and the robot subsystem communicate through the UDP protocol. The BCI subsystem is composed of a wireless EEG acquisition system, LCD display and online processing of EEG data.

The EEG acquisition equipment is responsible for converting the recorded EEG signals into digital signals through AD, and through online real-time processing (feature extraction and conversion algorithms) to generate control instructions that can be recognized by the robotic arm. The reason for joining the application of robotic arm in this article is to study whether the robotic arm can provide convenience for the disabled and solve the simple grasping movements necessary in life.

The steps of the application design in this article are: (1) collect the subject's EEG data and record it as the subject's EEG intent, (2) perform data processing and analysis on the saved data in the MATLAB software, (3) and finally convert it into the control command of the robotic arm and send it to the controller.

In this experiment, the EEG data of 5 students in school was collected, feature extraction and recognition and classification were performed, and finally it was transformed into instructions that can control the robotic arm and sent to the robotic arm controller.

This experiment analyzes the data of 5 subjects. The number of harmonics is 2 and the data length is 750 sampling points to compare the classification effects of the two algorithms eCCA and eCCA-Y, as shown in Fig.5.

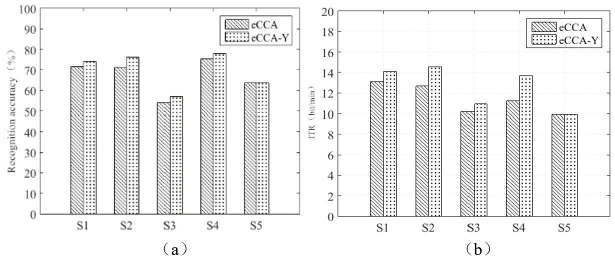


Fig.5 eCCA-Y schematic diagram

It can be seen from Fig.5(a) that the optimized eCCA-Y algorithm in this paper has improved recognition rate. Among them, S5 has the same accuracy rate for the two algorithms, and the recognition rate of S3 is relative to the other four. The test is low, S2 subjects have the largest difference in recognition rates between the two algorithms. In terms of ITR, as shown in Fig.5(b), the eCCA-Y algorithm is also improved, and S5 also shows the same effect as the recognition accuracy. The two algorithms are the same in terms of ITR. In addition, S4 subjects have the largest difference in ITR for the two algorithms.

6. Conclusion

Based on the SSVEP-BCI system to carry out application research experiments on the robotic arm, a multi-target recognition stimulus paradigm for the robotic arm is designed. The UDP protocol is used to realize the communication between the robotic arm and the brain-computer interface system. It is added at different frequencies. Different phases simplified the control method, refined the control of the robotic arm movement, and weakened its complexity.

Two feature recognition algorithms, eCCA and eCCA-Y, are used to process data, verifying the robotic arm and the brain-computer interface. The application realizes the interactive control of the robotic arm, and adds a stroke to the feature processing algorithm, which provides experience for the development of the BCI system.

Acknowledgements

The research is partly supported by the Project of Tianjin Enterprise Science and Technology Commissioner to Tianjin Tianke Intelligent and Manufacture Technology Co., Ltd (19JCTPJC53700). It is also supported by the Industry-University Cooperation and Education Project (201802286009) from Ministry of Education, China.

References

1. Chang H C, Lee P L, Lo M T, et al. Independence of amplitude-frequency and phase calibrations in an SSVEP-based BCI using stepping delay flickering sequences. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2011, 20(3) : pp.305-312.
2. Mak J N, Wolpaw J R. Clinical applications of brain-computer interfaces: current state and future prospects. *IEEE reviews in biomedical engineering*, 2009, 2: pp.187-199.
3. Zhang Y.et al. Multiway canonical correlation analysis for frequency components recognition in SSVEP-based BCIs. *International conference on neural information processing. Springer, Berlin, Heidelberg*, 2011: pp.287-295.
4. Wong CM, Wan F, Wang B, et. al. Learning across multi-stimulus enhances target recognition methods in SSVEP-based BCIs. *J. Neural Engineering*, 2020, 17: 016026.
5. Nakanishi M, Wang YJ, Chen XG, Wang YT, Gao XR, Jung TP. Enhancing detection of ssveps for a high-speed brain speller using task-related component analysis. *IEEE Transactions on Biomedical Engineering*, 2018, 65: pp. 104-112.
6. Yin D, Dai F, Song Y, et al. On Improvement of Target Recognition eCCA Method Based on SSVEP. *Chinese Intelligent Systems Conference. Springer, Singapore,* 2020: pp. 742-749.
7. Chen X, Wang Y, Nakanishi M, et al. High-speed spelling with a noninvasive brain-computer interface. *Proceedings of the national academy of sciences*, 2015, 112(44): pp. E6058-E6067.

Research on Bad Driving Detection Based on Behavior Recognition

Yasheng Yuan[1], Fengzhi Dai[1, 2, *], Di Yin[1], Yuxuan Zhu[1]

[1]Tianjin University of Science and Technology, China;
[2] Tianjin Tianke Intelligent and Manufacture Technology CO., LTD, China;
E-mail:[*]daifz@tust.edu.cn
www.tust.edu.cn

Abstract

Dangerous driving behavior is considered to be the direct or indirect reason of road accidents. Although artificial video surveillance is good to prevent bad driving, it wastes too much time and manpower. How to effectively identify behavior becomes the focus of the research. In recent years, deep learning showed the huge advantage in the field of computer vision. This paper adopt a number of deep learning network models, mining video integration of space and time features, introduction of analogy in human visual attention mechanism, improve the model deeply, using the LSTM to accurate and efficient video behavior analysis technology.

Keywords: deep learning, fatigue testing, convolutional neural network, action recognition

1. Introduction

With the maturity and development of internet technology, especially the development of video surveillance system, people tend to use more and more monitoring method to prevent some accidents. Video monitoring is an important part of prevention measures, the importance and actual efficiency were higher than human security. In the highly security demand of modern life, it is an essential part of a high and new technology, the action of driving of the driver is that people need to use one of monitoring. Research [1] shows that bad driving behavior of drivers is an important cause of traffic accidents. How to effectively monitor the driver's behavior is the key to safe driving.

Compared with traditional detection technology, deep learning technology shows the strong feature extraction ability. For target detection that performs well in deep learning, although it can more accurately detect the smoking [2], but for the cryptic action of fatigue, no time-level characteristics of target identification is obviously weak [3]. Especially in the video, the driver's side head, bowed their heads and normal driving operation will greatly improve the target detection model miscalculation, so network have time-level characteristic behavior identification, will have high research significance.

2. Main research method

Based on the deep learning on the driver's action recognition classification, through the analysis of the behavior recognition algorithm and the network model. Two-stream network is our choice as our main network, on the basis of the Two-stream network add amount Conv layers, improve the effect of feature extraction and increase the nonlinear. Adding BN layers and Dropout layers to prevent over fitting, and chooses appropriate activation function to further reduce the amount of calculation. The LSTM (Long Short-Term Memory) layer is added to further learn the timing information between behaviors and mine the long time dependence relationship of video sequence.

2.1 Data set processing

The training data set for this project contains 13,360 YouTube clips from the UCF101 data set and driving videos from volunteers. The videos mainly include five types of movements: human and object interaction, only body movements, human interaction, playing music equipment, and various kinds of sports. 22 kinds of videos of facial movements from UCF101 data set is selected. About 300 short videos in total, this videos have bad driving behaviors, such as smoking, talking on the phone, taking a nap and looking left and right, which were collected by ourselves from the vehicle camera during driving.

First, each video is divided into several short videos of about 120 frames (10s). After that, each short video is extracted frame by frame into RGB images, which are used as the input of spatial flow. Then, through the OpenCV and CUDA, the RGB frames extracted from the video are converted into optical flow diagrams [4].

The optical flow is the instantaneous velocity of pixels moving in space on the observed imaging plane. The optical flow method uses the change of pixels in the time domain and the correlation between adjacent frames. The method find the corresponding relationship between the previous frame and the current frame. It is a method to calculate the motion information of objects between adjacent frames. The projection of the motion for the object in the three-dimensional space on the two-dimensional imaging plane is shown in Fig.1. The result is a two-dimensional vector describing the change of position. In the case of minimal motion interval, we usually regard it as a two-dimensional vector describing the instantaneous velocity of the point u = (u, v), which is called the optical flow vector [5].

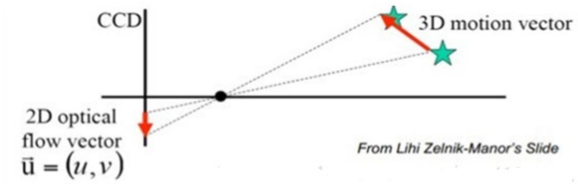


Fig.1. Optical flow method

TVL1 optical flow method in OpenCV to extraction is used in this paper. Each image obtained an optical flow diagram extracted from the X and Y directions respectively. The resulting optical flow diagram is shown in Fig.2.

Fig.2. Optical flow frame and image frame

2.2 Net selection

The two-stream convolutional network [6] is added by some network in this paper, which is divided into spatial convolutional network and time convolutional network.

The spatial-flow Convolutional network operates on a single video frame. Because some behaviors are strongly related to specific scenes and objects, it is effective to recognize behaviors from still images. As CNN (Convolutional Neural Networks) network is already a powerful image recognition algorithm, it can build a video recognition network based on large-scale image recognition algorithm. It can also utilize the pre-training network on the existing image classification data set. Time-flow convolutional network is different from ordinary CNN network. The input of time-flow convolutional network is formed by stacking optical flow displacement field between several consecutive frames. This input is characterized by the motion between the video frames, making the recognition process easier (no network implicit estimation of the motion is required).

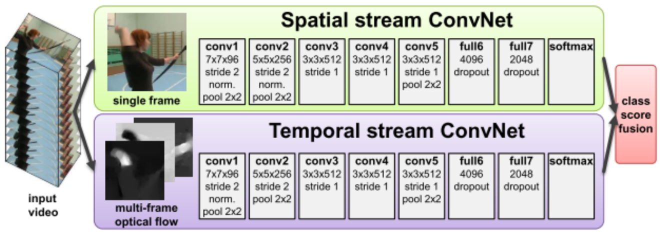


Fig.3. Two-stream Convolutional network

The two-stream convolutional network is shown in Fig.3. The algorithm is divided into two branches of convolutional neural network for feature extraction. First, a frame of RGB image in the video stream was input. After five layers of convolution and three layers of sampling, it was passed into the full connection layer of the two layers for feature extraction. The extracted spatial feature was input into the Softmax classifier for category prediction. Second, the optical stream features extracted from the video stream were input into the convolutional neural network with the same feature extraction branch of pixel space. Finally it passed into the Softmax classifier for category prediction.

2.3 Identify face areas

The network behavior identification network selected in this paper is improved based on two-stream, and the number of network layers is deepened on this basis. Before each lower sampling layer, a batchsize normalization layer, which means BN (Batch Normalization) & Relu [7], is added to prevent the over-fitting caused by too few training samples and too deep network. The network structure is shown in Table 1.

Table 1. Network structure

Layer name	Convolution kernel size	Convolution kernel number
Flow_conv_begin	7	64
BN &RELU	4	
Conv_begin_pool	3	---
Layer_1_Conv1_1	1	64
BN &RELU	4	
Layer_1_Conv1_2	3	64
BN &Relu	4	
Layer_1_Conv1_3	1	256
Layer_1_Conv1_expand	1	256
Layer_1_Conv1_sum	1	256
BN &Relu	4	
Layer_1_Conv2_1	1	64
BN &Relu	4	
Layer_1_Conv2_2	3	64
BN &Relu	4	
Layer_1_Conv2_3	1	256
Layer_1_Conv2_sum	1	256
BN &Relu	4	
Layer_2_Conv1_1	1	128
BN &Relu	4	
Layer_2_Conv1_2	3	128
BN &Relu	4	
Layer_2_Conv1_3	1	320
Layer_2_Conv1_sum	1	320
Layer_2_Conv1_expand	1	320
BN &Relu	4	
Layer_2_Conv2_1	1	128
BN &Relu	4	
Layer_2_Conv2_2	3	128
BN &Relu	4	
Layer_2_Conv2_3	1	320
Layer_2_Conv2_sum	1	320
BN &Relu	4	
Layer_2_Conv3_1	1	123
BN &Relu	4	
Layer_2_Conv3_2	3	128
BN &Relu	4	
Layer_2_Conv3_3	1	320
Layer_2_Conv3_sum	4	
Conv_final		512
BN &Relu	4	
Pool	8	---
Dropout	0.2	
Global_pool	Avg	
FC	16	
Loss		

The BN layer and Dropout mechanism is added, which is used to a certain extent and avoid network of fitting. Relu is selected as the activation function, change the sigmoid disappeared as a result of activation function gradient (sigmoid close to the saturated zone, change is too slow, derivative tends to zero) to complete the deep net training. The interdependencies between parameters are reduced, which can reduce the happen of over fitting. Through the proper superposition of 1x1 Convolutional layer and 3x3 Convolutional layer, the features are further extracted while the computation is reduced as much as possible. Finally Global pool is added to simplify the calculation.

2.4 Facial image processing

LSTM is an improvement based on the traditional cyclic neural network (RNN). It is very suitable for processing video sequences with time-dimension characteristics. In addition, LSTM network can avoid the gradient disappearance and gradient explosion problems in the practical application of RNN network.

In the classical LSTM structure, there are three "gates": (1) Oblivion gate, which is used to help RNN selectively forget some historical information; (2) Memory gate, which is used to strengthen the memory of some historical information; (3) Output gate, which is used responsible for comprehensive consideration of long and short term memory information to generate output signals.

In this paper, LSTM network (Fig.4) is also used for video signal timing modeling.

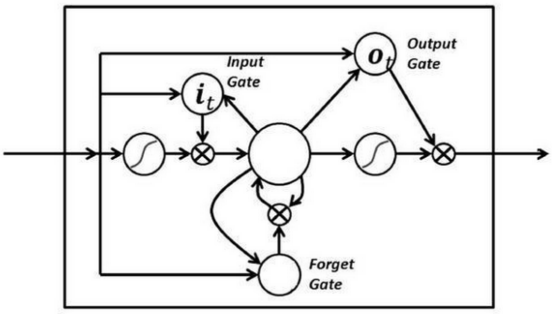


Fig.4. LSTM network structure

The key of the LSTM is cell state. The unit state is kind of like a conveyor belt. It goes all the way down the chain in a straight line, with only a few smaller linear interactions. It is very easy for information to flow unaltered.

State of cell transformation, namely value C transformation, have two operations. The two operation determines the forgotten and update of the computed tomography C_{t-1}. First of all, C_{t-1} through the multiplication of point-wise operation, this step determines whether the C_{t-1} value is forgotten. The number multiplied by C_{t-1} comes from the lower layer, the bottom of the data through a sigmoid layer, through the data values are below (0, 1). In general, when the input value is greater than 3 or less than -3, sigmoid value is close to the 1 and 0. That is to say, the value of C is generally going to be close to C_{t-1} unchanged, or 0, which means that you remember C or you forget C_{t-1}. Of course, a number between (0, 1) multiplied by C_{t-1} means how many C_{t-1} values you need to remember. Next, the value of C_{t-1} encounters an addition, which is a change to the value of C_{t-1}, such as the addition of new information, which is generally understood as something new to be remembered [8].

3. Improved PERCLOS algorithm

The network structure of this paper is shown in Fig.5. First, after the image flow and optical flow respectively pass through the convolutional network, the timing context information of the input video sequence is obtained through the LSTM layer. Two Softmax layers and two cross entropy loss layers are deployed after the feature output layer and the LSTM layer, respectively. The supervised information of the video category is used to drive the joint training of the LSTM and CONV networks. The hidden state of the LSTM layer can capture the time evolution of a specific category in the video sequence.

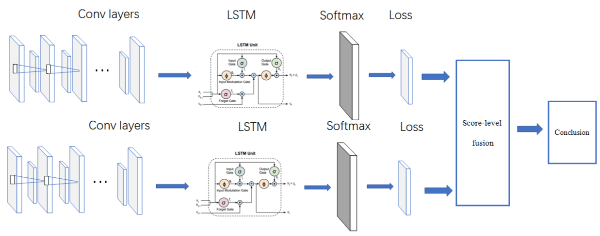


Fig.5. Two-stream-LD network structure

4. Experiment and Conclusion

All test sets are used to experiment on the model trained by the initial network, and the normalized confusion matrix of various behavior recognition rates is shown in Fig.6.

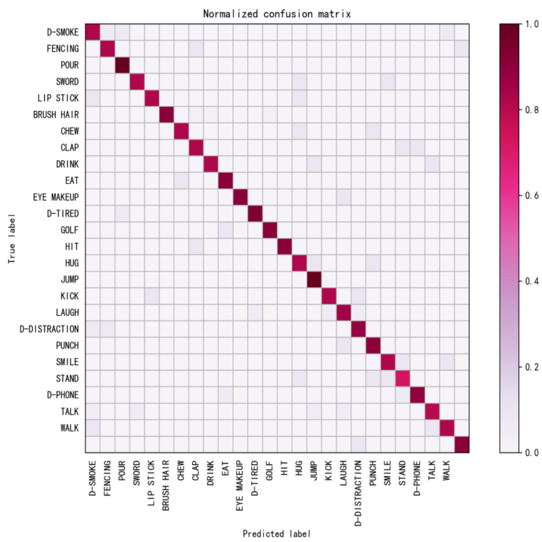


Fig.6. The normalized confusion matrix of various behavior recognition rates

In the matrix, the vertical axis represents the real label of a certain video behavior, and the horizontal axis represents the predictive label of a certain video behavior.

The darker the grid matrix color is, the higher the recognition rate is. And the lighter the grid matrix color is, the lower the recognition rate is. In the darker diagonals, the data contained represent the correct recognition rate for a certain type of behavior, while the data in the other light colored squares represent the wrong recognition rate when a certain type of behavior is predicted to be other behaviors.

Altogether 25 types of human behaviors were identified in this experiment. Among them, there are similar frames among the behaviors that need to use hands in front of the face, such as smoking, making phone calls and making up, so the error recognition rate among these people may be relatively high.

Based on the data in the confounding matrix, the dual flow convolutional neural network designed in this paper can effectively complete the task of recognizing and classifying bad driving behaviors.

Finally, the results are fused at the decision level. For the classification results obtained by multiple classifiers in the classification problem, if each classifier has different influences on the final prediction results, then the fusion can be considered in the way of weighted sum. We will find that, compared with speech and subtitle, the visual features contained in the image sequence are more helpful to determine the accurate behavior contained in the video. Therefore, we choose to give a larger weight to the classifier results of the video images, and give a smaller weight to the classification results of other classifiers.

The final result is shown in Table 2, from which can be seen that the Two-stream did well in UCF101 data set, this means that the Two-stream network in extracting time has good extraction effect when the flow characteristics, but in RGB space flow identification process is not ideal, but in our network of Two-stream-LD, both the space and time flows recognition effect has a significant improvement.

In the detection of bad driving behaviors, the Two-stream-LD network can be used to better identify various actions of the driver in the video. After comprehensive comparison, the method proposed in this paper is of certain research significance.

Table 2. Final result

Network	UCF101		
	RGB	Flow	RGB+Flow
Two-stream	83%	85.6%	89%
TSN	81.4%	79.6%	87.2%
IDT	78.2%	81%	86.4%
C3D	85.2%	--	--
LSTM	81.3%	--	--
3D Fused	82%	83.8%	86%
Two-stream-LD	88%	86%	90.3%

References

1. State of the States Report on Drowsy Driving, *National Sleep Foundation*, Nov. 2007, U.S.A.
2. Yasheng, Y, Fengzhi, D, Yunzhong, S, et al., On Fatigue Driving Detection System Based on Deep Learning. *Proceedings of 2020 Chinese Intelligent Systems Conference (CISC). Lecture Notes in Electrical Engineering*, 2020, pp.734-741.
3. Yasheng, Y., Fengzhi, D., Lingran, A, et al., Research on fatigue detection method based on deep learning. *Proceedings of 2020 International Conference on Artificial Life and Robotics*, Oita, Japan, 2020: pp. 640–643.
4. Zach C, Pock T, Bischof H. A Duality Based Approach for Realtime TV-L1 Optical Flow. *Proceedings of the 29th DAGM conference on Pattern recognition,* 2007: pp. 214–223.
5. Wang L, Xiong Y, Wang Z, et al. Temporal Segment Networks: Towards Good Practices for Deep Action Recognition, *European Conference on Computer Vision. Springer,* Cham, 2016: pp.20-36.
6. Feichtenhofer C, Pinz A, Wildes R P. Spatiotemporal Residual Networks for Video Action Recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, 2017: pp.3476-3484.
7. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society*, 2016: pp.770-778.
8. Donahue J, Hendricks L A, Rohrbach M, et al. Long-term Recurrent Convolutional Networks for Visual Recognition and Description. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39(4): pp.677-691.

Research on Crop Image Recognition Technology - Take Daylily as an Example

Jichao Zhao, Fengzhi Dai

College of Electronic Information and Automation, Tianjin University of Science and Technology, Tianjin, China,

*China E-mail: * daifz@tust.edu.cn*
www.tust.edu.cn

Abstract

Digital image recognition technology is the core technology of agricultural robots. The key research content is to judge the maturity of crops through image recognition. In this paper, the daylily is used as the crop to be identified, and digital image processing technology is used to separate the daylily from the background image, and then send the processed information to the picking actuator of the robot to assist the actuator in picking tasks. This method can not only be used in the recognition of daylily but also can be used in the recognition of tomatoes, cucumbers, and other crops by adjusting the recognition pictures and recognition parameters. It has high scalability.

Keywords: Crop Picking; Daylily; Digital Image Recognition; Maturity

1. Introduction

Agricultural robot refers to a kind of flexible automation or semi-automation equipment that takes agricultural products as the operating object, and has some human information perception and limb action functions. It is a kind of flexible automation or semi-automation equipment that integrates a variety of disciplines [1]. Agricultural robots can gradually replace human labor and continuously help agricultural production to reduce labor intensity. At the same time, they can also improve labor efficiency and help solve the problem of labor scarcity faced by many countries. Agricultural machinery has been paid more and more attention by developed and developing countries with less agricultural population, and has become one of the focal points of international agricultural equipment industry technology competition.

Agricultural machinery picking operations, first of all through the visual system for the identification and positioning of agricultural products [2]. The performance of the visual system determines the accuracy and efficiency of identification and positioning of agricultural products, and has a direct impact on the picking effect of the robot.

The visual system is generally composed of camera, image processing device, distance measuring device and computer. At present, the number of commonly used cameras is 1 or 2, namely monocular or binocular vision system. Images of the produce are captured by a camera before picking, then processed to identify the produce and its branches and leaves, and finally to determine the exact spatial coordinates of the produce [3-5]. Finally, actuators such as manipulator are driven to pick agricultural products at a certain Angle and strength. In order to avoid the influence of natural light, artificial light sources or optical filters are often needed to reduce the influence of surface shadows on the recognition effect of agricultural products.

In this paper, a crop recognition algorithm is proposed, which can extract the crops to be recognized from the crop image collected from the camera of agricultural robot. The extraction effect of the algorithm was tested with daylily as the main example. The generality of the algorithm is verified by testing on tomato.

2. Recognition Algorithm

According to their growth characteristics, crops can be divided into two categories: color changes before and after maturity, and color does not change. For instance, daylily and tomato fall into the category of color change, while cucumbers and beans fall into the category of color un-change.

Daylily is mainly distinguished from the color of its flower, which is cyan when immature and light yellow when mature. As shown in Fig.1, ripe daylilies are on the left and immature daylilies are on the right.

Fig.1. Daylily, left is ripe and right is immature

Our algorithm is mainly applicable to the extraction of crops with different colors before and after maturity. The algorithm flow chart is shown in Fig.2.

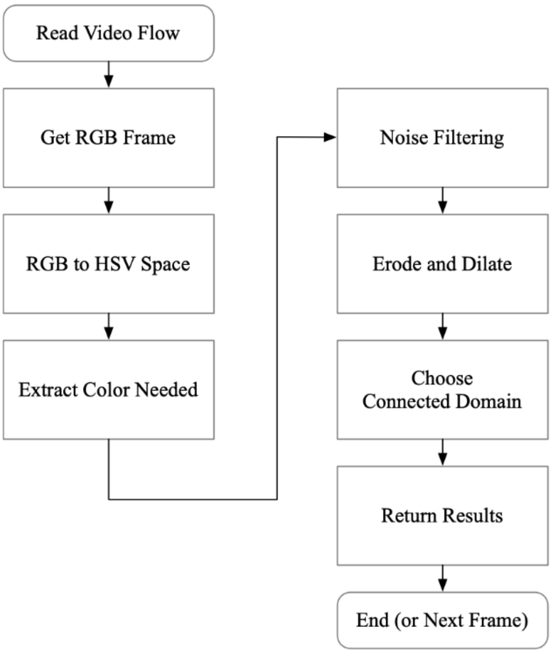


Fig.2. Algorithm Flow

2.1 HSV Color Space

The commonly used color space is RGB color space. Each pixel of the image contains three channels consist of Red, Green, Blue, and the final pixel color is composed of three colors mixed together. RGB color space is the cube model. HSV is a kind of color space created based on the intuitive characteristics of colors. It is also known as a Hexcone Model. The RGB color model is hardware-oriented, while the HSV color model is user-oriented. RGB space model and HSV space model are shown in the Fig.3.

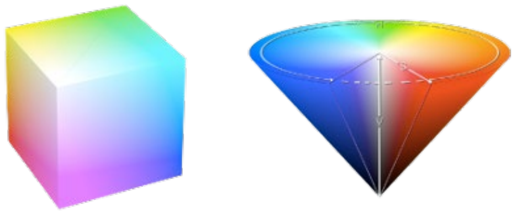


Fig.3. Color-Space, left is RGB-Space, right is HSV-Space

In HSV space, H represents Hue, which is measured at an Angle ranging from 0° to 360°. The value is calculated counterclockwise from red, which is 0°, 120° in green, and 240° in blue. Their complementary colors are: yellow is 60°, cyan is 180° and purple is 300°. S is the Saturation, which indicates the degree of Saturation of colors. If the white component of spectral color is 0, the saturation value is highest, the color is deep and bright. Generally, the value range of S is 0% ~ 100%. The larger the value, the more saturated the color is. V is Value, representing the brightness of the color, ranging from 0 to 1.

Compared with RGB space, HSV space can intuitively express the brightness, tone and brightness of colors, which is convenient for color contrast and emotional transmission. Therefore, HSV color space is selected to identify daylily. In addition, the main parameter that affects the identification is Hue, and we constantly adjust the H value to identify the daylily with good maturity. The effect of daylily extracted through HSV color space is shown in Fig.5(a).

2.2 Noise Filtering

When crops are extracted by color, similar color areas that do not belong to the crops to be extracted will also be selected, so further screening is needed. Look at Fig.5(a), the characteristic of the noise area is relatively fragmentary and disconnected, so the noise can be filtered through the filtering algorithm.

Common filtering algorithms include median filtering and mean filtering. In this paper, median filtering algorithm is used to filter the image noise after testing the two effects. The effect of median filtering is shown in Fig.5(b).

2.3 Dilate and Erode

Dilate is the operation of finding the local maximum. In contrast to dilate, erode is the operation of finding a local minimum. Erode followed by dilate is called open operation that can remove isolated small points, while the total position and shape remain unchanged. After dilate, erode is called close operation that can fill the small holes and bridge the small gaps, which also does not change the shape and position of the image.

In this paper, we will start with the open operation to remove the noise other than the extraction of daylily. Then the close operation is used to fuse the small distance area and fill the blank area inside the image. Finally, the morphologic images are obtained.

2.4 Choose Large Connection Domain

After the morphological processing is completed, there are still large areas in the image that have not been filtered out, so the image can be further screened by selecting the large connection domain.

Moreover, since the mature daylily is yellow-green, there is a clear color difference from the immature daylily. The color extraction method not only separates the day-lily from the branches, but also can judge whether the day-lily is mature or not. When it is mature, it can be detected and recognized, while when it is immature, it will automatically ignore the effect. The extracted daylily is shown in Fig.5(c).

3. Experiments

In this experiment, daylily was selected as the identified crops and were extracted from the pictures by adjusting the values of the algorithm parameters. Meanwhile, the extraction effect of tomato was also verified. But the parameter value of tomato is different from that of daylily, which also proves the universality of the algorithm.

3.1 Daylily

The daylily pictures used in this experiment were taken by our team members using mobile phones in the daylily planting base. The daylily is shown in Fig.4.

Fig.4. Daylilies original image

(a) After HSV color space extraction

(b) After morphological processing and noise filtering

(c) After selecting the large connection domain

Fig.5. Extraction results of daylilies

The parameters used to extract daylilies are shown in Tab.1 below.

Tab.1. Parameters of daylily extraction

Step Name	Param	Value / Range
HSV-space extraction	H	[0.10, 0.20]
	S	[0.26, 1.00]
	V	[0.50, 1.00]
noise filter	median	-
erode and dilate	radius	5px
large connection domain	>threshold	[50, +\infty]

3.2 Tomato

At the same time, we adjusted the parameters of the algorithm to carry out the experiment of tomato extraction from Fig.6 (left), the result of which is shown in Fig.6 (right).

Fig.6. Tomatoes extraction

Since the extraction parameters of tomato are the same as daylily except for values of HSV, only these three items are shown in Tab.2.

Tab.2. Parameters of tomatoes extraction

Step Name	Param	Value / Range
HSV-space extraction	H	[0.00, 0.10]
	S	[0.50, 1.00]
	V	[0.50, 1.00]

4. Conclusions

This paper presents an algorithm for crop extraction, which can extract crops with different colors before and after ripening. The algorithm was tested on daylily and tomato, and the extraction effect of the algorithm was demonstrated.

The validity and generality of the algorithm are verified by experiments. In addition, it can be seen from the experiment that this algorithm can be applied to the extraction of other crops only by adjusting the parameter values in the HSV color space.

Acknowledgements

The research is partly supported by the Project of Tianjin Enterprise Science and Technology Commissioner to Tianjin Tianke Intelligent and Manufacture Technology Co., Ltd (19JCTPJC53700). It is also supported by the Industry-University Cooperation and Education Project (201802286009) from Ministry of Education, China.

References

1. D. Khort, A. Kutyrev, et al. Robotized Platform for Picking of Strawberry Berries. 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), Kyiv, Ukraine, 2019, pp. 869-872.
2. Patel, Alakh and Jaymin Bhalani. A Review on in Field Cotton Recognition for Cotton Harvesting Robot based on Image Processing Technique. International Journal of Computer Applications 180 (2018): 6-8.
3. K. N. e. Alam Siddiquee, M. S. Islam, et al. Detection, quantification and classification of ripened tomatoes: a comparative analysis of image processing and machine learning. IET Image Processing, vol. 14, no. 11, pp. 2442-2456, 18 9 2020.
4. A. Wang, Y. Xu, et al. Semantic Segmentation of Crop and Weed using an Encoder-Decoder Network and Image Enhancement Method under Uncontrolled Outdoor Illumination. IEEE Access, vol. 8, pp. 81724-81734, 2020.
5. M. W. Liu, M. Ozdogan and X. Zhu. Crop Type Classification by Simultaneous Use of Satellite Images of Different Resolutions. IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 6, pp. 3637-3649, June 2014.

Design of Daylily Agricultural Picking Robot

Jichao Zhao, Fengzhi Dai *

College of Electronic Information and Automation, Tianjin University of Science and Technology, Tianjin, China

*China E-mail: * daifz@tust.edu.cn*
www.tust.edu.cn

Abstract

Agricultural picking activities occupy an important role in daily life, and the research and development of agricultural picking robots play a more crucial role and increasingly become an advantageous tool to improve agricultural productivity. Daylily, as a crop for daily consumption, can significantly reduce serum cholesterol and has extremely high nutritional value. However, the picking conditions of daylily are harsh, and long-term picking is likely to cause extremely serious harm to the human body, especially the hands. This paper designs an agricultural picking robot that can pick daylily automatically, which can greatly free hands and improve labor productivity. Especially in health, it can significantly reduce the harm to the human body.

Keywords: Picking Robot, Daylily, Agricultural, Robot Design

1. Introduction

Daylily, also known as Golden Needle, has been cultivated in China since ancient times. The flower of daylily has high ornamental value, and its buds are picked and dried to become vegetables with high nutritional value. Daylily usually grows buds at dusk every day, the flowers bloom the next morning and wither around noon.

According to daylily's physiological characteristic, during the flower picking season, a special person must be fixed at 5～8 a.m. for picking activities. Otherwise, nutrients will be lost after blooming, and both the quality and the price will be low. The harsh picking conditions have greatly affected the picking efficiency of daylily, and long-term picking activities will cause certain harm to the human body. Therefore, agricultural robots that automatically pick daylily have become an important and urgent need.

Regarding the research of automated daylily picking, in 2012, Shanxi Province, China, developed a riding style daylily picker. The production efficiency of this picking device is 1.45 times higher than the traditional production method, and the production cost is reduced by 25.2%. This effectively solves the problem that daylily pickers are soaked by dew and harms their bodies, significantly improves the working conditions of daylily pickers, reduces labor intensity, and has obvious economic and social benefits.

This paper designs a daylily picking robot [1]. The robot uses a camera to collect images, determines the position of the daylily, and uses a parallel robotic arm with cutting and grabbing functions to pick the daylily and put it into the designated position. The equipment will have a set of image processing algorithms based on machine learning neural networks to determine the horizontal and vertical positions of daylily [2, 3]. The algorithm can autonomously learn the characteristics of a daylily, such as color, contour shape, size, etc., and then judge whether the photographed object is a daylily based on the above features [4]. If it is daylily, then execute the picking procedure, if not, then ignore and continue to identify the next one.

Fig.1. Pick daylily

2. Robot Design

The agricultural robot designed in this paper includes a vision system, picking structure, mobile structure, microcontroller, and power system. The functional structure of the robot is shown in Fig.2.

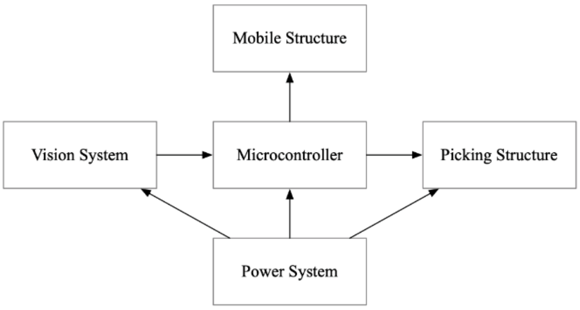


Fig.2. The functional structure of the robot

Through the vision system to identify the maturity of daylilies and the location information on the branches, the picking structure uses the recognized image information to separate the ripe daylilies from the branches. The mobile structure is responsible for the movement of the robot. The microcontroller is used to process the information sent back by the recognition module, and issue the instruction whether the picking structure picks, and how the driving mechanism moves. The power system is responsible for powering the entire robot.

2.1. Vision System

In this design, the vision system needs to meet the following requirements: 1) Can effectively identify the maturity of daylily, selective picking, to ensure the quality of picked daylily; 2) Identify the position coordinates of daylily in space, including horizontal and vertical coordinates and the distance between the camera module and daylily, assist the picking of the manipulator; 3) It can judge whether the robot is still in the field by identifying the environment, and assist the robot in walking to a certain extent.

Therefore, we choose the RGB-D binocular camera with depth function as our visual system [5].

Fig.3. The main global currents

The camera module has the following characteristics:

- Depth range: 0.6-8m
- Depth map resolution: 1280*1024 max
- Depth filed angle: 58.4*45.5cm
- Time delay: 30-45ms
- RGB: 1080P
- Connection type: USB

2.2. Picking Structure

When using a manipulator to pick daylily, it is important to ensure that it will not cause secondary damage to the crop, thus the common rigid manipulator cannot be used. Here we use an under-actuated flexible manipulator to perform the picking work. When picking, the manipulator relies on the cylinder to provide pressure to drive the fingers to close. This manipulator adopts a pneumatic drive. Compared with an electric drive, it can further adjust the gripping force of the manipulator by adjusting the pneumatic pressure to protect the daylily from being damaged [6]. The manipulator is shown in Fig.4.

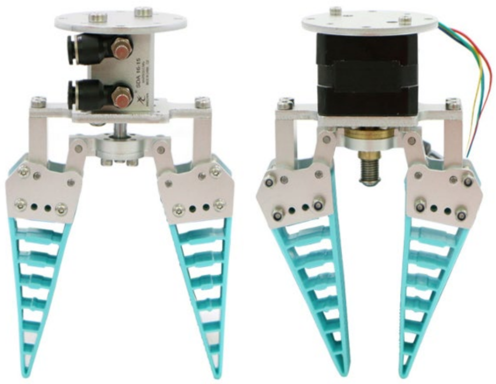


Fig.4. The manipulator

2.3. Mobile Structure

As is shown in Fig.5, the planting base of daylily is sandy land, which has a high requirement for the tire of the picking robot. If the wheel drive, it is easy for the tire to slip. And if it is cloudy and rainy, the ground will become muddy, and wheeled robots will be unable to move at all.

Fig.5. Daylily planting base

We use a crawler drive. The contact area between the crawler and the ground is large, the conflict force grip is strong, and it is not easy to sink. The crawler type can walk on any messy terrain without any influence, and can also walk normally in any muddy and soft ground. Whether it is the rainy southern part of China or the monotonous north, whether it is a mountain or a plain, whether it is sandy or muddy, the crawler drive can make the robot seem to walk on flat ground.

2.4. Microcontroller

The microcontroller adopts the design scheme of dual controller STM32-Raspberry Pi. The two controllers are shown in Fig.6 and Fig.7 respectively.

Fig.6. STM32 main global currents

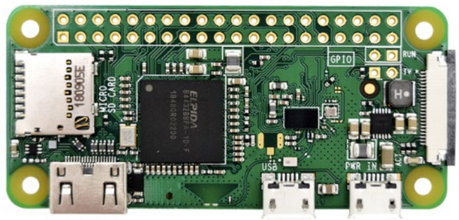


Fig.7. Raspberry Pi main global currents

STM32 microcontroller is the first choice for embedded control chips. This microcontroller has the characteristics of strong real-time, low power consumption, high integration, and a rich peripheral library, which is convenient for development. Here we use it to connect the mobile structure that drives the robot and the picking structure.

Raspberry PI internally runs a customized version of the Linux system, can be installed rich processing software, software features rich, to capture video and images can be arbitrarily modified. We use the Raspberry PI to connect to the vision system to process video information.

2.5. Power System

The power system is the power source of the whole system and is responsible for providing power for STM32, Raspberry PI, vision system, etc. The L78 series of three-terminal positive regulators (Fig.8) is available in TO-220, TO-220FP, D^{2}PAK, and DPAK packages and several fixed output voltages, making it useful in a wide range of applications [7].

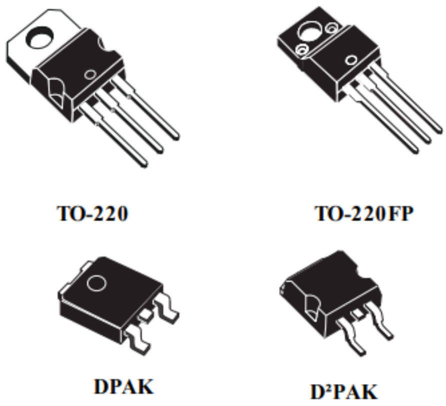


Fig.8. The L78 serial package

Different systems require different voltages, this requires the power system to be able to output different voltages stably. Therefore, L78 meets our power demand, so the L78 series is adopted as our power chip.

3. Conclusion

In this paper, we designed an agricultural picking robot of daylily according to its growth characteristics. The robot recognizes the maturity of daylily and analyzes the spatial coordinates through image recognition, and then drives the manipulator to pick. The robot can effectively improve people's labor production efficiency, and more importantly, avoid the harm caused to the human body by the daylily picking.

References

1. Zhao JC, Dai FZ, et al. Device Design Based on TDS Water Quality Detection, *International Conference on Artificial Life and Robotics*, Oita, Japan, 2020, 1: P620-623.
2. J. B. Han, K. M. Yang, et al. A Modeling and Simulation based on the Multibody Dynamics for an Autonomous Agricultural Robot. *7th International Conference on Control, Mechatronics and Automation*, Delft, Netherlands, 2019, pp. 137-143.
3. U. Rao. Design of automatic cotton picking robot with Machine vision using Image Processing algorithms. *International Conference on Control, Automation, Robotics and Embedded Systems*, Jabalpur, 2013, pp. 1-5.
4. Y. Yu, K. Zhang, et al. Real-Time Visual Localization of the Picking Points for a Ridge-Planting Strawberry Harvesting Robot. *IEEE Access*, 2020, 8: pp. 116556-116568.
5. Z. Xu, R. Jia, Y. Liu, et al. Fast Method of Detecting Tomatoes in a Complex Scene for Picking Robots. *IEEE Access*, 2020, 8: pp. 55289-55299.
6. G. Bao, P. Yao, et al. Flexible pneumatic end-effector for agricultural robot: Design & experiment. *IEEE International Conference on Robotics and Biomimetics*, Zhuhai, 2015, pp. 2175-2180.
7. A. J. Scarfe, R. C. Flemmer, et al. Development of an autonomous kiwifruit picking robot. *4th International Conference on Autonomous Robots and Agents*, Wellington, 2009, pp. 380-384.

Development of Intelligent Public Trash Can Based on Machine Vision and Learning

Longyu Gao [1], Fengzhi Dai [1, 2] *, Zhiqing Xiao [1], Jiangyu Wu [1], Zilong Liu [1]

[1] *Tianjin University of Science and Technology, China;*

[2] *Tianjin Tianke Intelligent and Manufacture Technology CO., LTD, China;*
*E-mail: * daifz@tust.edu.cn*
www.tust.edu.cn

Abstract

At present, with the maturity of machine vision technology and the continuous expansion of application fields, there have been many intelligent trash cans based on machine vision, which can realize certain garbage identification and automatic classification. However, due to certain technical limitations of machine vision, it is impossible to identify all garbage. In this paper, a smart public trash can based on machine vision and auxiliary sensors is proposed. In addition to realizing machine vision to identify and automatically classify garbage, sensors will also be used to assist in identifying garbage to solve problems such as the same garbage classification of different shapes. At the same time, enhanced learning will be added to realize the self-learning of the trash can, so as to achieve the goal of continuously increasing identifiable types.

Keywords: machine vision, machine learning, garbage classification, auxiliary sensors

1. Introduction

With the development of society and the increasing improvement of people's living standards, society is paying more and more attention to the problem of garbage classification, and various regions have also implemented relevant garbage classification policies. However, due to various types of garbage and people's poor awareness of autonomous classification, the classification effect is difficult to achieve the expected effect. And as a result, intelligent sorting trash bins based on various technologies have emerged.

Most the intelligent sorting trash can lies in automatically identifying the type of trash. At present, with the maturity of machine vision technology and a wide range of applications, an intelligent classification of trash cans based on this technology has been born, which can realize certain garbage identification and automatic classification. However, due to certain technical limitations of machine vision, it is impossible to recognize all garbage. For example, it is difficult to recognize the same garbage under different shapes.

In summary, the smart trash can not only needs to be based on machine vision, but also needs to be supplemented when machine recognition cannot be effective. Therefore, a smart public trash can based on machine vision and auxiliary sensors is proposed. In addition to realizing to recognize garbage and automatically classify it by machine vision, sensors will also be installed to assist in identifying garbage to solve problems such as the same garbage classification of different shapes. At the same time, enhanced learning will be added to realize the self-learning of the trash can, so as to achieve the goal of continuously increasing identifiable types.

2. Overall Design Scheme of the System

According to actual application scenarios and market requirements, combined with machine vision, sensor detection technology, and wireless communication technology, this paper designs a smart public trash can based on machine vision to realize automatic garbage sorting and machine learning (continuously increasing the number of identifiable types), the classification of different states of the first class of garbage and the real-time monitoring of garbage filling.

When the smart trash can detects that the rubbish is input, the trash can uses a visual camera to capture and identify the dropped rubbish [1]. Then compare and analyze the data in the database to determine the type of garbage, and then use its own mechanical structure to sort it into the corresponding garbage bin.

For the unrecognizable garbage, the trash can will work through the interaction of metal sensors, humidity sensors, VOC odor sensors and capacitive sensors to sort the garbage into the corresponding garbage cans, mainly to solve the problem of garbage occupying the identification position. At the same time, the unrecognizable garbage is photographed, uploaded to the server, and re-trained to determine the type of garbage. When the garbage is recognized again next time, the garbage classification will be performed directly through machine vision. At the same time, each trash can will share data, which can greatly improve the learning speed of the trash can and shorten the learning cycle. The flow chart of the intelligent public trash can based on machine vision is shown in Fig.1.

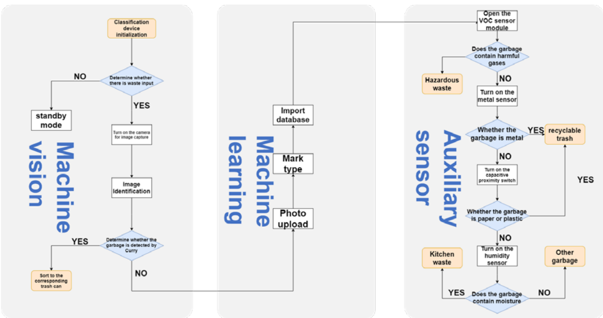


Fig.1. The Overall flow chart of intelligent public trash bin based on machine vision

3. Design of Hardware Control System

This system selects STM32F103 single-chip microcomputer as the core controller, and combines OpenMV and peripheral modules to realize the control system function. It mainly include: STM32, OpenMV, metal sensor, humidity sensor, VOC smell sensor and capacitance sensor, communication module, motor, etc. The hardware control part of the system is introduced below.

3.1. Main control chip

STM32f407zet6 is selected as the core controller of the hardware control system, which is mainly used to realize the functions of data acquisition and analysis and execution of component motion control [2].

The processor of STM32f407zet6 is Arm Cortex-M4 high-performance core, the maximum operating frequency is 168MHz, and the package form is 1qfp144. Its peripheral configuration is powerful, supporting communication interfaces such as SP interface, I2C interface, USB interface, USART interface, and peripherals such as ADC and timer. STM32f407zet6 is shown in Fig.2.

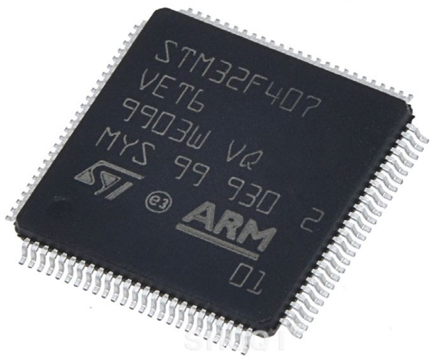


Fig.2. Main control chip STM32f407zet6

3.2. Metal sensor

The function of material recognition is realized through the metal sensor module. The system uses metal sensors to identify metals. This system uses LJ18A3-8-Z/BX type metal sensor. The sensor has the advantages of fast response speed, strong anti-interference ability, waterproof and corrosion resistance, and very stable performance.

It is widely used in automated assembly lines, intelligent hardware and security systems. The operating voltage of the sensor is 6-36V (DC), and the switch contains a high-frequency oscillation circuit, a detection circuit, an amplifying circuit, a de-transmitting circuit and an output circuit. The metal sensor is shown in Fig.3.

Fig.3. Metal sensor

3.3. Humidity Sensor

Use SHT3x temperature and humidity sensor module to communicate with STM32, and turn it on when machine vision cannot identify the garbage, so as to realize the identification of wet garbage.

The module applies digital module acquisition technology and temperature and humidity sensing technology, and the product has high reliability and long-term stability. It had the built-in humidity and temperature sensor elements, analog to digital converters, signal processing, calibration data and I2C host interface. The humidity sensor is shown in Fig.4.

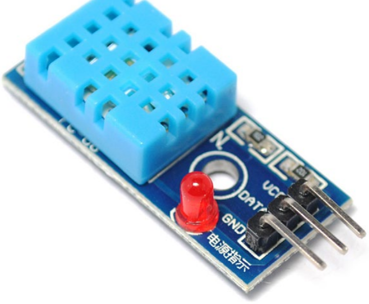


Fig.4. Humidity Sensor

3.4. Capacitive proximity switch

The second part of the material identification is mainly realized by the capacitive proximity switch M8M12. It can detect any dielectric substance, including conductors, semiconductors, insulators, and even can be used to detect liquid level and powdered materials. For non-metallic substances, the movement distance is determined by the dielectric constant of the material. The greater the dielectric constant of the material, the farther the sensing distance is.

The distance between it and the pallet is controlled by the motor to realize the classification of different materials. The capacitive proximity switch M8M12 is shown in Fig.5.

Fig.5. Capacitive proximity switch M8M12

3.5. VOC air quality sensor

The detection of hazardous waste that cannot be identified by machine vision is mainly achieved through the ZP16 digital VOC air quality sensor module. The air quality module uses advanced chip thick film semiconductor gas sensors. The gas sensors are resistant to formaldehyde, benzene, carbon monoxide, and ammonia, Hydrogen, alcohol, cigarette smoke, flavors and other organic volatile gases have extremely high sensitivity. After aging, debugging, calibration and adjusting, the module has good consistency and extremely high sensitivity. Extremely high sensitivity, excellent long-term stability, factory calibration, convenient and quick to use; with sensor fault self-diagnosis function, low power consumption and long life. The VOC air quality sensor is shown in Fig.6.

Fig.6. VOC air quality sensor

4. OpenMV

The vision part uses the openMV single-chip microcomputer to communicate with the STM32. The OpenMV camera is a small, low-power, low-cost circuit board that can be scripted by a high-level language Python (to be precise, MicroPython). We can use an external terminal to trigger shooting or execute an algorithm, or use the result of the algorithm to control the IO pin.

OpenMV uses STM32H743II ARM Cortex M7 processor, 480 MHz, 1MB RAM, 2 MB flash. All I/O pins output 3.3V and are 5V tolerant [3]. OpenMV is shown in Fig.7.

Fig.7. OpenMV

5. Overflow Reminder Module

The overflow detection uses the infrared distance measuring sensor module GP2Y0A21YK0F 10-80cm distance sensor, the integrated combination of PSD (position sensitive detector), IRED (infrared light emitting diode) and signal processing circuit, the reflectivity of various objects, The environment temperature and working time are not easily affected by the distance detection. When the garbage overflows, the return value of the module will have a big change. When the return value meets certain conditions, the main control will transmit the corresponding information to the management terminal via the WIFI module. The overflow reminder module is shown in Fig.8.

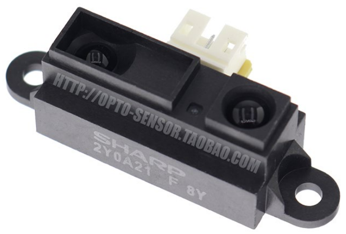


Fig.8. Overflow reminder module

6. Conclusion

Aiming at the problem of public garbage classification, this paper proposes a classification method based on the combination of machine vision and auxiliary sensors; designing an automatic garbage classification hardware system to realize the automatic classification and recycling of common garbage in public places in life. Tests show that the classification method is effective. The hardware and software system works stably. It shows that the broad application prospects of the research results of this paper in waste classification and environmental protection.

In the research process of this article, we found that there are some problems worthy of further digging. First, we can make greater improvements in the detection algorithm to improve the accuracy of small garbage recognition. Second, we can study the design of multi-bin garbage cans or other structures for garbage collection device to meet the sorting needs of more target categories.

Acknowledgements

The research is partly supported by the Project of Tianjin Enterprise Science and Technology Commissioner to Tianjin Tianke Intelligent and Manufacture Technology Co., Ltd (19JCTPJC53700). It is also supported by the Industry-University Cooperation and Education Project (201802286009) from Ministry of Education, China.

References

1. Golnabi, H., A. Asadpour. Design and application of industrial machine vision systems. *Robotics and Computer-Integrated Manufacturing*, 2007, 23(6): 630-637.
2. Liu, Jianghai, Yadong Jiang. Design of intelligent trash can be based on machine vision. *2020 International Conference on Image, Video Processing and Artificial Intelligence*. Vol. 11584. International Society for Optics and Photonics, 2020.
3. Vernon, David. Machine vision-Automated visual inspection and robot vision. STIA 92 (1991): 40499.

The Optimized Intelligent Algorithms on Face Recognition and Tracking for ROS-based Robots

Yue Chen, Shuhao Tian, Huailin Zhao, Shengyang Lu
School of Electrical and Electronic Engineering, Shanghai Institute of Technology, No. 100, Haiquan Road, Shanghai 201418, China

E-mail: 405838753@qq.com
https://www.sit.edu.cn/

Abstract

With the development of artificial intelligence, face recognition and tracking technology have been widely used in many fields such as target positioning, automatic driving, and human-computer interaction. Recently, a large number of face detection, recognition and tracking algorithms have emerged, but there are still many shortcomings in practical applications, such as slow face detection, low detection accuracy, and face recognition and tracking for ROS robots Algorithms are rare. This paper improves the traditional Haar-like algorithm and LK optical flow tracking algorithm, and designs a ROS robot platform based on the improved algorithm. By comparing the accuracy and timeliness of face detection and tracking between the improved algorithm and the traditional algorithm, the superiority of this design algorithm is obtained.

Keyword: Face Tracking and recognition; ROS robots; LK optical flow; Haar-like algorithm

1.Introduction

Video tracking[1] is an important branch of computer vision. With the increasing impact of artificial intelligence on human production and life, video tracking is widely used in many fields such as target positioning, video tracking, automatic driving, and human-computer interaction. Recognition and tracking of mobile faces is an important technology to improve the intelligent level of mobile robots. Researchers have developed a variety of computer face recognition and tracking algorithms. The ultimate goal is to solve noise interference, target occlusion of rapid movement and the surrounding environment More complex issues and real-time tracking. With the continuous development of computer technology, a large number of algorithms for real-time video tracking have emerged, mainly including three types of algorithms based on target external characteristics, target contours, and target local areas. Face detection and tracking is an extremely important research field in video tracking technology, and it is the basis of many face applications. Therefore, research, processing and tracking of faces is a very meaningful direction in face video surveillance. However, in practical applications, face detection technology still has some shortcomings, such as slow face detection speed, low accuracy of face detection, and face recognition and tracking systems for ROS robots are rare. This paper designs a set of face recognition and tracking system suitable for ROS mobile robots to improve the intelligence of ROS robots.

Ebied [2] proposed the application of Principal Component Analysis (PCA) in face feature extraction and dimensionality reduction to solve the problem of reduced face recognition rate caused by the large dimension of face features. This method applies mirroring technology to the parity analysis of feature maps. The final simulation results show that this method effectively improves the accuracy of face detection and face recognition. But the algorithm runs slower and has higher requirements on the training set. TFCootes[3] and his research team proposed an Active Shape Model (ASM) based on statistical parameterization. This method builds a statistical parameterized model based on the statistical analysis results of the target contour feature vector, which can reflect Draw out the law of target shape change. Because of its simplicity and high efficiency, the algorithm has been well applied in face detection and tracking. Dejun Tang from the School of Computer Science and Technology of Dalian Maritime University[4] studied the image feature extraction and matching technology in face recognition, using the improved Singular Value Decomposition (SVD) method of class estimation space for feature extraction to solve face recognition In the process, the singular value vector is directly used as the face feature, which leads to the problem of low recognition rate. In addition, in order to reduce the instability of face recognition caused by changes in external factors such as illumination, expression, noise, posture and other factors, a face feature representation method that combines multi-scale global features and local features is proposed, and then rough integration is used. The simplified algorithm performs feature selection on the extracted features. Through the above overview and analysis of existing face detection and face recognition algorithms, the existing methods have achieved certain success in face recognition and detection in complex environments and occlusions, but there is still much room for improvement. It is very important to further improve the accuracy and robustness of face recognition and detection under complex conditions.

In the process of realizing mobile face recognition and tracking, there are mainly three issues involved: 1) Recognizing the face, 2) Marking the key points of the face, 3) Detecting the key points before and after the face moves, and judging whether it is the same face. Achieve correct tracking. Therefore, after analyzing and comparing most face recognition algorithms, this design improves the Haar-like[5] face recognition algorithm to improve the accuracy and speed of face recognition. Then use the improved LK optical flow tracking algorithm[6] to track a specific moving face by detecting the key points before and after the face moves. Finally, the Camshift[7] color tracking algorithm is used for face tracking, and the improved algorithm is compared to realize the face recognition and tracking of the ROS robot. The main innovations of this article are as follows:

(i) When marking the key points of the face, we have improved on the basis of the Harris algorithm for corner detection and adopted a more accurate marking method, which is called the better key point marking algorithm.
(ii) This paper uses an improved LK optical flow tracking algorithm to detect the key points before and after the face moves, so as to realize the tracking of a specific moving face. Reduce the impact of the surrounding environment on face tracking.
(iii) As a control, we also applied the Camshift color tracking algorithm to research and experiment, and compared and analyzed the experimental results of the LK optical flow tracking algorithm.
(iv) This paper designs and implements the face recognition and tracking system of the ROS mobile robot platform, on the ROS robot platform, which can meet the accuracy and real-time performance of face recognition and tracking in complex environments through real environment tests.

2. Face recognition and tracking algorithm design

Taking into account the recognition speed, accuracy and compatibility with the robot system, the face recognition and tracking design system is mainly composed of Haar-like feature recognition, better key point marking and adding, and LK optical flow tracking. We call it the LK optical flow tracking scheme.

2.1 Face recognition based on Haar-like detector

Haar-like classifier detection target is to use the features of known training samples to compare features with unknown images to achieve target detection. This article is mainly used in face detection, generally using two rectangular features, by making the difference between the white and black area pixels to meet the detection needs. The Haar-like detector is composed of Haar-like features, Haar integral map, and cascade classifier.

Haar-like features can be roughly divided into three types: edge features, linear features, central features and diagonal features. According to the above-mentioned feature group, the Haar-like detector consists of black and white rectangles as feature templates, and its feature value is defined as the difference between the white rectangular pixels and the black rectangular pixels. The Haar-like detector uses the feature template to arbitrarily change its specifications in the image, process the image, and use the exhaustive method to calculate the feature value of the face area and the feature value of the non-face area, and judge the image gray by the feature value Degree changes and detect target recognition. The application effect is shown in Fig.1:

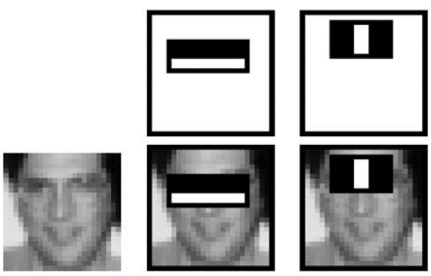


Fig.1 Haar module face recognition application

The Haar-like detector uses a sliding window to detect the area of the face. In order to calculate the pixel difference between the feature maps, OpenCV is used to grayscale the RGB feature maps detected by the Haar detector and normalize them. The pixel value of each sliding window is grayed out. For each pixel $i(x,y)$, the average gray value of the pixel value $mean$ and the square average gray value sq_{mean} are calculated as follows:

$$mean = \frac{\sum_{x=0}^{w}\sum_{y=0}^{h} i(x,y)}{w \bullet h} \tag{1}$$

$$sq_{mean} = \frac{\sum_{x=0}^{w}\sum_{y=0}^{h} i^2(x,y)}{w \bullet h} \tag{2}$$

Then the normalization factor (varNormFactor) and the normalized haar feature value (normValue) are obtained by normalizing the image. The calculation formula is as follows:

$$\mathrm{var}\,NormFactor = \sqrt{sq_{mean} - mean^2} \tag{3}$$

$$normValue = \frac{featureValue}{\mathrm{var}\,NormaFactor} \tag{4}$$

Where $featureValue$ is the feature value of each pixel in the feature map. In order to improve the effect of face detection, this paper adopts a cascade classifier called $Adaboost$, which is achieved by cascading multiple weak classifiers to form a strong classifier. The specific algorithm of the $Adaboost$ cascade classifier is as follows:

(a) The picture to be input is preprocessed to obtain the characteristic value corresponding to each pixel of the picture.

(b) Use the Haar classifier to calculate the Haar feature value ($normValue$) corresponding to the feature value and compare it with the No. 1 face feature. If not, go to step(f), if yes, go to step(c).

(c) Compare the Haar feature value ($normValue$) with the No. 2 face feature, if it does not match, go to step(f), if it matches, go to step(d).

(d) Compare the Haar feature value ($normValue$) with the No. 3 face feature, if it does not match, go to step(f), if it matches, go to step(e).

(e) Confirm that it is a human face.

(f) Confirm that it is not a human face.

The specific algorithm flow chart is shown in Fig. 2:

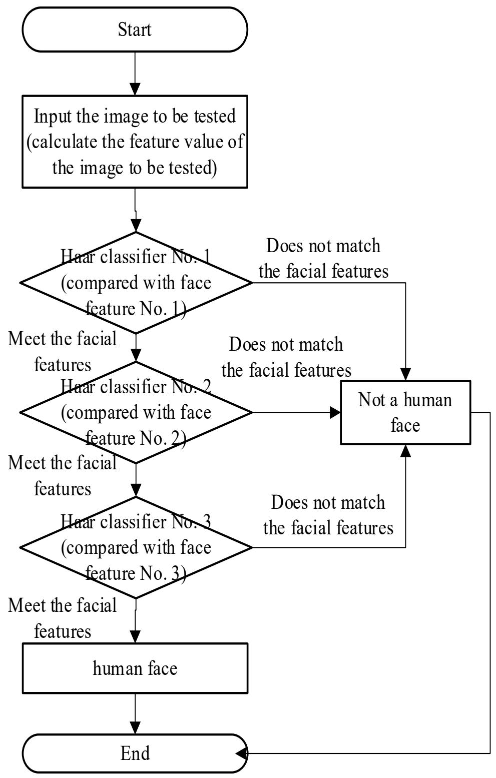


Fig.2 Flow chart of cascade classifier judgment

According to the results of face detection at each level, the image area similar to the face is screened to accurately locate the face. The scalefactor variable parameter is introduced to set the window size. The smaller the value of this parameter, the smaller the size of the sliding window and the more detailed the scan of the face, but it will take more time and memory to scan.

2.2 Better key point marking algorithm

The better key point marking algorithm mainly uses the gray-scale change relationship between the pixels between the marked corner points. The corner points are mainly the connecting points of the contour lines of the objects in the image, as shown in Fig.3. Use a fixed window to slide in any direction on the image to be tested, and compare the degree of pixel grayscale changes in the window before and after the slide. If there is a large gray scale change, then it is determined that the window has corners.

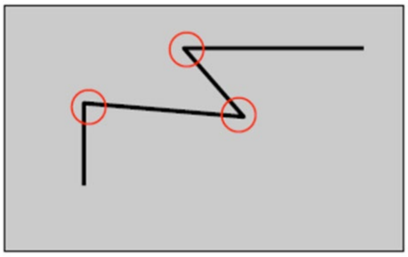


Fig.3 Corner detection

Use corner detection[9] scoring function for better key point marking algorithm, the specific scoring function is as follows

$$R = \min(\lambda_1, \lambda_2) \tag{5}$$

Manually set two feature values λ_1 and λ_2. When the score of the picture to be detected exceeds the threshold values λ_1 min and λ_2 min, it will be judged as a corner point. The specific application of ST scoring function is shown in Fig.4.

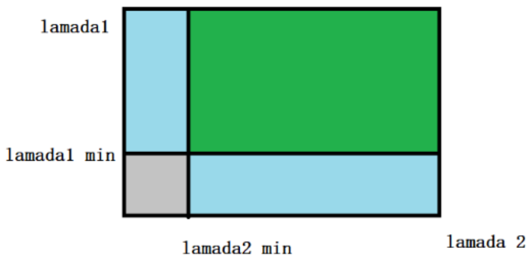


Fig.4 ST scoring function to judge corner points

After multiple tests, the appropriate number of corner points is selected, and the selection of key points is determined by the number of corner points, so that the face can be more accurately recognized and detected.

2.3 LK optical flow tracking algorithm

After extracting the key point features, the LK optical flow tracking algorithm [11] is used for real-time face tracking.

We assume:

(1) The brightness is constant;

(2) The time is continuous or the motion is "differential motion" and the space is consistent;

(3) The adjacent pixels remain adjacent and have similar motions.

According to the above three assumptions, there is a picture at time t, the internal coordinates of the picture are a point at the position (x, y), and the pixel gray level is I (x, y, t); according to the assumption (2) after differential motion , That is, after moving (dx, dy) distance after dt, its gray value is I (x+dx, y+dy, t+dt); finally, according to assumption (1), the previous t time and t+dt The gray scale is the same at all times. The mathematical expression is as follows:

$$I(x, y, t) = I(x+dx, y+dy, t+dt) \quad (6)$$

After differential transformation and equation transformation, the following formula is obtained:

$$\begin{bmatrix} \left[\dfrac{dI_1}{dx}, \dfrac{dI_1}{dy}\right] \\ \dots\dots\dots\dots\dots \\ \left[\dfrac{dI_k}{dx}, \dfrac{dI_k}{dy}\right] \end{bmatrix} \begin{bmatrix} \dfrac{dx}{dt} \\ \dfrac{dy}{dt} \end{bmatrix} = \begin{bmatrix} \dfrac{dI_1}{dx} \\ \dots\dots \\ \dfrac{dI_k}{dx} \end{bmatrix} \quad (7)$$

Among them, dI/dt represents the gray changes around t and $t+dt$ at the two moments, dx/dt and dy/dt , represents the movement speed of the point in the axis x、y direction, and dI/dx and dI/dy represents the gray changes in the axis x、y direction.

Although the equation cannot be solved based on the position information of a single key point, the position of the key point in the picture after the elapse of time dt can be obtained by combining multiple key point information and hypothesis (3).

In the algorithm design, this design takes LK optical flow tracking as an extension of the better key point. Therefore, the LK optical flow tracking node can be used to subscribe to the key points that have been obtained in the better key point node, and the algorithm can be used to pass the current key point. And two grayscale pictures to predict the next set of key points, and perform backward prediction.

2.4. Optimal design of LK optical flow tracking algorithm

In the process of key point tracking, since the key points of the tracking are judged by the ST scoring function to have low scores, these key points are automatically discarded by the optical flow tracker. This phenomenon leads to the loss of faces during the tracking process. And misjudgment. Therefore, the algorithm is improved and optimized for the above problems. In the process of tracking, use the tracked results to make a statistical set, search for each frame in the video tracking process, and count, delete the abnormal feature points, update the statistical set, and search for the next frame. The specific flow chart is as follows:

(a) Make a data set based on the data points obtained by searching for the first k frames;
(b) Search for k+1 frames, compare and distinguish the feature points. If it is determined to be an abnormal point, go to step(c). If it is not an abnormal point, go to step(d);
(c) Remove anomalies;
(d) Keep feature points;
(e) Judge whether it is the last frame, if it is the end of the last frame, if it is not the last frame, go to step(f);
(f) Update the statistics set.

Based on the above algorithm, the detection accuracy of the LK optical flow algorithm can be improved, and the impact of this method on the complex environment can also be minimized. In order to improve the accuracy of the recognition and tracking algorithm, this algorithm is used to improve the LK optical flow algorithm. Although this algorithm has a certain increase in the amount of calculation, the accuracy and robustness have been improved a lot. The specific algorithm flow chart is shown in Fig. 5 below:

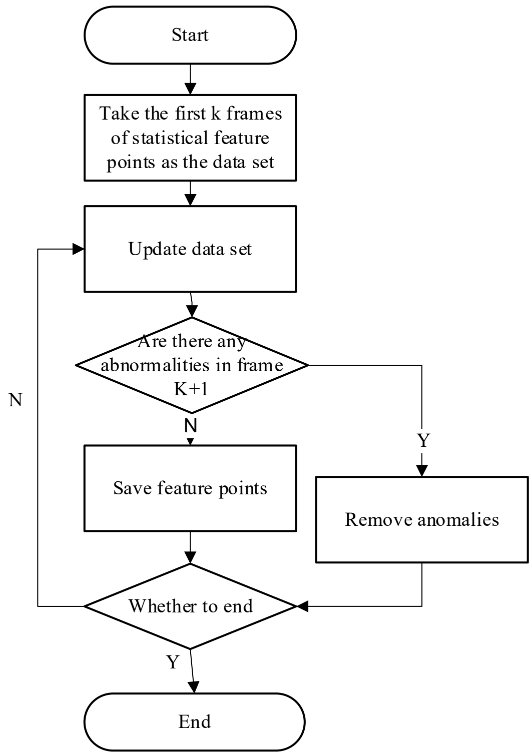


Fig.5 Optical flow statistics algorithm flow chart

3. Design of ROS robot platform based on face recognition and tracking algorithm

The experimental environment is Ubuntu system, which is realized by programming under the terminal. The hardware platform is Turtlebot2 robot, Kinect camera and host computer. After programming the upper computer loaded with the Ubuntu system and completing the basic face recognition and tracking algorithm design, the corresponding Python program was designed and tested on the ROS robot Turtlebot2.

3.1 Feature point tracking test

Process the video information received by Kinect, recognize the face and track the face in real time, and finally achieve the purpose of face recognition and tracking in a complex environment. Fig. 6 is the hardware platform built for this experiment.

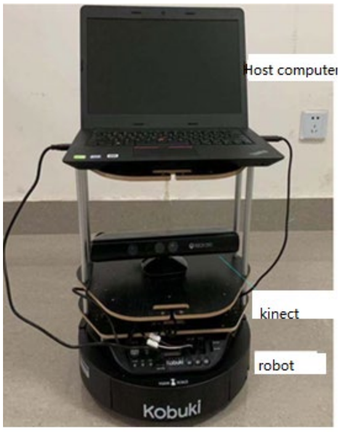


Fig.6 The mobile robot Turtlebot2

The system uses the traditional Harr-like algorithm and LK optical flow algorithm to meet the basic requirements of face recognition and tracking, but there are some problems. As shown in Fig. 7, when the face recognition and tracking program runs for a certain period of time, some of the key points will drift from the ROI area (the face recognition area) to other non-target areas. And as the running time goes by, the number of key points in the recognition screen is significantly reduced, until a few key points are left in the face area, which cannot allow the robot to track well.

Fig.7(a) 103th frame   Fig.7(b) 107th frame

Fig.7 Face tracking effect before optimization

In response to the above problems, we modified and optimized the algorithm. Fig.8 uses the algorithm proposed in this article. It can be seen from Fig.8 that with the increase of time, the size and shape of the frame tracked by the face in the process of continuous movement remain unchanged, and the tracked feature points are also relatively increased and successfully completed the tracking. It can be seen that compared with the traditional algorithms, the algorithm proposed in this paper greatly reduces background information, makes full use of target information, and effectively reduces false feature points. Not only that, it also ensures that the number of feature points makes the tracking accuracy and robustness greatly improved.

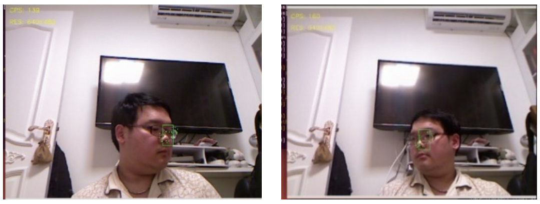


Fig.8(a) 139th frame   Fig.8(b) 160th frame

Fig.8 Face tracking effect after optimization

As shown in Fig.8 above, the improved face recognition and tracking algorithm in the ROS robot platform has a greatly improved tracking effect compared with the previous one.

3.2 Face recognition test

The ROS robot platform designed this time can achieve the tracking test of the human face, that is, when the human face is moving, the robot also follows the human face to rotate and detect and recognize the human face. In order to verify the accuracy of the face recognition of the robot in the process of moving, the vertical upward is adopted, and the left and right swings at the same speed are used to measure the maximum detectable angle $0°$ of the robot. As shown in Fig.

9 below:

Fig.9(a) 0°　　Fig.9(b) 15°

Fig.9 Swing tracking process

From the above figure, it can be seen that when the human face rotates at a certain angle, the robot is also rotating with the human face. In order to obtain the tracking range of the robot, the human rotation angle is used as a reference for comparative analysis. In this experiment, the traditional Haar-like+optical flow tracking algorithm is compared with the improved Haar-like+optical flow tracking algorithm and the Camshift algorithm, and the superiority of this algorithm is obtained. The comparison is shown in Table 1:

Table 1 Algorithm effect comparison table

Algorithm name	Face swing angle	Recognition speed(Frames/sec)
Traditional Haar-like + optical flow tracking	±10°~±15°	92
Camshift tracking algorithm	±24°~±26°	72
Ours	±26°~±28°	87

From the above table, it can be found that the detection angle of the algorithm in this paper has been greatly improved compared with the traditional Haar-like+optical flow tracking algorithm, but the detection speed has been reduced due to the increase in calculation. But compared to the Camshift tracking algorithm, both the accuracy and the detection speed are improved. The algorithm in this paper has a good effect.

4.Conclusion

This paper improves the traditional Haar-like algorithm and optical flow tracking, and designs and optimizes a set of face recognition and tracking system based on ROS mobile robot. The superiority of the improved algorithm is obtained by comparing with the traditional algorithm. The specific work is summarized as follows:

1. Implement and improve the traditional Haar-like algorithm and LK optical flow tracking algorithm.

2. Designed a face tracking platform based on the ROS system, and compared the improved algorithm with the traditional algorithm, and got the following conclusions:

A. The processing accuracy of the traditional Haar-like algorithm+optical flow tracking algorithm is very different from the algorithm in this paper, but the speed is faster, and the Camshift algorithm is not as good as the algorithm in both processing accuracy and speed.

B. The algorithm in this paper is applied to the ROS robot platform, and the accuracy of robot face recognition is greatly improved compared with traditional algorithms.

5. References

1. Schnell Anna Elisabeth,Van den Bergh Gert,Vermaercke Ben,Gijbels Kim,Bossens Christophe,de Beeck Hans Op. Face categorization and behavioral templates in rats[J]. Journal of vision,2019,19(14).
2. Chen Jeng. A new patch-based LBP with adaptive weights for gender classification of human face[J]. Journal of the Chinese Institute of Engineers,2020,43(5).
3. Zhang G, Vela P A . Good features to track for visual SLAM[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2015.
4. Xiaochen L, Xiaojie L, Yufeng X, et al. Evaluation of Lucas-Kanade based optical flow algorithm[C]// 2018 IEEE CSAA Guidance, Navigation and Control Conference (GNCC). IEEE, 2018.
5. Li Wang, Huailin Zhao, Yaoyao Li. Research on Research on the multi-scale network crowd density estimation algorithm based on attention mechanism. The 2019 International Conference on Intelligent Informatics and BioMedical Science (ICIIBMS2019), Nov.21-24, 2019, Shanghai, China, pp272-278.
6. Yawei Hou, Huailin Zhao. Handwritten Digit Recognition Based on Depth Neural Network. 2017 International Conference on Intelligent

Informatics and BioMedical Science (ICIIBMS2017), Nov.24-26, 2017, Okinawa, Japan. pp35-38. (EI)

7. Shunzhou Wang, Huailin Zhao, Weiren Wang, Huijun Di, Xueming Shu. Improving Deep Crowd Density Estimation via Pre-Classification of Density. The 24th International Conference on Neural Information Processing (ICONIP2017), Nov.14-18, 2017, Guangzhou, China. pp260-269. (EI)
8. 8.Muhammad Asif, Nadeem Haider, Qasem Al-Mdallal, et al. A Haar wavelet collocation approach for solving one and two-dimensional second-order linear and nonlinear hyperbolic telegraph equations. 2020, 36(6):1962-1981
9. Multimedia; New Multimedia Study Findings Reported from National Taipei University (Fast Iris Localization Using Haar-like Features and Adaboost Algorithm). 2020,
10. Xirui Ma, Yizhou Lin, Zhenhua Nie et al. Structural damage identification based on unsupervised feature-extraction via Variational Auto-encoder[J] Measurement, 2020, 160
11. Guangqiang Li, Lei Yu, Shumin Fei. A Binocular MSCKF-Based Visual Inertial Odometry System Using LK Optical Flow. 2020, :1-16.
12. Multimedia; Investigators at Jiangsu University Describe Findings in Multimedia (Fast Stereo Visual Odometry Based On Lk Optical Flow and Orb-slam2). 2020.
13. Timotej Gašpar, Miha Deniša, Primož Radanovič, et al. Smart hardware integration with advanced robot programming technologies for efficient reconfiguration of robot workcells. 2020, 66
14. 고동현, 이재현, 김정인, et al. Integration of Mobile Robot and Manipulator Using the ROS. 2020, 29(3):182-188.
15. Iván del Pino, Miguel Á. Muñoz-Bañon, Saúl Cova-Rocamora, et al. Deeper in BLUE Development of a roBot for Localization in Unstructured Environments. 2020, 98(6):207-225.
16. Enrique Coronado, Gentiane Venture. Towards IoT-Aided Human–Robot Interaction Using NEP and ROS: A Platform-Independent, Accessible and Distributed Approach. 2020, 20(5)
17. Umberto Maniscalco, Antonio Messina, Pietro Storniolo. ASS4HR — An Artificial Somatosensory System for a Humanoid Robot. The ROS package. 2020, 11
18. Optical Research; New Data from China Nuclear Power Engineering Co. Ltd. Illuminate Findings in Optical Research (High Precision Indoor Positioning Method Based On Visible Light Communication Using Improved Camshift Tracking Algorithm). 2020, :337-.
19. Optical Research; New Data from China Nuclear Power Engineering Co. Ltd. Illuminate Findings in Optical Research (High Precision Indoor Positioning Method Based On Visible Light Communication Using Improved Camshift Tracking Algorithm). 2020,
20. Wanchao Mao, Hongyun Xie, Zequn Tan, et al. High precision indoor positioning method based on visible light communication using improved Camshift tracking algorithm. 2020, 468
21. Zhang Xiangnan, Gong Wenwen, He Qifeng, et al. Camshift tracking method based on correlation probability graph for model pig. 2020, 2020(1)
22. Gao Chaoqian, Chen Hu, Li Tianping. Target Tracking Based on Camshift Algorithm and Multi. 2020, 1550(3)
23. 양병호, 이명진. Study on the Video Stabilizer based on a Triplet CNN and Training Dataset Synthesis. 2020, 25(3):428-438.Ruiheng Zhang, Lingxiang Wu, Yukun Yang, et al.
24. Multi-camera multi-player tracking with deep player identification in sports video. 2020, 1020tabek Khudayberdiev, Muhammad Hassaan Farooq Butt.
25. Fire detection in Surveillance Videos using a combination with PCA and CNN. 2020, 3(3).

Control of a Novel 5D Hyperchaotic System

Qiang Wei
Army Military Transportation University, China

Hong Niu
College of Electronic Information and Automation,
Tianjin University of Science & Technology, China
E-mail: spots@163.com
www.tust.edu.cn

Abstract

In this paper, a novel five-dimensional (5D) autonomous hyperchaotic system is presented, and the characteristics of the 5D system are given in brief. For control of the 5D hyperchaotic system, a linear feedback controller is designed via the Lyapunov stability theory, so that the 5D system is no longer hyperchaotic but globally asymptotically converges to the equilibrium point at the origin. The numerical simulation results are given to illustrate the feasibility and effectiveness of the method.

Keywords: the novel 5D hyperchaotic system; hyperchaos control; Lyapunov stability theory; global asymptotic stability

1. Introduction

Hyperchaos was first proposed by Otto Rössler in 1979.[1] Since then, many novel hyperchaotic systems have been formulated. In order to obtain hyperchaos, the system need to satisfy the following two important requisites. Firstly, the minimal dimension of the phase space that embeds a hyperchaotic attractor should be at least four, which requires the minimum number of coupled first-order autonomous ordinary differential equations to be four. Secondly, the number of terms in the coupled equations giving rise to instability should be at least two, of which at least one should have a nonlinear function.[2] Therefore, hyperchaos is much more complicated than chaos, and it has greater engineering significance and application prospect in signal processing, secure communication and so on.

In this paper, a novel 5D hyperchaotic system, which has been introduced in Ref. 3, is reviewed. Stability control of the 5D system would be discussed, and some simulation results would be given to demonstrate the validity of the designed linear feedback controller.

2. The Novel 5D Hyperchaotic System

The dynamic equations of the novel 5D hyperchaotic system are formulated as

$$\begin{aligned}
&\dot{x} = a\left(y - x\right),\\
&\dot{y} = \left(c - a\right)x + cy + w - xz,\\
&\dot{z} = -bz + xy,\\
&\dot{v} = mw,\\
&\dot{w} = -y - hv,
\end{aligned} \tag{1}$$

where $x, y, z, v, w \in R$ are state variables, and a = 23, b = 3, c = 18, m = 12 and h = 4.[3]

Let the initial values of the 5D system (1) be (x_0, y_0, z_0, v_0, w_0) = (1, 1, 1, 1, 1), then the Lyapunov exponents respectively are $\lambda_1 = 0.8732 > 0$, $\lambda_2 = 0.1282 > 0$, $\lambda_3 = -0.0013 \approx 0$, $\lambda_4 = -0.5770 < 0$ and $\lambda_5 = -8.4231 < 0$. It indicates that the 5D system (1) is hyperchaotic. The phase portraits of the 5D hyperchaotic system (1) are shown in Fig. 1.

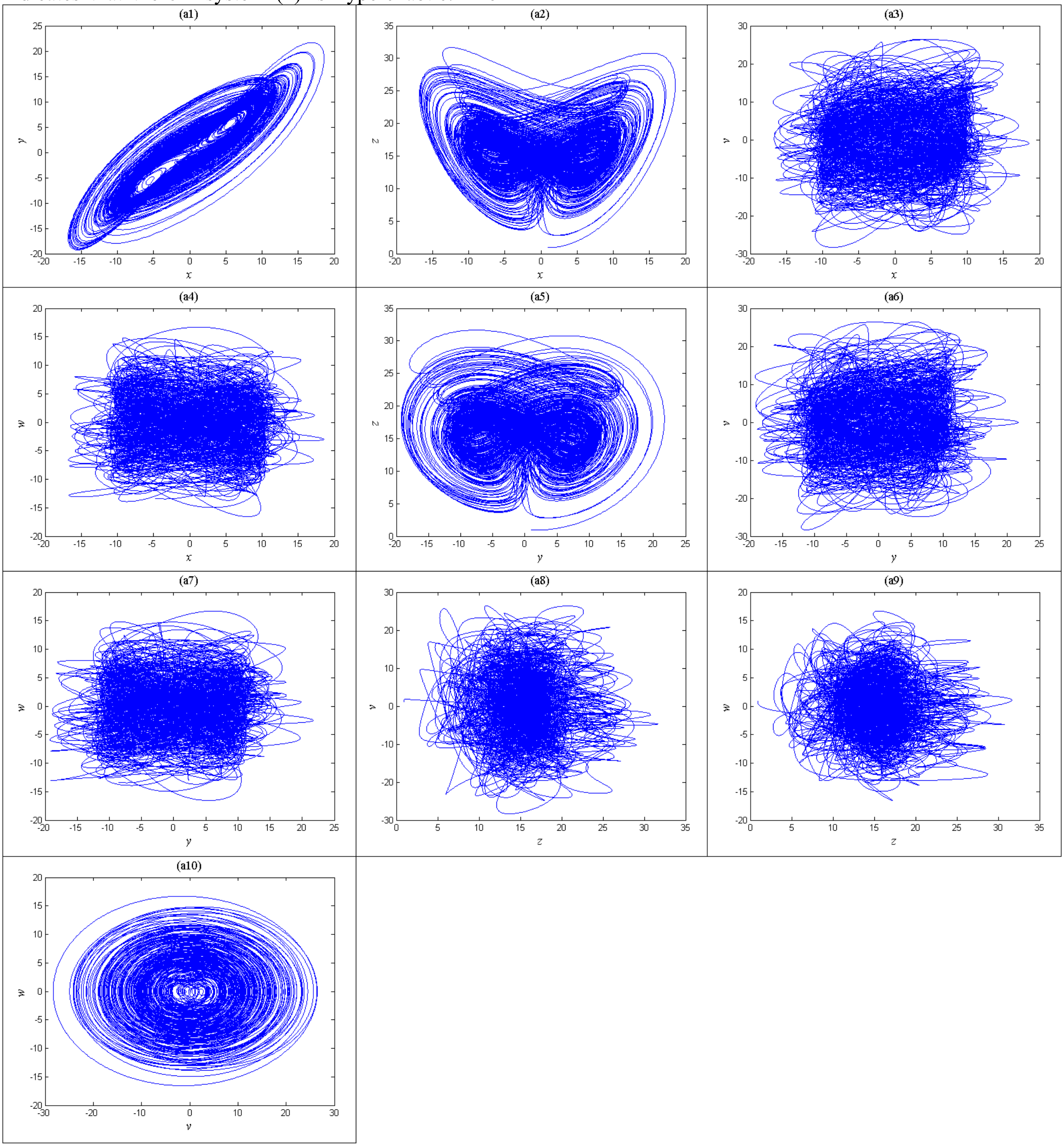


Fig. 1. Phase portraits of the 5D hyperchaotic system: (a1) *x-y*; (a2) *x-z*; (a3) *x-v*; (a4) *x-w*; (a5) *y-z*; (a6) *y-v*; (a7) *y-w*; (a8) *z-v*; (a9) *z-w*; (a10) *v-w*

3. Hyperchaos Control of the 5D System

3.1. *Formulation of the controlled system*

The controlled system is represented as

$$\begin{aligned} \dot{x} &= a(y-x)+u_{c1}, \\ \dot{y} &= (c-a)x+cy+w-xz+u_{c2}, \\ \dot{z} &= -bz+xy+u_{c3}, \\ \dot{v} &= mw+u_{c4}, \\ \dot{w} &= -y-hv+u_{c5}, \end{aligned} \qquad (2)$$

where

$$\begin{aligned} \boldsymbol{u}_c &= \begin{bmatrix} u_{c1} & u_{c2} & u_{c3} & u_{c4} & u_{c5} \end{bmatrix}^{\mathrm{T}} \\ &= \begin{bmatrix} -k_1 x & -k_2 y & -k_3 z & -k_4 v & -k_5 w \end{bmatrix}^{\mathrm{T}}, \end{aligned}$$

and $k_1, k_2, k_3, k_4, k_5 \geq 0$.

3.2. *Design of the linear feedback controller*

Theorem 1. *Let* $\boldsymbol{x} = \boldsymbol{0}$ *be an equilibrium point for* $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x})$, *where* $\boldsymbol{f}: D \to R^n$ *is a locally Lipschitz map from a domain* $D \subset R^n$ *into* R^n. *Let* $V: R^n \to R$ *be a continuously differentiable function such that*

$$V(\boldsymbol{0}) = 0 \text{ and } V(\boldsymbol{x}) > 0, \ \forall \boldsymbol{x} \neq \boldsymbol{0}$$

$$\|\boldsymbol{x}\| \to \infty \Rightarrow V(\boldsymbol{x}) \to \infty$$

$$\dot{V}(\boldsymbol{x}) < 0, \ \forall \boldsymbol{x} \neq \boldsymbol{0}$$

then $\boldsymbol{x} = \boldsymbol{0}$ *is globally asymptotically stable.*[4]

From Theorem 1, take a continuously differentiable function

$$V = \frac{1}{2}\left(x^2 + y^2 + z^2 + \frac{h}{m} v^2 + w^2 \right)$$

as a Lyapunov function candidate for the controlled system (2). Then, the derivative $\dot{V}$ is given by

$$\begin{aligned} \dot{V} &= x\dot{x} + y\dot{y} + z\dot{z} + \frac{h}{m} v\dot{v} + w\dot{w} \\ &= -(k_1+a)x^2 + cxy - (k_2-c)y^2 \\ &\quad -(k_3+b)z^2 - k_4\frac{h}{m}v^2 - k_5 w^2 \\ &\leq -\left(k_1 + a - \frac{c}{2}\right)x^2 - \left(k_2 - \frac{3}{2}c\right)y^2 \\ &\quad -(k_3+b)z^2 - k_4\frac{h}{m}v^2 - k_5 w^2. \end{aligned}$$

For $\dot{V} < 0$, the parameters k_1, k_2, k_3, k_4 and k_5 should satisfy that

$$\begin{aligned} k_1 + a - \frac{c}{2} &> 0, \\ k_2 - \frac{3}{2}c &> 0, \\ k_3 + b &> 0, \\ k_4 \frac{h}{m} &> 0, \\ k_5 &> 0, \end{aligned} \Rightarrow \begin{aligned} k_1 &> \frac{c}{2} - a, \\ k_2 &> \frac{3}{2}c, \\ k_3 &> -b, \\ k_4 &> 0, \\ k_5 &> 0, \end{aligned} \Rightarrow \begin{aligned} k_1 &= 0, \\ k_2 &= 30, \\ k_3 &= 0, \\ k_4 &= 1, \\ k_5 &= 1. \end{aligned}$$

As a result, the linear feedback controller $\boldsymbol{u}_c$ can be designed as

$$\begin{aligned} \boldsymbol{u}_c &= \begin{bmatrix} u_{c1} & u_{c2} & u_{c3} & u_{c4} & u_{c5} \end{bmatrix}^{\mathrm{T}} \\ &= \begin{bmatrix} 0 & -30y & 0 & -v & -w \end{bmatrix}^{\mathrm{T}}. \end{aligned}$$

From Theorem 1, the controlled system (2) is globally asymptotically stable at the origin.

3.3. *Numerical simulation*

Let the initial values still be $(x_0, y_0, z_0, v_0, w_0) = (1, 1, 1, 1, 1)$, then the curves of the state variables of the controlled system (2) before and after adding the controller $\boldsymbol{u}_c$ are shown in Fig. 2 and Fig. 3 respectively. Comparing Fig. 3 with Fig. 2, it can be found that the state variables x, y, z, v and w converge to zero asymptotically and rapidly. It implies that the controlled system (2) is no longer hyperchaotic but asymptotically stable at the origin. It illustrates that the linear feedback controller $\boldsymbol{u}_c$ is feasible and effective for hyperchaos control of the 5D system (2).

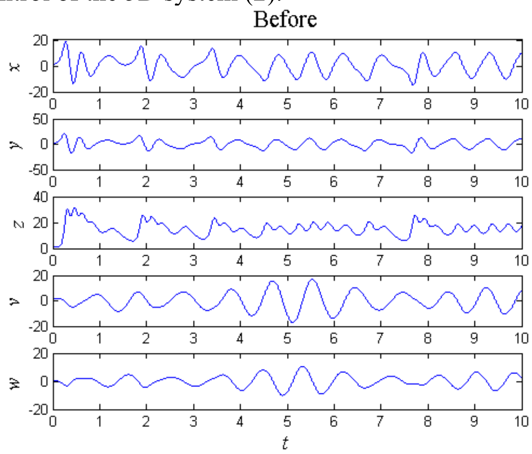


Fig. 2. Curves of the state variables of the controlled system before adding the controller $\boldsymbol{u}_c$

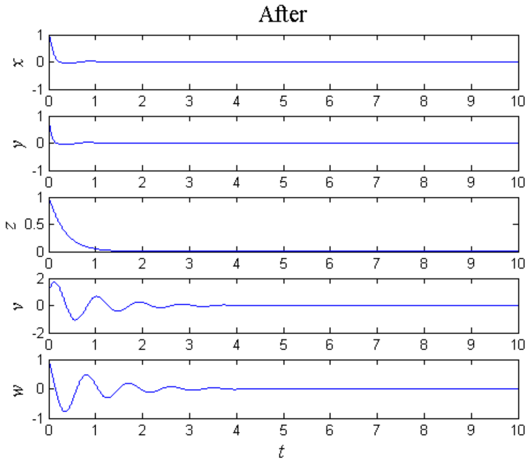


Fig. 3. Curves of the state variables of the controlled system after adding the controller $\boldsymbol{u}_c$

4. Conclusions

In this paper, a novel 5D hyperchaotic system is reviewed. For hyperchaos control of the 5D system, a linear feedback controller is designed. The numerical simulation results demonstrate the validity of the controller. The study in this paper has some engineering significance.

References

1. O. E. Rössler, An equation for hyperchaos, *Physics Letters A* **71**(2–3), 1979, pp. 155–157.
2. T. G. Gao, G. R. Chen, Z. Q. Chen and S. J. Cang, The generation and circuit implementation of a new hyperchaos based upon Lorenz system, *Physics Letters A* **361**(1–2), 2007, pp. 78–86.
3. Q. Wei and H. Niu, Analysis and circuit design of a novel 5D hyperchaotic system, *Dynamical Systems and Control* **8**(2), 2019, pp. 118–128.
4. H. K. Khalil, *Nonlinear Systems*, 3rd edn., Prentice Hall, New Jersey, 2002, pp. 112–126.

Design of Multifunctional Vehicle Interior Environment Monitoring System Based on Wireless Communication

Yuqi Yan *, Jialin Yang , Zhongxu Qin

Wuhan University of Technology, China

*E-mail: * 1814696080@qq.com*

www.whut.edu.cn

Abstract

The interior environment of the car affects the driver's mental state to a certain extent, so it is necessary to design a multi-functional interior environment monitoring system. In this paper, STM32F103 single chip microcomputer is used as the core controller, which integrates multiple sensors to collect various environmental information in the vehicle, and transmits the data to the monitoring platform through WiFi wireless communication. When the detected data exceeds the preset threshold, the monitoring system will generate an alarm to remind the user to carry out relevant operations. The system has the advantages of low cost, convenient use and high precision, and has certain commercial value and market prospect.

Keywords: car, sensors, STM32F103, WiFi

1. Introduction

At present, the car is an indispensable means of transportation in daily life. The internal environment of the car affects the physical condition of the driver and passengers to a certain extent. Therefore, it is necessary to design a multifunctional vehicle interior environment monitoring system. The multi-function internal environment monitoring system can collect the temperature, humidity, CO concentration, PM2.5 concentration and other internal environment information in the car in real time, and display them on the LCD screen. When the detected data exceeds the preset threshold, the monitoring system will generate an alarm to remind the user to carry out corresponding operation. The system can also transmit various data information to the monitoring platform through WiFi wireless communication, and users can view the historical data information on the monitoring platform. The vehicle interior environment monitoring system can effectively monitor and alarm the vehicle interior environment information in real time. The system is convenient to use and has a certain market value.

2. Main control chip and peripheral circuit

The overall hardware structure of the system is composed of main control chip and peripheral modules, as shown in Fig.1. The multifunctional internal environment monitoring system selects STM32F103VET6 single-chip computer as the core controller of the hardware control system. The processor of STM32F103VET6 is ARM Cortex-M3 high performance kernel, with a maximum operating frequency of 72Mhz, Flash memory of 512 KB, SRAM of 64KB, and encapsulated in LQFP64[1]. Its peripheral configuration is strong and supports many communication interfaces including SPI interface, I2C interface, USB interface, USART interface, ADC, timer and other peripherals[2]. With 80 IO ports, it is easy to connect peripheral modules for function expansion. The powerful function ensures that the chip can fully meet the hardware functional requirements of the whole system.

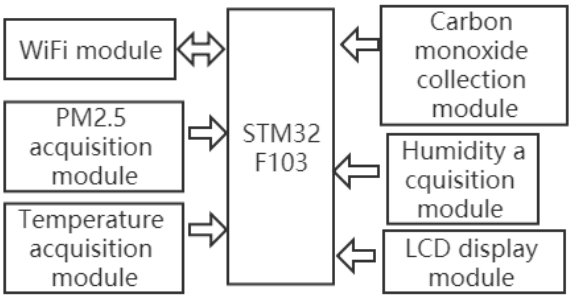


Fig.1 Overall architecture of hardware system

2.1. AD converter circuit

The STM32F103 series has three ADC, each of which has 16 conversion channels, with a conversion accuracy of 12 bits and a resolution of 1/4096. The ADC conversion circuit is shown in Fig.2. The dynamic contacts of the patch sliding rheostat of the AD conversion circuit in the monitoring and control system are connected to the ADC channel pins of the MCU chip. When the user rotates the sliding rheostat adjustment knob, the magnitude of the dynamic contact voltage in the AD conversion circuit part of the system will change accordingly. The range of voltage change is 0~3.3V, which is also the default ADC voltage collection range of the hardware end circuit board of the system.

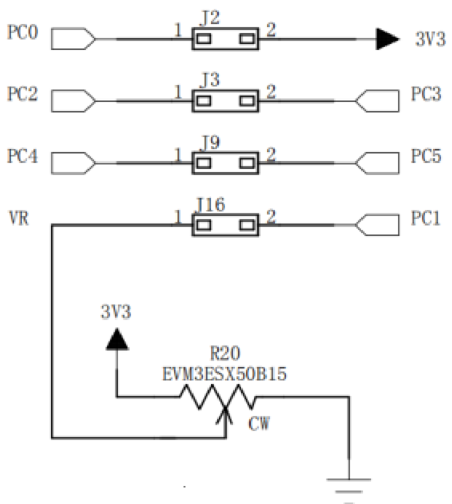


Fig.2. ADC converter circuit

2.2. JTAG interface circuit design

JTAG (Joint Test Action Group) is an international standard test protocol. It mainly tests the internal MCU chips of the monitoring and control system, and simulates and debugs the hardware control part of the system. As shown in Fig.3, there are four main lines in the JTAG interface part: TMS, TCK, TDI, TDO, which are mode selection, clock, data input and data output lines respectively. In the design of the monitoring and control system, the JTAG interface circuit uses 20-pin interface wiring connection. In order to enhance the anti-jamming capability of the download side of the system, ground wires are added between each signal line.

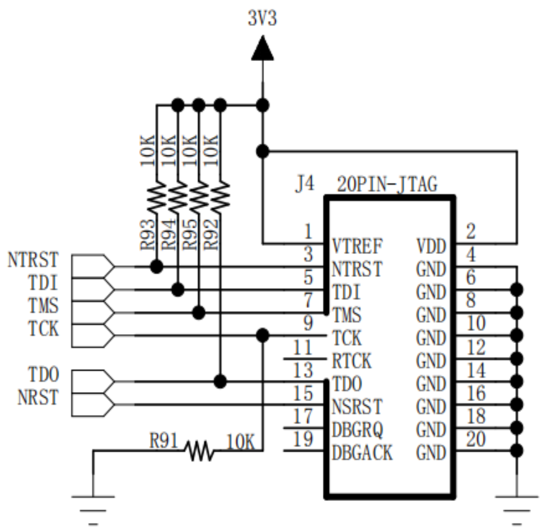


Fig.3. JTAG Download circuit

3. Peripheral module of hardware system

The peripheral modules of the hardware system mainly include temperature and humidity sensor, smoke sensor, interactive display screen, WiFi module and so on.

3.1. Temperature and humidity sensor

DHT11 digital temperature and humidity sensor is a temperature and humidity composite sensor with calibrated digital signal output. By using special digital module collection and temperature and humidity sensor technology, the environmental monitoring and control system has strong reliability and stability for temperature and humidity detection of the environment. The DHT11 module has low cost and strong anti-jamming capability, and it can measure the relative temperature and humidity of the automotive interior environment quickly. DHT11 uses full range calibration, which allows the module to be interchanged directly without having to be re-calibrated if a failure occurs. As shown in Fig.4, the DHT11 module contains a capacitive humidity sensor to measure ambient humidity and an NTC temperature sensor to measure ambient temperature. Besides, a small, high-performance 8bit MCU is connected to the module.

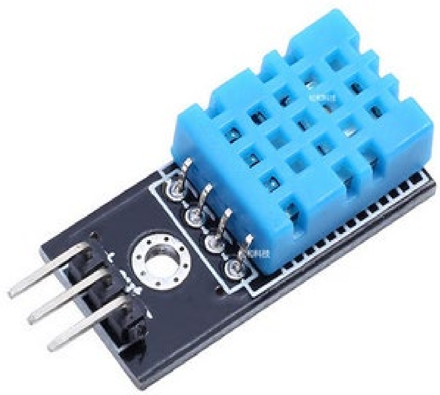


Fig.4. DHT11 sensor

3.2. MQ-2 sensor

The physical picture of the MQ-2 is shown in Fig.5. The MQ-2 gas sensor used in the environmental monitoring and control system can detect the presence of liquefied gas, butane, methane or alcohol in the environment which are potentially dangerous to human health. The conductivity of gas-sensitive materials increases with the increase of the concentration of combustible gas. The current smoke concentration value is output from the sensor signal by the conditioning circuit and sent to the MCU for processing. The alarm concentration of this sensor is 0 when it is not set, and the alarm concentration threshold can be set by the user with an upper limit of 100%. The MQ-2 sensor has a wide detection radius, maintains high sensitivity and fast response in a variety of complex environments and excellent stability in harsh environments.

Fig.5. MQ-2 sensor

3.3. ESP8266 module

ESP8266 WIFI wireless communication module is selected as the data transmission bridge between smart hardware and server in this system, as shown in Fig.6. The ESP8266 module integrates a low power ESP8266 chip developed by LEXIN Information Technology Company. The chip is highly integrated with a built-in TCP/IP protocol stack and TR switches, amplifiers, regulators, power management components, etc. Its on-chip storage and processing power is strong, and it requires very few peripheral circuits, so it is easy to embed and develop[3].

The ESP8266 module is powered by a 3.3V voltage and operates in three modes: SoftAP mode, Station mode and SoftAP+ Station coexistence mode. In SoftAP mode, this module is equivalent to a wireless access point providing wireless access services, similar to a wireless router. Other terminal devices can connect to set up a LAN with ESP8266 for data transmission; in Station mode, this module is equivalent to a wireless terminal, which can connect to the Internet through a router and transmit data to the server; in SoftAP+Station coexistence mode, ESP8266 module can establish LAN communication with other terminal devices as a wireless access point, and can also connect services for Internet communication[4].

The ESP8266 module usually uses TCP protocol and UDP protocol for data transmission. UDP is a connectionless, unreliable protocol which does not require the receiver to confirm before sending the data and data transmission efficiency is high, but the reliability is low, which makes it unable to determine whether the data is complete in the transmission process.

The TCP protocol is a link-oriented and reliable protocol which needs to wait for the receiver to confirm before the connection can be established. It is secure, reliable and the integrity of data can be guaranteed. The ESP8266 module in this system uses TCP protocol to set the WIFI application mode of ESP8266 as STA mode through AT instruction, and chooses TCP connection to realize the function of establishing data connection with the server.

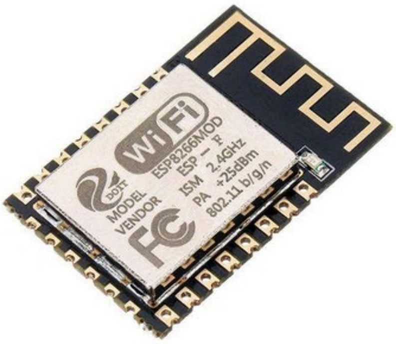


Fig.6. ESP8266 module

3.4. User interactive display module

The physical picture of the display is shown in Fig.7, which consists of an LCD display panel, a capacitive touch panel and a PCB substrate. The touch panel is equipped with a touch control chip, which is specially designed to handle the touch signals generated by the user when he touches the display, and then transmits the information through the outgoing signal lines. Because the MCU of the system cannot control the LCD panel directly, a LCD controller is needed to deal with the process of displaying. MCU only needs to give the data to the controller to display.

Fig.7. User interactive display module

4. Software development environment

This system chooses to develop the STM32F103VET6 controller in Keil5 development environment, and chooses C language as the development language. Keil5 is an integrated development environment with powerful compilation and simulation debugging capabilities.

STM32 is developed in two ways: register development and library development. This system is developed in the way of Library development, calling function interface directly to configure registers, which is fast and readable. Use STMISP downloader to program the STM32 core controller, and use TCP serial debugging assistant for functional testing.

The core controller of STM32 needs to collect data information of each sensor module, and control each peripheral module to realize the sub-functions of the system through internal logical analysis and processing.

5. Conclusion

The multi-functional vehicle interior environment monitoring system designed in this paper can achieve the functions designed in advance after testing. The LCD can display the data information collected by each sensor in real time. When the data exceeds the set threshold, the system will give an alarm.Through the serial debugging assistant, the WiFi communication function is normal, and the hardware device can connect and send data to the cloud server through WiFi module. The monitoring system designed in this paper has certain practical value.

References

1. Ibrahim, Dogan, Applying the ESP8266 processor in IoT application, *Electronics World*, 2017, 123(1976): pp.14-17.
2. Honghua Liao, Hao Fu, Weichuang Yu, et al., Design and Implement of Pyroelectric Energy Harvester Experimental Measurement System Based on STM32F103VET6, *Sensors & Transducers*, 2016, 96(1): pp.69-74.
3. Fengzhi Dai, Yuxing Ouyang, Yiqiao Qin, et al., Development of Integral Smart Home Appliances, *Journal of Robotics, Networking and Artificial Life*, 2018, 4(4): pp.291-294.
4. Craig A, Lindley, Meet the ESP8266, *Nuts & volts*, 2015, 36(10): pp.42-47.

Robot Structure and Motion Control Design Based on UG and Proteus

Yuhang Sheng

Tianjin University of Science and Technology, China

E-mail: 1346733865@qq.com

www.tust.edu.cn

Abstract

The paper is to design a six-degree-of-freedom biped robot by the research on the humanoid characteristics of the biped robot. Our biped robot chooses the steering gear ASMC-03B as its power unit, we restricts the model parameters by analyzing the function relationship between the steering gear torque and the volume, and regulates the size of every parts, and create a motion analysis model. In the hardware part, Arduino UNO, which is used as the main control chip, realize the communication between the main control chip and the servo drive module PAC9685 through the IIC bus protocol, which saves the main control chip resources and ensures the execution efficiency. In the simulation part, it is to simulate the steering angle of the steering gear, and output it in the form of a waveform.

Keywords: Keywords: Biped robot,6-degrees of freedom, Mechanical structure, Motion control

1. Introduction

With the rapid development of science and technology in today's society, robots are playing an increasingly important role in human life. The invention of robots is one of the greatest inventions in the history of human development. At the same time, it is a very inclusive cross-disciplinary. Biped robot, as one of the hot topics in the field of robot research, has been paid much attention, and its research results are more outstanding than other types of robots.

Relative to domestic, foreign biped robot research field started earlier. Among them, Japan is the first country to invest in biped robot research[1]. The DOF of biped robot has experienced the development from PLANAR DOF ROBOT TO SPATIAL DOF robot, the working mode of biped robot has developed from "teaching-reproduction" to autonomous decision-making robot. Recently, the most interesting biped robot research abroad is the Atlas Robot developed by Boston Power Company. It has a strong perception system and decision-making system, when encountered obstacles to self-judgment to adjust the center of gravity, and by updating the gait to plan the landing site to get away from obstacles. The latest version of Atlas adds body coordination and extreme sports such as cool running and triple jumping on one foot, as shown in Fig.1.

Fig.1. Atlas biped robot

Compared with foreign countries, domestic research on biped robot technology started late. In 1985, "863" plan included the research of intelligent robot in "the outline of China's high-tech Research and Development Program". Among the universities with outstanding achievements in the field of biped robots are the Harbin Institute of Technology, the National University of Defense Technology, and the Beijing Institute of Technology. In recent years, the outstanding achievements in the field of biped robot research in China are the "human-shaped robot Butler" Walker launched by

Shenzhen Youbixuan Co. , Ltd. in 2019, as shown in Fig.2, its range of motion ability includes all-round free walking, stable climbing and going up and down stairs, the use of three-dimensional visual navigation positioning system to achieve navigation and obstacle avoidance functions, applications such as remote monitoring, education and smart home.

Fig.2. Humanoid Machine Butler

But, the robot product production cycle is long, in the production process unpredictable factor is many. Based on this premise, this paper designs a 6-dof biped robot which meets the requirements by doing the pioneering work to the biped robot model.

2. Mechanical design

Before modeling the biped robot, the Mechanical Structure Diagram of the biped robot should be constructed, and then the joint and leg structure should be defined. The mechanical structure of biped robot can be designed according to the anatomy and physiology of human body, the left and right legs are symmetrically distributed, and 6 joints are 2 hip joints, 2 knee joints and 2 ankle joints. The mechanical design of the biped robot is shown in Fig.3.

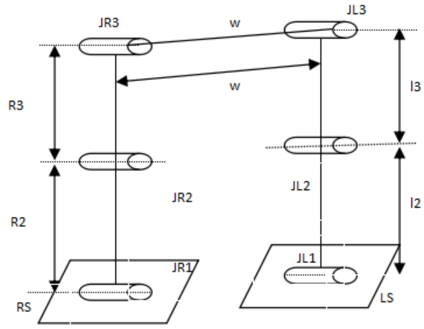


Fig.3. mechanical structure of robot

3. Design UG model

Computer aided technology (CAD) is widely used in the field of mechanical design. The application of CAD design combines computer application with engineering design. parametric design is the Yenshin of CAD technology. This article uses the Ug drawing software, which has the powerful compound modelling function, the modularization is more centralized. The mechanical structure design of biped robot can be divided into two parts: actuator selection and parameter design. The steering gear is Asmc-03b, and the Working Torque is 0.38N·M. Parameter design[2] refers to the design method of controlling the size and installation position of the parts through the core parameters, normalizing the size of the parts to a reasonable range. Among them, the working torque of the steering gear is an important index to standardize the size of the parts. According to the physical definition of Torque, the Torque is equal to the cross product of the radial vector of the rotating shaft and the acting force, and the expression of the Torque is as shown in (1).

$$\mathrm{M} = \mathrm{L} \times \mathrm{F} \tag{1}$$

Bring $\mathrm{F} = \mathrm{m} \times \mathrm{g}$ and $\rho = m/V$ into available

$$\rho \times \mathrm{V} \times \mathrm{g} \times \mathrm{L} < 0.38\mathrm{N} \cdot \mathrm{M} \tag{2}$$

When the size of the robot is designed, the parameters of the Cross section can be set according to the constraint conditions, and the size is reasonable.

The Mechanical Structure Model of biped robot is composed of steering gear, rudder plate, waist u-shaped support, foot plate, a-shaped parts, leg structure function support, z-shaped function support and other parts. By establishing the connection relation of each part and the constraint relation of the parts, the parts added to the assembly environment can be located precisely. The biped robot model assembled by constraint command is shown in Fig.4.

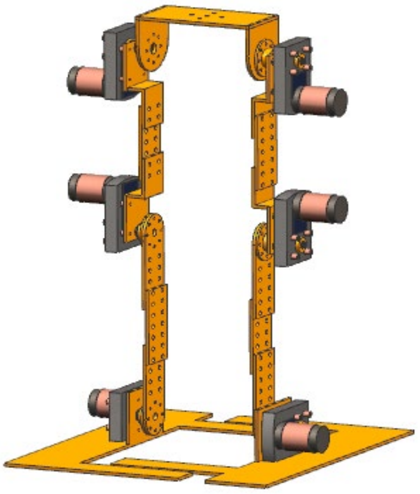


Fig.4. robot assembly drawing

4. Circuit design of Proteus

Based on the Proteus simulation circuit software, the hardware circuit of biped robot is designed to drive the steering gear. Proteus circuit design part mainly includes: Power supply voltage-stabilizing circuit, mcu Minimum System Circuit, rudder drive circuit, rudder circuit.

The power supply module is composed of three parts of regulated power supply circuit. The components of each regulated power supply circuit are composed of 78 series three-terminal regulator, capacitor, polar capacitor and power supply terminal. The regulated power supply circuit supplies power to the minimum system circuit, SERVO0, SERVO1, SERVO2, SERVO3, SERVO4 and SERVO5 rudder circuit respectively.

Single-chip microcomputer circuit[3] is the core of the minimum system circuit of single-chip microcomputer, and is an indispensable part of the design control circuit. The main control chip takes the ATmega328[4] core as the microprocessor, and completes the tasks of data storage, computation, peripheral device control and interrupt processing.

The SERVO drive circuit uses PAC9685 chip, in which the data transmission pin of SDA and SCL clock pulse pin communicate with the minimum system circuit of single chip microcomputer through IIC bus, and the LED pin is used as the output of PWM wave by connecting the row of pins of 16 pins, A0, A1, A2, A3, A4, A5 are the chip addresses for the driver chips and are connected to the enable signal.

The steering gear circuit is controlled by a micro-controller, which communicates with the steering gear drive module via IIC bus. The steering gear drive module outputs PWM wave to control the steering gear. The hardware circuit diagram is shown in Fig.5.

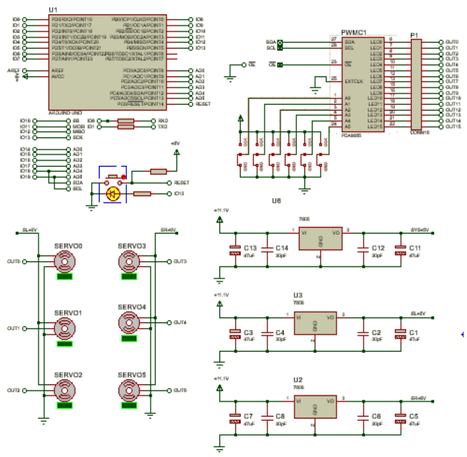


Fig. 5. Hardware circuit

5. Simulation experiment of Proteus circuit

The hardware-in-the-loop simulation is illustrated in Fig.6 with an example of a single steering gear experiment. The adjustable potentiometer (a) and the display screen (b) are added to make the circuit simulation visual. By adjusting the potentiometer to output analog voltage, analog voltage signal transmission to the microcontroller minimum circuit, SCM internal voltage value according to register configuration, output the corresponding PWM wave drive actuator rotation angle.

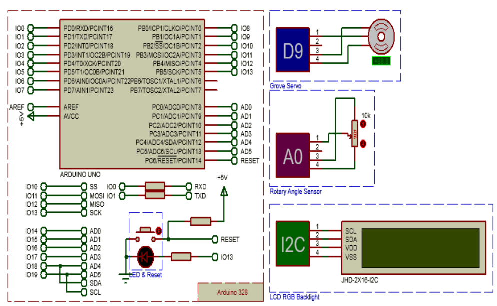


Fig.6. Single steering engine circuit diagram

By adjusting the adjustable potentiometer to control the pulse width of the steering gear, the Steering Gear Angle from -90 ° to 90 ° rotation, as shown in Fig.7.

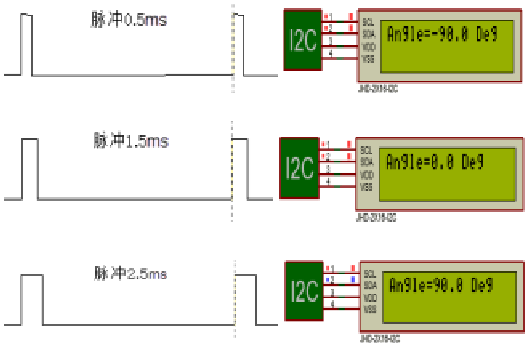


Fig.7. Control diagram of pulse width and angle

The experiment shows that the pulse width is controlled by an adjustable potentiometer, the duty cycle of the actuator is changed from 0 to 100% , and the rotation angle of the actuator is displayed by a display screen, so as to achieve precise control.

6. Conclusion

The structure and Motion Control Circuit of biped robot are designed in this paper. In the part of mechanical structure design, through the analysis of the relationship between the rudder torque and material, the dimensions of the parts are standardized to reach a reasonable range. Modeling and assembly of parts using UG software. In the part of Motion Control Circuit, the Proteus circuit is mainly designed, the Arduino Uno is used as the control chip, PAC9685 is used as the actuator module, and 78 series three-terminal voltage regulator is used as the regulator module. In the part of hardware circuit design, the steering gear rotation is simulated by visual programming to realize the precise control of the steering gear.

References

1. Chen C, Shih B, Shih C, et al. Design, modeling and stability control for an actuated dynamic walking planar bipedal robot. *Journal of Vibration and Control*, 2013, 19(3): 376-384.
2. Yunqiang Li, Guobin Lin. Design of intelligent parking lot based on Arduino. *IOPConference Series: Materials Science and Engineering*, 2019, 490(4).
3. Yuan M, Kang M, Ding Y. Parametric Model Of Excavator Working Equipment Based On The Advanced Development Of UG. *International Conference on Mechanical Engineering & Intelligent Systems*. 2015.
4. Marwa F Areed. A keyless Entry System based on Arduino board with Wi-Fi technology. *Measurement*, 2019, 139.

Design of Brushed Motor Position Loop Control System Based on Incremental PID

Tianyi Zhang *, Peng Lu

Tianjin University of Science and Technology, China;

*E-mail: * 2679483816@qq.com*

www.tust.edu.cn

Abstract

The motors are an important part of the transmission and control system. With the development of technology, the focus of motors in practical applications has shifted to precise control of speed, position and torque. This paper focuses on the control system design of the brush motor position loops, using STM32 microcontroller and incremental PID to achieve accurate control of the motor stopping moment position. The software is designed in C programming. It is divided into four parts: System Clock Configuration Functions, Principal Functions, System Tick Timer Interrupt Callback Function, Position closed loop PID Control Design Function. The hardware part is based on the STM32F103VET6 core board.

Keywords: DC brush motor, STM32, closed-loop control, PID

1. Introduction

For many reliable technical reasons, DC-powered brush motors have been largely replaced by electronically controlled brushless motors. In spite of this, in less demanding or cost-sensitive applications, Brushed motors are still an effective solution.

By combining the Brushed Motor with the Basic Motor Driver IC, brushed motors and end products allow for many additional operations and better protection. This paper describes the design of position loop hardware and software control system, for brush motor incremental PID. The function of this system is to rotate the motor, and change the stopping position after pressing the key.

For YS-F1Pro demo board hardware, we have obtained a program on the rotation of hard stone motors. We have optimized the changes based on this program. The obtained source program is optimized to be migrated to the core board containing the STM32F103ZET6 chip. This is used to drive the motor to rotate according to programmed instructions. control intelligent household device and monitor the information of intelligent household device.

2. The Hardware Selection

The entire hardware system consists of three parts: the motor, the hardstone driver board, and the black core board containing the STM32F103VET6.

2.1. Driver board selection

For the driver board, choose the motor driver ULN2003 from the YS-F1Pro development board. ULN2003 is a high-current drive array. It is mostly used in control circuits of microcontrollers, smart meters, PLCs, digital output cards, etc. It can directly drive relays and other loads. Its input 5V TTL level and output up to 500mA/50V. ULN2003 can be directly connected to TTL and CMOS circuits at 5V operation, it can directly process data that would otherwise require a standard Logical Buffer to process.

The ULN2003 is a family of high-voltage, high-current Darlington transistor arrays, it features high current gain, high operating voltage, wide temperature range, and high load capacity. It is suitable for all types of systems requiring high speed and high power drives.It can drive brushed DC motors of up to 600W (60V10A).

The schematic diagram of the driver chip ULN2003 is shown in Fig.1.

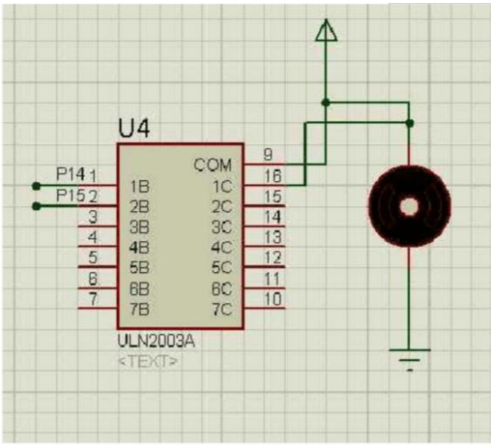


Fig.1. The schematic diagram of the driver chip ULN2003

2.2. Motor selection

For the motor, We chose the GM37-545 Hall-coded geared motor, which has a precision reducer for better coil amplification, and features high power and high torque. the GM37-545 Hall-coded geared motor is shown in Fig.2.

Fig.2. the GM37-545 Hall-coded geared motor

2.3. Core chip selection

For the core broad, the chip selected is the STM32F103VET6. which is the Cortex-M3 ARM core and has a 100-pin. The chip has 128KB flash memory and 20KB RAM. It can through SWD or JTAG with KEIL 4 to download the program and online debugging [1]. And it has more than 80 general-purpose IO ports, multiple timers and serial ports, which fully meet the design of the required pin and memory requirements. It has the following characteristics:

- Supported peripherals: Timer, ADC, SPI, I2C and USART.
- Operating Voltage 2.0V ~ 3.6V.
- Operating temperature range: -40 ºC ~ 85 ºC.

The design of the main control chip is shown in Fig.3.

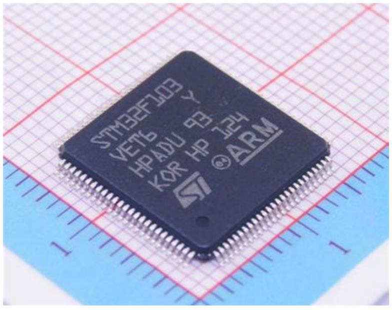


Fig.3. STM32F103VET6 chip

3. Software Design

Software design mainly includes the design and writing of programs, the application of functions in programs, etc.

3.1. Program Block Diagram

The program block diagram of system is shown in Fig.4.

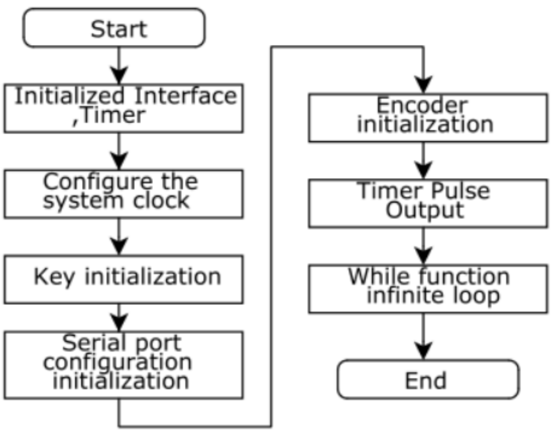


Fig.4. program block diagram

The first step is to configure the Flash interface and system ticking timers, keys, serial communication, encoder initialization, and so on. Then configure the timer pulse output. Finally, set the LCD program to an infinite loop with the while statement.

3.2. System Timer Interrupt Callback Function

This function calculates PID results based on a velocity loop period of 60ms ,and no change in limiting speed. It uses PWM duty cycle to determine the current motion direction. Note that every time a timer occurs, the interrupt enters the callback function once.

3.3. Position closed loop PID control function

The function is enter the current control amount and output the target control amount. First, the deviation is calculated, and the deviation dead-band is used to determine if the deviation exists. The " iIncpid" is calculated according to the formula for the incremental PID, and the final error is stored for the next calculation [2]. Main procedures for incremental PID control is shown in Fig.5.

```
/* Dead Zone Limits */
  if ((iError<50) && (iError>-50))
     iError = 0;
 iIncpid=(int32_t) ((sPID.Proportion * (float)iError)            // Term E[k]
                   - (sPID.Integral * (float)sPID.LastError)       // Term E[k-1]
                   + (sPID.Derivative * (float)sPID.PrevError) );// Term E[k-2]

  sPID.PrevError = sPID.LastError;        // Store errors for the next calculation
  sPID.LastError = iError;
  return(iIncpid);                          // Return the incremental value

}
```

Fig.5. Main procedures for incremental pid control

This program uses the PID formula to calculate the deviation iError：

Deviation = set value - measured value

If -50 < deviation < 50, it can be considered 0.

The PID formula is as follows：

$$u_k = K_P * e_K + K_i \sum_{j=0}^{k} e_j + K_d\,(e_k - e_{k-1}) \quad (1)$$

PID incremental value = current error E[K] - last error E[K-1] + last error E[K-2]

$$\Delta u_k = u_k - u_{k-1} = \mathrm{A}e_k - \mathrm{B}e_{k-1} + \mathrm{C}e_{k-2} \quad (2)$$

The last error is stored for the next calculation, returning an infinite loop of incremental values.

4. Project Experiment

4.1. Motor control theory

The drive board adopts H-bridge driving circuit, the directional control of brushed DC motor is shown in the figure, when Q1, Q4 conduction, Q2, Q3 closed, the current direction is Q1 → motor→ Q4, then the motor is positive, when Q2, Q3 conduction, Q1, Q4 closed, the current direction is Q3 → motor→ Q2, then the motor is reversed. Motor Forward Rotation Diagram is shown in Fig.6 .Motor Reverse Diagram is shown in Fig.7.

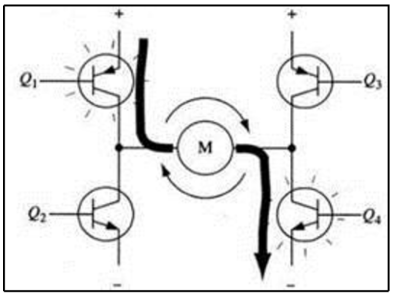


Fig.6. Motor Forward Rotation

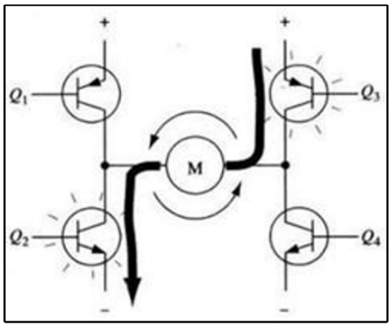


Fig.7. Motor Reverse

4.2. Experimental procedures

After hardware selection and software design, hardware and software connections are needed to complete the experiment.

There are the following steps：

- Establishment Project
- Create new Usart workgroups and files and link settings.
- Importing software design programs.

  Change the optimization program according to experimental requirements and schematic diagrams, design main functions and sub-functions, program the system clock configuration function, main function, system drop timer interrupt callback function, and position closed-loop PID control design function.

- Connecting Hardware Devices.
  Connect the core board, motor and YS-F1Pro drive board according to the given pin diagram and datasheet.

The experimental results that can be obtained：

1. Press KEY1: motor start up
2. Press KEY2: motor stop
3. Press key 3 and add 1 from the motor's stop position.
4. Press key4 and subtract 1 from the motor stop position.

5. Conclusion

This experiment is to write software design for position ring control of brushed motor (incremental PID) using C program. The program includes the main function, timing interrupt callback function, clock configuration function, position ring design function, etc. Connect hardware devices to complete the experiment. Use PID function to calculate the motor stop position.

References

1. Yuxing Ouyang , Fengzhi Dai , Yiqiao Qin , et al. Integral Design of Intelligent Home Equipment. *Proceedings of International Conference on Artificial Life and Robotics,* 2017, 22: pp.148-151.
2. Zhiwang Chen, Baohua Liu, et al, *Getting Started with STM32 Embedded Microcontrollers*, YanShan University, Hebei, China, 2014.5.

Design of a Fire alarm system

Yuhui Cheng *, Tianyi Zhang

Tianjin University of Science and Technology, China;

*E-mail: * 13642017570@163.com*

www.tust.edu.cn

Abstract

This article introduces the design of the fire alarm system and analyzes its working principle. The system uses a distributed sensor network based on ZigBee to achieve fire alarm and real-time evacuation. Select the infrared flame sensor and MQ2 smoke sensor as the information source. The core control board is the Arduino Mage2560 board. Through the ESP8266 module and the GPRS module, the multi-channel information transmission function can be realized. The alarm can be classified according to the fire situation.

Keywords: ZigBee, Hierarchical alarm, crowd evacuation

1. Introduction

With the development of cities, the increasing number of high-rise buildings has led to an increase in fire hazards. Due to the difficulty of evacuation and fire fighting in high-rise buildings, once a fire occurs, it will cause serious losses. In this case, it is particularly important to call the police in time and evacuate the crowd.[1].

At present, most fire alarm devices are independent alarms. The wiring is cumbersome and the wiring is prone to aging and wear. Therefore, the probability of failure or false alarms of such alarm devices is relatively high. On the other hand, a system that only warns but does not provide specific fire conditions, nor can it guide people to evacuate.

Based on the above discussion[2], we designed a new type of fire alarm system. The system obtains data through the infrared sensor and MQ2 smoke sensor and uploads it to the client through the ESP8266WIFI module. The Zigbee protocol is used for wireless communication between monitoring nodes. When a fire breaks out, the system will alarm at different levels, and guide the crowd based on the location and scale of the fire. The overall structure of the system is shown in Fig.1.

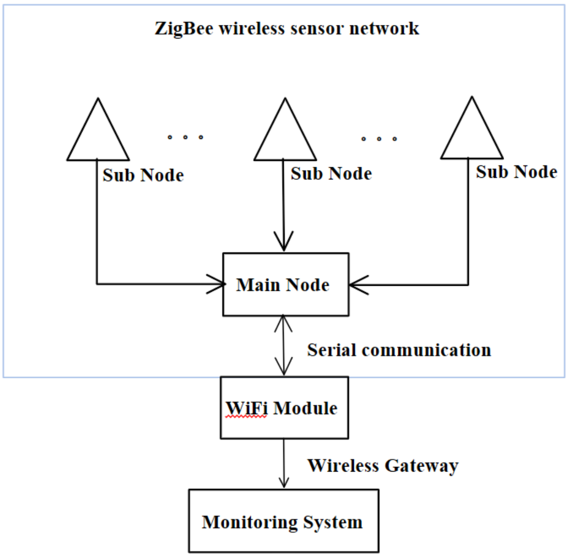


Fig.1 System structure

2. The Hardware Structure

The hardware of this system mainly has four parts. The Arduino Mage 2560 board and the Arduino Nano board constitute the control unit of the system node; the infrared sensor and the MQ2 smoke sensor on the node constitute the data monitoring unit; the CC2530 communication module, ESP8266 WiFi module and GPRS SIM300C module constitute the communication unit; the multi-directional indicator light and the voice module constitute the evacuation unit.

2.1. control unit

The system is composed of the main node and a sub-node. The main node uses the Arduino Mage2560 development board as the control board, and the sub-nodes are designed based on the Arduino Nano development board.

The Arduino Mega 2560 is a microcontroller board based on the ATmega2560. It has 54 digital input/output pins (of which 15 can be used as PWM outputs), 16 analog inputs, 4 UARTs (hardware serial ports), a 16 MHz crystal oscillator, a USB connection, a power jack, an ICSP header, and a reset button.

The Arduino Mega 2560 board is shown in Fig.2.

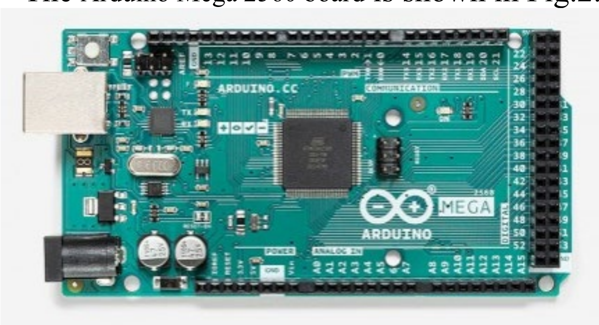


Fig.2 Arduino Mega 2560 board

The Arduino Nano is a small and complete board based on the ATmega328.

The Arduino Nano board is shown in Fig.3.

Fig.3. Arduino Nano board

2.2. Data monitoring unit

Fire alarm equipment with a single information source is prone to false alarms or untimely alarms. The system adopts infrared sensor and smoke sensor as the data monitoring unit of this system.

The infrared sensor can collect the light signal in the range of 750~1100 nanometers, and is very sensitive to the flame spectrum. It can judge whether the flame is detected by converting the light signal into a high and low-level signal by calculation. The detection angle of the sensor is about 60°, and the working voltage is 3.3V~5V.

The infrared sensor is shown in Fig.4.

Fig.4. infrared sensor

In the early stage of a fire, a lot of smoke is generated, so the detection of smoke can also be used as one of the methods of fire monitoring. The system uses the MQ2 smoke sensor as the smoke data collection module. The sensor has a response time within 10s and can monitor smoke data from 300PPM to 10000PPM.

The MQ2 smoke sensor is shown in Fig.5.

Fig.5. MQ2 smoke sensor

2.3. Communication unit

Aiming at the problem of complicated wiring of wired communication equipment and prone to failure, this system uses wireless communication for data exchange. For the communication between nodes, this system selects the CC2530 radio frequency module as the communication unit between nodes, which has the function of ZigBee wireless networking. This module has the following characteristics:

- With strong adaptability, nodes can be added or deleted according to the needs of the usage scenario
- Effective communication distance: 100m
- Low power consumption

After the information is summarized to the master node, the data will be sent to the client through the ESP8266 WIFI module to realize remote monitoring.

When a fire occurs, the system may face power outages and network outages. In order to send the alarm information to relevant persons in time, the system is equipped with a GPRS SIM300C communication module, which can carry out a fire alarm in the form of SMS notification.

2.4. Guiding unit

Multi-directional LED indicators and MP3 voice module constitute the system's guiding unit. Multi-directional LED indicator, the main body is composed of red and green LED lights. The MP3 voice module supports audio output in MP3 and WAV formats, which can be controlled through the serial port of the microcontroller.

The design of the sound and light evacuation module is shown in Fig.6.

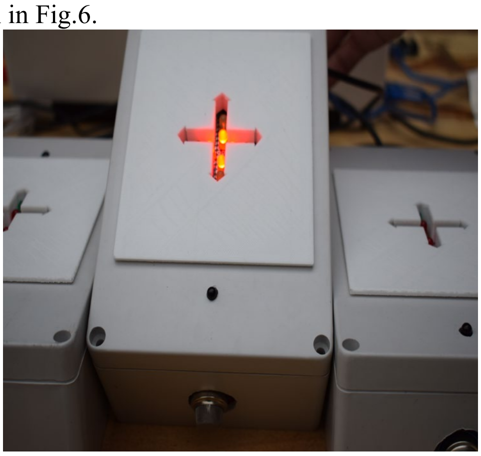


Fig.6. Sound and light evacuation module

3. Introduction of functional module

3.1. Data monitoring function

The system uses infrared sensors and smoke sensors to collect information and sends it to the client for digital display. The user can view the monitoring data, working status, and alarm status of each node through the system interface. The client is written in C# language and has good compatibility and stability.

The design of the monitoring software is shown in Fig.7.

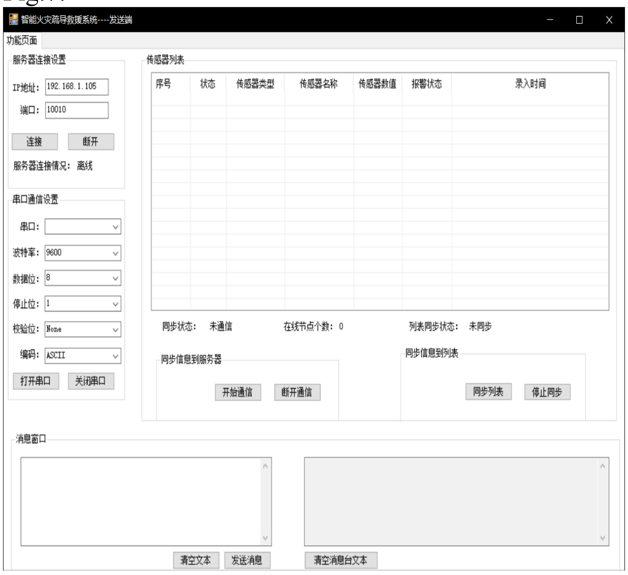


Fig.7. monitoring software

3.2. Design of Hierarchical alarm function

In order to prevent invalid alarms due to sensor failures or operating errors, this system has designed a hierarchical alarm function. The discrimination mechanism will divide the fire situation into two levels according to the number of alarm nodes and the data of alarm node.

When the fire is small and easy to extinguish, the system will send an alarm message to the corresponding person in charge to notify him to deal with it. When the fire situation is large or there is a continuous development trend, it will directly send an alarm message to the neighboring fire department through the alarm platform.

3.3. Sound and light guiding function

In order to effectively guide the crowd, the system has designed a light guide and voice prompt guide.

The light grooming function is completed by a multi-directional indicator, which will display red in the direction close to the fire point and green in the direction far away from the fire point according to the information of the alarm node.

The function of the voice prompt is realized by the voice module. When a fire breaks out, the system will broadcast an alarm and broadcast the approximate location of the fire based on the node location information entered in advance to provide guidance for crowd evacuation.

4. Conclusion

This fire alarm system is suitable for high-rise buildings or crowded indoor shopping malls. After a fire occurs, it can report to the police in time, and provide a reasonable evacuation route according to the location parameters of each fire detection node to facilitate the evacuation of the crowd.

References

1. Jianguo Wang, Baoyi Guo, Monitoring system based on ZigBee and WiFi buildings combining, *Journal of Xi'an Technology University*, 2008, 28(5): pp.460-465.
2. Fan Yang, Maolin Chen, Xun Wu, et al., Forest fire alarm system based on sensor information fusion technology, *Journal of Huazhong University of Science and Technology: Natural Science Edition*, 2013, 41(2): pp.22-25.

Design of Intelligent Curtain Control Circuit Based on Single Chip Microcomputer

Qianqian Zhang *, Jichao Zhao, Haokang Wen, Hongbo Hao

Tianjin University of Science and Technology, China;

E-mail: 577972810@qq.com

www.tust.edu.cn

Abstract

Since entering the 21st century, high technology has promoted the development of human beings and artificial intelligence has been popularized gradually. In this paper, the structure and principle of crystal oscillator circuit, reset circuit and photosensitive sensor circuit are introduced with 89C51 single chip microcomputer as the main control unit. The whole circuit system is analyzed in this paper. The circuit module of photosensitive sensor can be used to detect the external light intensity and automatically control the curtain. The user can set the temperature threshold with the remote control.

Keywords: The steering gear 89C51; Photosensitive; Embedded remote control; Temperature monitoring

1. Overall scheme design

This design starts from people's demand for the system design function, under the circumstance of comprehensive consideration of various factors, designs the overall framework of the automatic control system, and on the basis of the realization of the overall function, considers the scalability of the system as far as possible [1].

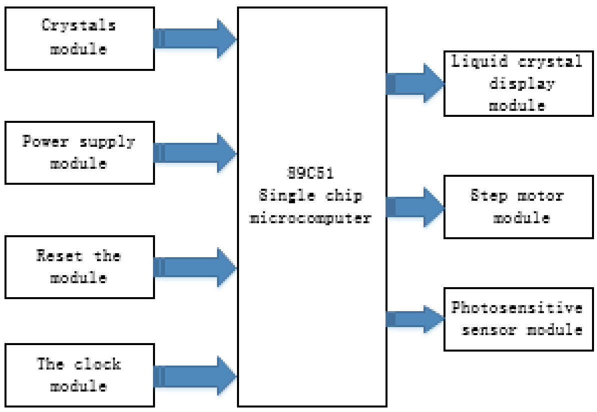


Fig.1. Overall scheme block diagram

2. Each module scheme design

(1)Photosensitive control module: This module firstly changes the resistance value of photoresist through the change of the external light intensity, and then causes the change of output voltage. When the light intensity weakens to the set value at night, the light intensity is sensed by the photosensitive sensor, and the stepping motor is controlled by 51 MCU to close the curtain. When the light intensity increases to the set value in the morning, the light intensity is sensed by the photosensitive sensor, and the 51 single-chip microcomputer controls the stepper motor to open the curtain.

(2)Temperature monitoring module: the infrared remote control is used to set the temperature value, and the temperature monitoring system is used to detect the ambient temperature. When the temperature exceeds or is lower than the given value, 51 single-chip microcomputer controls the stepping motor to turn forward to open or close the curtain [2].

(3)Infrared remote control module: The remote control of this module is realized through the infrared signal sent by THE HT6221 chip. When the receiver receives the infrared signal, the decoding operation is performed first, and then the code value corresponding to each button on the LCD is displayed. The software program allows the user to control the forward and reverse of the motor through the remote control, so that the curtain can be opened or closed. Here, another function of infrared remote control is to set the initial temperature value, and the button can change the initial temperature value, which makes people's life more convenient and comfortable.

3. Subsystem design

3.1Circuit design steps

The crystal oscillator in the circuit is used to generate the reference frequency and optimizes the frequency stability and anti-interference. The accuracy of the frequency in the circuit is controlled by the reference frequency. In addition, it can also produce the clock signal and oscillation current of the microcontroller, the operation of the single chip cannot do without crystal oscillator, it is the peripheral hardware that the single chip system cannot lack, if the circuit is abnormal, the single chip system will be paralyzed, unable to run, leading to the failure of the whole system. All the programs of the single chip microcomputer are written into the read-only memory when they are written.

The single chip microcomputer system must read the program after it is started. The time it takes the processor to read a piece of code is one machine cycle of the processor, and this time is the beat for the processor. Therefore, crystal oscillator circuit is the core peripheral circuit of processor operation. The 12 MHZ vibration source is selected in this design. Although the internal vibration source of the MCU is provided, external vibration source is selected because of the design requirement. Figure2 shows the microcontroller crystal oscillator circuit.

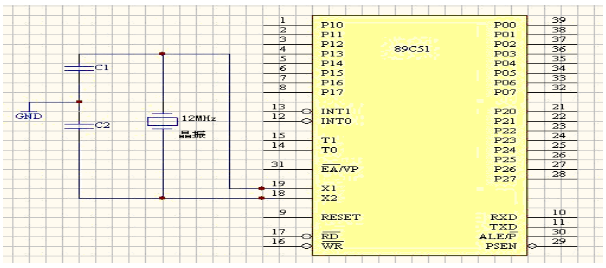


Fig.2.Single chip crystal oscillator circuit diagram

3.2Reset circuit

The processor designed the reset function circuit, which is used to reset the system when the system wants to return to the state of restart. In fact, when the processor starts to run on power, it is a reset state, so the processor startup is required to reset. You will also need to use the reset button to make pin 9 high to restart 89C51 in the "On" or "dead" locked state. Figure3 shows the reset circuit diagram.

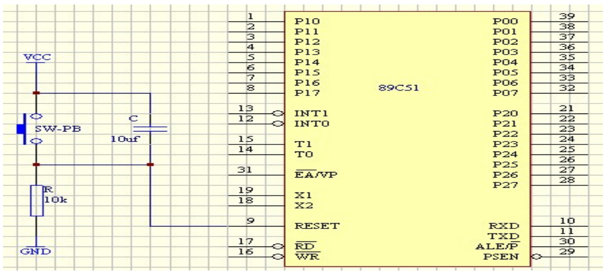


Fig.3.Reset circuit diagram

3.2 clock circuit

This design requires the curtain to be opened or closed within a specified time, so a timer is needed. To ensure that the microcontroller is consistent with the external clock, the DS12887 real-time clock chip is used. The chip data can be stored for more than 10 years without damage, and the signal on the RESET pin has no influence on the RAM in the MCU[4]. The clock circuit diagram is shown in Figure4. RESET is directly connected to the Vcc to ensure that DS12887 enters or exits the power off state. From this pin, it can output a square wave with a frequency of 2HZ-256Hz.

When the interrupt occurs, the microcontroller reads the input signal once or twice to check if the rolloff state has been entered and obtains the three-phase current and voltage at the entire point. MOT grounding, the high 8-bit of DS12887 is set to 7FH, while the low 8-bit is determined by the address of the chip (00H-3FH) in each cell.

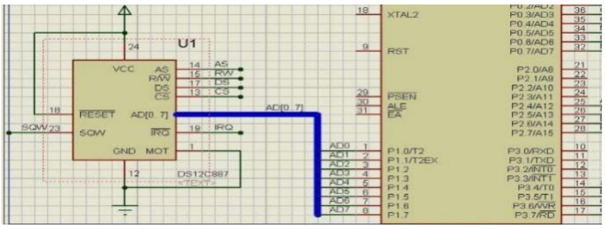


Fig.4. Clock circuit diagram

3.4Power circuit

The operation of the MCU requires 5V voltage, so it is necessary to design the power circuit for the MCU. Figure 5 shows the power circuit of the microcontroller.

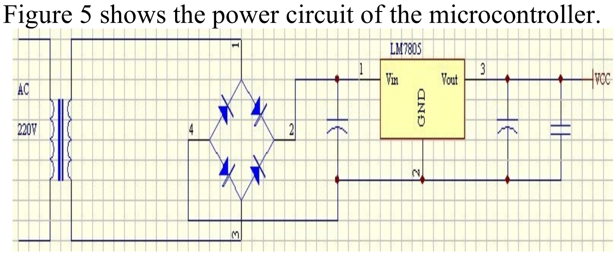


Fig.5. Power circuit diagram

3.5 Stepping motor circuit

Stepper motor is a digital servo actuator with simple structure, reliable operation and convenient control. SCM is controlled by the rotation Angle and displacement of the stepper motor [3]. This design adopts the model of three-phase stepless variable speed motor 130HZ308-450 to control the rotation Angle and displacement. Stepper motor has high torque, high impact resistance and high precision. The drive circuit of the stepping motor operates according to the control signal. Fig.6 and Fig.7 show the block diagram and circuit diagram of the stepper motor control system.

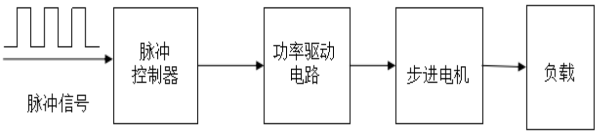


Fig.6. Block diagram of stepper motor control system

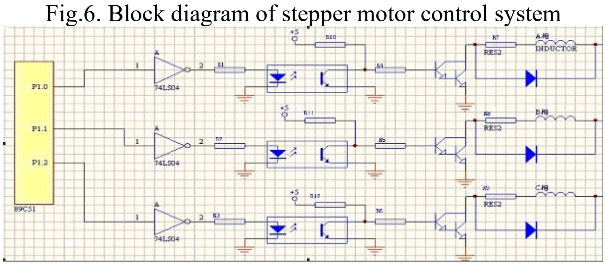


Fig.7.Step motor control system circuit diagram

The pulse distribution circuit is controlled in a predetermined order to open and close the phases. This design not only simplifies the circuit and reduces the cost, but also flexibly changes the control scheme of stepper motor according to the needs of the system. Software control pulse is described in the software design section.

In order to operate in the state of large pulse current, the stepper motor drive circuit works. The photoelectric coupler is used for separation to avoid interference between the SCM and the power supply circuit of stepping motor, and to prevent the main control system from strong electrical interference. In addition, if the driving circuit fails, the MCU at the higher voltage of the power amplifier will not be damaged, and the external resistor of the driving circuit consumes a lot of energy, which will affect the stability of the circuit [5]. Dual voltage drive circuit, as the name implies, has two supply voltages. The current displays on high and low voltage connections, which inevitably results in torque at valley points, which is not easy for motor normal operation to drop.

This disadvantage overcomes the drive to the chopper and the efficiency of the stepper motor can be improved. In this design, constant frequency PWM subdivision drive circuit is adopted, and the block diagram of this circuit is shown in Figure8.

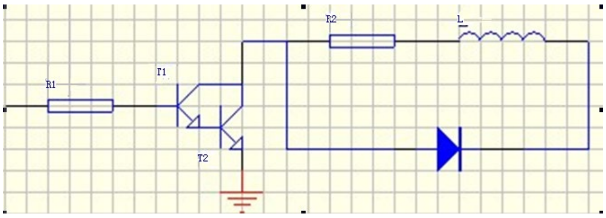


Fig.8.Constant pulse width modulation subdivision drive circuit

In SCM application system keyboard can input data and send instructions to the microcontroller. It is the main means of manual intervention by a microcontroller [7]. When the program is wrong or has an error, you can press the reset key to restore its original working state.

3.6LIQUID crystal display circuit

The connection of liquid crystal module is shown in Figure9. Liquid crystal is used for display in the system. This device can display all characters except Chinese characters, which can fully meet the design requirements. The single chip microcomputer can control the contents of the LIQUID crystal display through the data port and the sequential interface. Through the operation of the software can display any position of the liquid crystal specified display content, easy to operate, simple. The most important is the LCD requirements of the power parameters and SCM system consistent, so that the general power supply, power supply design does not have to reduce design work.

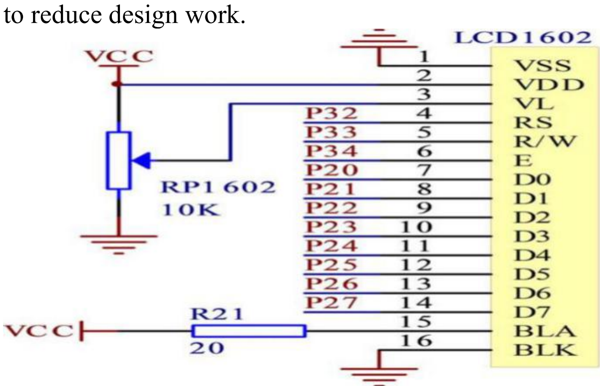


Fig.9.LCD module connection diagram

3.7Photosensitive sensor circuit

The electric curtains are automatically switched on and off depending on the lighting conditions, so photoelectric sensors are required. Photodiode is the main component of photosensitive sensor. It is a photoelectronic device operating on the basis of photoconductor, also known as the light tube[8]. The photodiode has no polarity. It is a purely resistive device. It can be used with a DC offset voltage or ac voltage. When no light is present, the dark resistance of the photodiode is large, and the current in the circuit is small. When there is light irradiation, its photodiode resistance drops sharply, then the current in the circuit increases rapidly.

Photodiode has the characteristics of high sensitivity and good light splitting. In addition, it is small in size, light in weight and stable in performance. Therefore, it is widely used in automation technology. Photodiodes are commonly used for optical measurement, optical control, and photoelectric conversion (light to electricity). Typically, photodiodes are sliced to absorb more light. When it is illuminated by light, the electron-hole pair is excited to participate in the conduction in the semiconductor wafer (photosensitive layer), reducing the resistance value and thus increasing the current in the circuit.

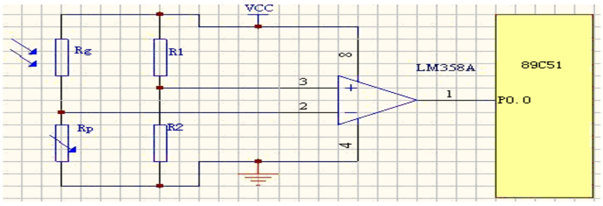


Fig.10.Schematic diagram of optical control circuit

Using the light control principle, the daylight curtain automatically opens and the dark screen automatically closes. The operational amplifier consists of a comparison circuit with two resistors. The two resistors are divided under the same input to obtain a voltage value, which is used to compare with the reference voltage. The reverse input USES a photodiode to collect the light intensity of the external environment. When the indicator light is on, the voltage value of the invert input is constant. Then the voltage values of the two resistors are compared, and the compared signals are sent to the MCU to determine the positive and negative rotation of the motor through the MCU to control the conversion of high and low voltage at port P0, so as to realize the functions of automatic opening of daylight curtains and automatic closing of evening curtains.

In the meantime, we can replace photodiodes with cheaper ones on the market, but the effect is the same. Photodiodes can change their physical properties according to the intensity of light. The signal is changed according to changes in the resistance inside the photodiode by using the unidirectional conductivity of the diode.

4. Conclusion

The stepper motor is used as a control component of a single-chip microcomputer to complete the main task of the opening and close the curtain; Photoresistors are used as detection kits to provide external illumination changes for single-chip microcomputers; Infrared detection circuit is used to achieve manual control, the main control chip using 89C51 MCU. In addition, with the help of buttons and display circuits, the intelligent requirements of the automatic curtain control system are finally realized in collaboration with each module.

The sensor part USES photodiodes, which can continuously detect changes in the external light intensity. After the signal from the bridge circuit enters the comparator, a signal can be obtained, and the operation of the stepper motor is controlled by the pulse signal of the single-chip microcomputer[7].This design of the stepper motor can execute the commands of a single chip very well. Stepper motor is a digital servo actuator, which has the advantages of simple structure, reliable operation, convenient control and good control performance. The designed clock circuit is compatible with the timing function of the single chip microcomputer, and the detection light intensity of the photoelectric sensor solves the problem of automatic control.

References

1. Li Chaoqing. *Principle and Interface Technology of Single Chip Microcomputer*. Beijing University of Aeronautics and Astronautics Press, 2013.
2. Li shengli. *Fundamentals of analog electronics technology*. School of electrical and electronic engineering, wuhan institute of engineering, 2013, 13.
3. Yuan Zengmin. *Basic Course of Analog Electronics technology*. Beijing: Electronics Teaching and Research Group, Tsinghua University, 2014.
4. Song jingbin. *Linux network programming* (2nd edition).Beijing: Tsinghua university press, 2014, 2.
5. Liu gang, Zhao jianchuan. *Linux system transplantation* (second edition). Beijing: Tsinghua university press, 2014, 2.
6. Hu Hancai. *Principle of Single Chip Microcomputer and its Interface Technology*. Beijing: Tsinghua University Press, 2010: 59-62.
7. Liu Zhuqin, Bai Zesheng. (in Chinese) *A Digital frequency meter based on single chip Microcomputer*. Modern Electronic Technology, 2010, (1):15-18.
8. Chen Ken, Yang Xiangdong, Liu Li, Yang Dongchao. *Robotics and application*. Beijing: Tsinghua University Press, 2006.

Research on the Algorithm of Flue Gas Desulfurization System

HongBo Hao

Tianjin University of Science and Technology, China

E-mail: 2844011096@qq.com

www.tust.edu.cn

Abstract

In this paper, the process of flue gas desulfurization and denitrification, which is a nonlinear, time-varying, large lag and strong coupling complex variable process, is analyzed and studied in depth. Based on the analysis of the coupling characteristics of the adsorption tower, the control model of the desulfurization and denitrification process of the adsorption tower is established, and the system identification adaptive PID Decoupling control algorithm is adopted. The Simulink toolbox corresponding to Matlab software is used to simulate and verify the effectiveness of the algorithm. The RBF neural network identifier is used to identify the plant model on-line to improve the adaptive ability of the controller.

Keywords: flue gas, desulfurization, PID, RBF neural network, Matlab

1. Introduction

With the continuous development of the country's social economy, the law enforcement of environmental protection has also been strengthened. In order to improve the regional air quality, the state has vigorously carried out the work of supporting the improvement of atmospheric environmental quality. In recent years, the concentration of inhalable particles in urban air has decreased significantly, but the situation of air pollution is still very serious. The power industry, as a major pollution reduction industry, it is facing great pressure from air pollution control. Activated carbon integrated desulfurization and denitrification technology has been widely used in flue gas purification in recent years. Desulfurization and denitrification process is a nonlinear, large lag, strong coupling complex industrial process control system, domestic and foreign scholars have done a lot of research on it, and its control has basically realized digital [1]. At present, intelligent algorithm and control technology are combined, and various optimization algorithms are used to optimize it. The desulfurization and denitrification process is based on the joint action of physical and chemical reactions.

The corresponding research institutions have formed a lot of mature theoretical literature on this reaction mechanism. However, for different projects, the process adopted is different, the diversity of equipment and the difference of process flow. So far, the desulfurization and denitrification system have not formed a unified mathematical model, nor has a complete set of intelligent control algorithm. In recent years, there are more and more researches on Modeling and control of flue gas desulfurization and denitrification system. The research direction is mainly divided into the following aspects: process improvement, updating and adjustment of desulfurization and denitrification chemical equipment, optimization of control and prediction algorithm.

The most important part of activated carbon desulfurization and denitrification system is adsorption tower. In this paper, the intelligent control combined with neural network and PID control is used to control the adsorption tower, which can overcome the problems of nonlinearity, large lag and coupling in the complex industrial process of desulfurization and denitrification.

In this paper, the control system of sintering flue gas desulfurization and denitrification in a domestic steel plant are taken as the research object.

2. Overall scheme design

First of all, this paper in-depth understanding of activated carbon integrated desulfurization and denitrification technology process, system structure, purification principle. Through the analysis of desulfurization and denitrification system, the main variables and internal coupling relationship are determined. The identification model is established by recursive factor least square method. The identification model of desulfurization and denitrification of adsorption tower under the integrated process of activated carbon is established.

Secondly, based on the combination of neural network and PID control, an intelligent decoupling control algorithm, namely RBF online identification adaptive PID Decoupling control algorithm are designed [2]. The basic adaptive controller NNC is composed of a single neuron PID with self-learning and self-adaptive ability. The object model is identified online by RBF neural network identifier to improve the adaptive ability of the controller. The intelligent decoupling control algorithm is simulated to verify the robustness and feasibility of the decoupling control scheme combining neural network with PID control.

Finally, the network structure of the desulfurization and denitrification control system is described. The designed self-tuning PID Decoupling Algorithm Based on RBF model identification is programmed. The algorithm is programmed. The control effect of the algorithm and the integrity of the algorithm function block are verified by simulation [3].

3. Composition of desulfurization system

The flue gas desulfurization and denitrification system mainly includes three parts: flue gas system, activated carbon system and ammonia supply system.

3.1. Flue gas system

The flue gas system is the main part of the desulfurization and denitrification process. The sintering flue gas is firstly heated by the heater at the inlet. Then it was sent to the adsorption tower through the adsorption tower. It is also sent to the chimney channel through the booster fan at the tail, and it is the main exhaust fan. After the reaction treatment in directly discharged to the atmosphere [4].

3.2. Activated carbon system

The activated carbon system is mainly composed of three parts: adsorption tower, desorption tower and activated carbon transportation system. The main task of the analytical column is to separate the decomposable substances adsorbed in the activated carbon at high temperature, so that the activated carbon can be fully recycled. Activated carbon transportation system involves the addition and transportation of activated carbon in adsorption tower and desorption tower.

3.3. Ammonia supply system

The main components of ammonia supply system include liquid ammonia storage facilities and ammonia injection facilities. The liquid ammonia from the storage tank enters the evaporator by its own pressure and it is heated and it is evaporated into ammonia. Ammonia and air from the dilution fan are mixed and diluted in the air mixer. Next it was injected into the flue through the injection system for chemical reaction, and nitrogen oxides are removed.

4. Design of desulfurization system

4.1. Multivariable desulfurization system

The multivariable system structure diagram, which represents a multivariable system with N inputs and N outputs [5]. It is supplemented by task list module, distribution plan module, task analysis module and personal center module. The desulfurization and denitrification process control of adsorption tower are a typical dual input and double output multivariable control system, where r_1 and r_2 are the set values of desulfurization and denitrification, and y_1 and y_2 are the conversion values of desulfurization and denitrification respectively. Multivariable system structure diagram is shown in Fig.1.

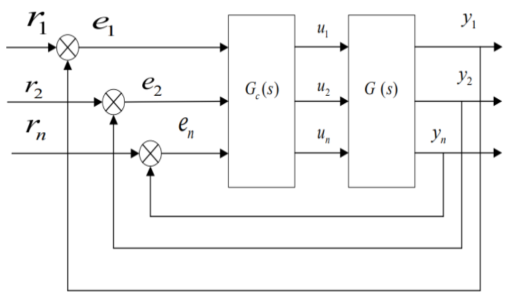


Fig.1. Multivariable system structure diagram

4.2. Analysis of denitration characteristics

Under the steady-state condition, the reaction conditions and the whole reaction process of the desulfurization and denitrification system are relatively stable. The relevant parameters in each reaction process have no obvious fluctuation, and the corresponding relationship between controlled variables is relatively clear.

The variation characteristics of generator power at a certain time, the load of sintering unit is stable in time period (0-t). It changes into transition state (t1-t2), and starts to stabilize at t2. It keeps stable for a period of (t2-t3). Finally it starts to change around t3 and restabilizes at t4.The actual steady-state diagram of desulfurization is shown in the Fig.2.

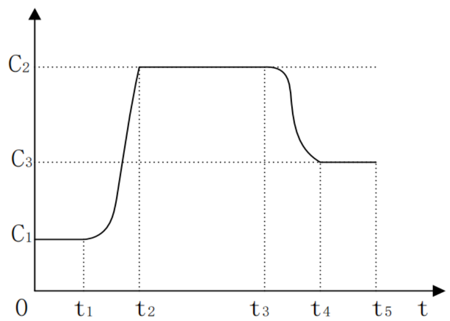


Fig.2. Actual steady state diagram of desulfurization

4.3. Data processing method

The actual process data contains a lot of industrial object information, so the amount of data collection should be as comprehensive as possible, so as to facilitate the later modeling and model verification. Data processing mainly includes data conversion and data error processing. Conversion includes scale, conversion and weight function. Conversion has a direct impact on the process accuracy, nonlinear mapping ability and numerical optimization algorithm.

4.4. Model of desulfurization and denitrification

System identification is to determine the model equivalent to the measured system according to the input and output data of the system. The purpose of system identification is to estimate the model structure and unknown parameters. It can be divided into two categories according to the model form, one is non parametric model identification method, and the other is parametric model identification method. This paper is based on the second model identification method, namely parameter model identification. The recursive factor least square method is used to identify the activated carbon desulfurization rate of the channel, and the identification curve is shown in the Fig.3.

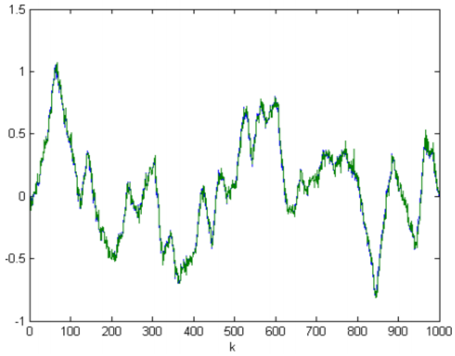


Fig.3. Comparison of fitting value and real value

4.5. Structure of neural PID control

The working principle of neural PID control is as follows: on the basis of online identification of controlled object by NNI, the weight coefficient of NNC is adjusted in real time to make the system adaptive and achieve effective control [5].

The adaptive PID control algorithm is as follows:

$$\Delta u(k)=K_p(k-1)x_{c1}(k)+K_i(k-1)x_{c2}(k)+K_d(k-1)x_{c3}(k)$$

A single neuron is used to construct PID controller, and its structure is shown in the Fig.4.

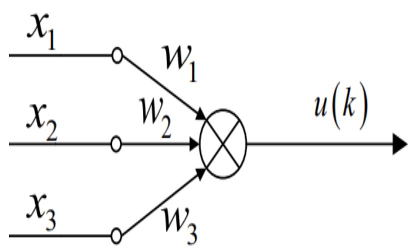


Fig.4. Single neuron PID controller

4.6. RBF adaptive model design

The basic idea of RBF neural network is that RBF as the "base" of hidden layer node constitutes the hidden layer space.n the case of over weight connection. The input vector is directly mapped to the hidden space. After the parameters of the neural network are determined, the nonlinear mapping relationship is determined.

The transformation from the input layer space to the output layer space is linear, and the output of the network is the linear weighted sum of the outputs of the hidden layer nodes. Neural network weights can be solved by various linear optimization algorithms. RBF neural network is a three-layer feedforward network, which is mainly composed of input layer, hidden layer and output layer [6]. The neuron structure of RBF neural network is shown in the Fig.5.below.

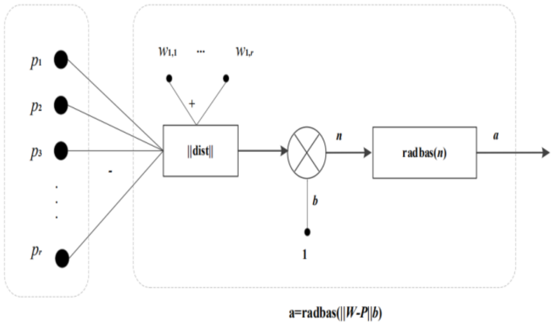


Fig.5. Neural structure of RBF neural network

5. Conclusion

This paper mainly analyzes the common methods and principles of decoupling, and uses RBF to identify adaptive PID online. Decoupling control is studied by control algorithm. In this paper, firstly, the classical feedforward decoupling controller is designed according to the identification model. Then, an adaptive single neuron PID identification based on RBF model is proposed [7]. The design and Simulation of intelligent decoupling controller are carried out to verify the effectiveness and superiority of the scheme.

Step signal is used to trigger the two channels at the same time, and R1 = 1 and R2 = 1 are set to simulate the triggering of the two channels. The simulation results are shown in the Fig.6.

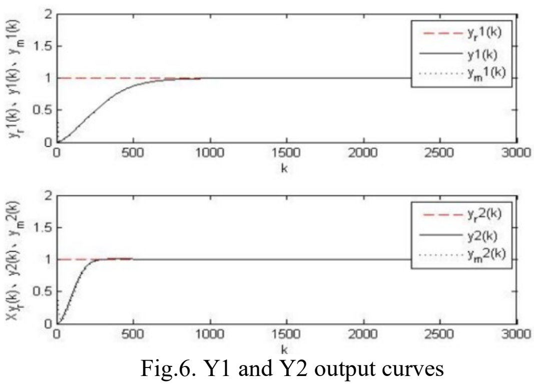


Fig.6. Y1 and Y2 output curves

According to the analysis of the above Fig.6, the two channels can respond quickly and achieve good steady-state results. The steady-state error is small. PID realizes the adaptive adjustment process. According to the image analysis, RBF realizes the online identification and adaptive adjustment of the model [8].

References

1. Li G, Wu Q, Wang S, et al. The Influence of Flue Gas Components and Activated Carbon Injection on Mercury Capture of Municipal Solid Waste Incineration in China. *Chemical Engineering Journal*, 2017, 326: 561-569.
2. Shi-He Chen, Hong-Ming He, Xiao-Feng Li, et al. Application of Fuzzy PID Controller with Self-Adaptive Algorithm and Non-uniform Grid Scheduling to WFGD. *IFAC Proceedings Volumes*, 2009, 42(9):20-25.
3. Zhang Y, Luan T, Yao Y. MtsPSO-PID Neural Network Decoupling Control in Power Plant Boiler. *IFAC Proceedings Volumes*, 2013, 46(20):101-105.
4. Sarah Sühnholz, Frank-Dieter Kopinke, Barbara Weiner. Hydrothermal Treatment for Regeneration of Activated Carbon Loaded with Organic Micropollutants. *Science of the Total Environment*, 2018, 644:854-861.
5. Yonggang Wang,Xinfu Pang,Zailin Piao,et al. Intelligent Decoupling PID Control for the Forc ed-circulation Evaporation System. *Chinese Journal of Chemical Engineering*, 2018, 23(12):2075-2086.
6. Shetty K, Srinath, R, Shivashankar, Shetty, Rashmi P. Use of Radial Basis Function Neural Network for Bearing Pressure Prediction of Strip Footing on Reinforced Granular Bed Overlying Weak Soil. *World Academy of Science Engineering & Technology*, 2012, 1(2):281-294.
7. Junwei Huang, Qingtai Xiao, Jingjing Liu, et al. Modeling Heat Transfer Properties in an ORC Direct Contact Evaporator Using RBF Neural Network Combined with EMD. *Energy*, 2019, 173:306-316.
8. Deepti Deshwal, Pardeep Sangwan, Divya Kumar. A Language Identification System Using Hybrid Features and Back-Propagation Neural Network. *Applied Acoustics*, 2020, 164:67-92.

A Survey of Low Visibility Image Enhancement Based on MSRCR Algorithm

Haokang Wen*, Hongbo Hao
Tianjin University of Science and Technology, China
*E-mail: * haokangwen@foxmail.com*
www.tust.edu.cn

Abstract

With the development of computer vision systems in the fields of traffic and safety monitoring, image enhancement has become an important research direction of computer vision. After development, the Retinex algorithm has a better enhancement effect, and the MSRCR algorithm has now developed into one of the important methods in image enhancement. This article introduces the development of the Retinex algorithm, and focuses on the main process and steps of the MSRCR algorithm, and uses the algorithm to enhance the low-visibility images of haze and night in multiple scenes. The actual results show that the MSRCR algorithm has a better enhancement effect and has a wide range of application values.

Keywords: image enhancement, MSRCR, Retinex algorithm

1. Introduction

Sometimes there will be some low visibility in daily life, such as the influence of haze caused by air pollution, and the situation of insufficient illumination at night, which will make the image acquired by the image acquisition system produce more obvious degradation phenomenon, such as low color saturation, poor definition of edge details, blurred image and so on. This will reduce the efficiency of later research and analysis. In order to improve the quality of image acquisition under low visibility and reduce the serious impact of fog weather on outdoor imaging system, many scholars and researchers have carried out a lot of research on image enhancement technology.

For the image defogging technology, the multi-scale information characteristics of the original image are usually taken as the main consideration factor, and the impurities in the non-sensitive area of the image are eliminated by compensating the contrast, brightness and color saturation of the image scene, so as to improve the visual effect of the image. At present, there are many algorithms for image enhancement, including global and local histogram equalization, wavelet transform enhancement and Retinex based on color constancy theory. This paper introduces the Retinex algorithm.

2. Development of Retinex Algorithm

2.1. The basic model of Retinex

Retinex is a commonly used image enhancement method based on scientific experiments and scientific analysis. It was proposed by Edwin H. land[1]. Retinex is a word made up of two words: retina and cortex. Land's Retinex model is based on three assumptions:

- The color we perceive is the result of the interaction between light and matter. The water we see is colorless, but the water film soap film is colorful, which is the result of light interference on the surface of the film.
- Each color region is composed of red, green and blue primary colors of a given wavelength.
- The primary colors determine the color of each unit area.

The basic theory of Retinex theory is that the color of an object is determined by its ability to reflect long wave (red), medium wave (green) and short wave (blue) light. Color uniformity is the basis of color uniformity. Different from the traditional linear and nonlinear methods, Retinex can achieve a balance in dynamic range compression, edge enhancement and color constancy. Therefore, Retinex can enhance various types of images adaptively.

The main steps are as followed:

According to the image formation model, an image can be expressed as:

$$I(x, y) = S(x, y) \cdot R(x, y) \quad (1)$$

(x, y) is the coordinates of the pixels in the image, "S" represents the incident light, "R" represents the reflection characteristics of the object, "I" is the reflected light, which is captured by the camera as an image.

Take the logarithm to separate the incident light component S and the reflected light component R. I'(x,y) is the logarithm of I(x,y).

$$I'(x, y) = \log[S(x, y) \cdot R(x, y)]$$
$$= \log\big(S(x, y)\big) + \log\big(R(x, y)\big) \quad (2)$$

Convolve the original image with a Gaussian template, which is equivalent to low-pass filtering the original image to obtain a low-pass filtered image D(x,y), where F(x,y) represents Gaussian filtering function:

$$D(x, y) = I'(x, y) \cdot F(x, y) \quad (3)$$

Subtract the low-pass filtered image from the original image to obtain the high-frequency enhanced image G(x,y).

$$G(x, y) = I'(x, y) - D(x, y) \quad (4)$$

In the previous steps, the incident light component S and the reflected light component R are separated, so the antilog of the resultant high-frequency enhanced image G(x,y) must be taken to obtain the enhanced image R(x,y).

$$R(x, y) = \exp\big(G(x, y)\big) \quad (5)$$

The contrast enhancement of R(x,y) is performed to obtain the final result image. As mentioned above, the flowchart of the Retinex algorithm is shown in Fig.1.

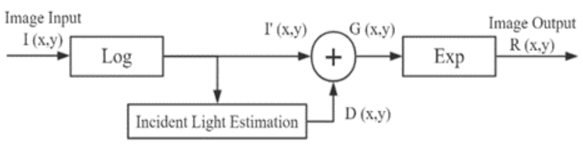


Fig.1 Flow chart of Retinex algorithm

2.2. SSR algorithm

Researchers such as Jobson proposed the single-scale Retinex (SSR) algorithm[2,3]. The specific formula is as follows:

$$R_i(x, y) = \log I_i(x, y) - \log[I_i(x, y) * F(x, y)] \quad (6)$$

In the formula, $R_i(x,y)$ is the output of Retinex in the "i" color spectrum, $I_i(x,y)$ is the image distribution, that is, the brightness value at the position (x,y). * represents the convolution operation, F(x,y) is a wraparound function that is defined by equation (7).

$$F(x, y) = K \cdot e^{\frac{-(x^2+y^2)}{\sigma^2}} \quad (7)$$

Among them, σ is the wrapping scale, K is the normalization constant. The wrapping function satisfies:

$$\iint F(x, y)dxdy = 1 \quad (8)$$

The stronger the dynamic compression capability of SSR is, the better the details of the dark part of the image can be enhanced, but the color distortion of the output image is more serious.

2.3. MSR algorithm

MSR is developed on the basis of SSR algorithm[4]. The specific formula of MSR is as followed:

$$R_i(x, y) = \sum_{k=1}^{K} w_k \{\log I_i(x, y) - \log[I_i(x, y) * F(x, y)]\},$$
$$i = 1, \ldots, N \quad (9)$$

Among them, i represents the i-th color channel, and (x,y) represents the coordinates of the pixel in the image. N is the number of color channels in the image. N=1 represents a grayscale image, N=3 represents a color image, $i \in (R,G,B)$. $I_i(x,y)$ represents the i-th color channel in the input image, $R_i(x,y)$ represents the output result of the MSR of the i-th channel. F(x,y) is a Gaussian function, k represents the number of Gaussian surround functions or the number of surround scales, w_k represents the weight related to the Gaussian function, $\sum_{k=1}^{K} w_k = 1$, in general, MSR takes three scales of high, medium and low, K=3. As shown in the Fig.2, it is the flow chart of MSR algorithm.

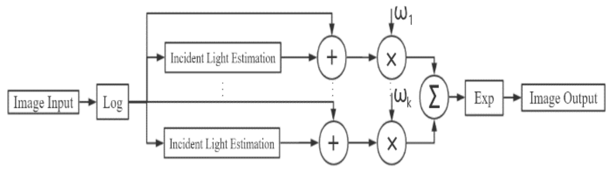


Fig.2 Flow chart of MSR algorithm

2.4. Basic theory of MSRCR algorithm

The algorithm with color recovery factor originated from Retinex theory, which adds a color recovery factor to the multi-scale algorithm[5,6].

$$\begin{cases} R_i^{MSRCR}(x,y) = C_i(x,y)R_i^{MSR}(x,y) \\ C_i(x,y) = \beta\log[\alpha I_i'(x,y)] \\ I_i'(x,y) = I_i(x,y)/\sum_{i=1}^{3} I_i(x,y) \end{cases} \tag{10}$$

Among them, $i \in (R, G, B)$, C_i is the color recovery coefficient of the i-th channel, which is the key factor, β is the gain constant, α is the controlled nonlinear enhancement coefficient, R_i^{MSRCR} is the output image of the i-th channel processed by this algorithm.

The MSRCR algorithm is based on MSR algorithm, adding the adjustment factor C_i which considers the ratio of R, G and B in the original image. Because the color recovery factor has color compensation for the image enhanced by MSR algorithm, the image enhanced by MSRCR can better maintain the original color without distortion.

3. Test Result

The main purpose of image enhancement is to suppress the useless noise in the process of image acquisition and transmission, highlight the useful information in the image, make the image conform to the visual effect of human eyes as much as possible, or make the image transform into a form that is conducive to computer recognition and analysis, and improve the subsequent processing ability and application value of the image.

In this paper, under the Windows 10, based on MATLAB 2015a software, the MSRCR algorithm is used to process the fuzzy images in haze and night. as shown in the Table 1 below, it is the computer environment.

Table 1. Computer environment

GPU	NVIDIA GeForce MX150
Video memory	2 GB
Operating system	Windows 10
Software version	MATLAB R2015a

As shown in the Fig.3 below, the image on the left is the original image, and the image on the right is enhanced by this algorithm. The image in haze weather is processed, and the details of the enhanced image are more obvious. In particular, the color of the object in the original image is grayish white, but after enhancement, the color is more vivid.

Fig.3 Comparison of haze image enhancement effect

As shown in the Fig.4 below, the edge outline and details of the object in the original image are very fuzzy and difficult to distinguish. The enhanced image on the right has better visual effect and clearer object contour after processing the image at night.

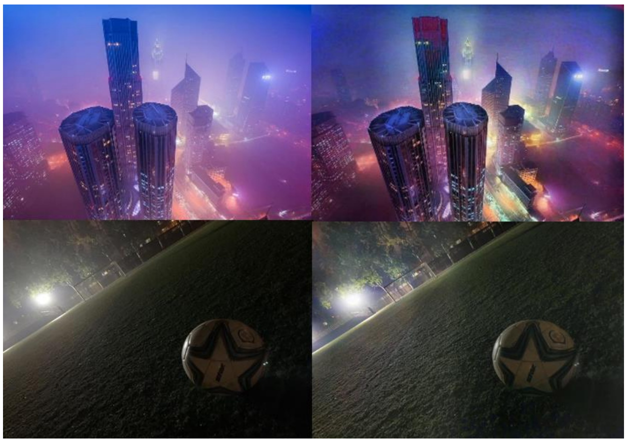


Fig.4 Comparison of night image enhancement effect

As shown in the two figures, it can be seen from the comparison that, although there are halos in some areas of the images, the overall enhancement effect is satisfactory due to the clearer details and brighter colors of the objects.

4. Conclusion

The theory and development of Retinex algorithm are introduced in this paper. Especially the process and method of SSR, MSR, MSRCR algorithm are introduced. At the end of the article, the MSRCR algorithm is used to for practical test and the experimental results are remarkable and satisfied.

The test results show that the enhancement effect of this method is obvious and has certain practicability. Retinex image enhancement algorithm based on human visual system has the advantages of color constancy, image sharpening and high fidelity, which cannot be compared with traditional enhancement algorithm. The color information of the enhanced image can be restored or maintained well.

References

1. Land E H. The retinex theory of color vision. *Scientific american*, 1977, 237(6): pp. 108-129.
2. Jobson D J, Rahman Z, Woodell G A, "Properties and performance of a center/surround retinex," *IEEE Transactions on Image Processing*, 1997, 6(3): pp. 451-462.
3. Hines G, Rahman Z, Jobson D, et al. Single-scale retinex using digital signal processors. 2005.
4. Rahman Z, Jobson D J, Woodell G A. Multi-scale retinex for color image enhancement//*Proceedings of 3rd IEEE International Conference on Image Processing. IEEE*, 1996, 3: pp. 1003-1006.
5. Jobson D J, Rahman Z, Woodell G A. Retinex image processing: Improved fidelity to direct visual observation//Color and Imaging Conference. *Society for Imaging Science and Technology*, 1996, 1996(1): pp. 124-125.
6. Girija M G, Shanavaz K T, Ajith G S. Image dehazing using MSRCR algorithm and morphology based algorithm: A Concise Review. *Materials Today: Proceedings*, 2020, 24: pp. 1890-1897.

Design of a Taxi Meter Based on Single Chip Computer

Haokang Wen *, Min Wang

Tianjin University of Science and Technology, China

*E-mail: * haokangwen@foxmail.com*

Abstract

As a convenient means of transportation, taxis greatly facilitate people's daily travel. Based on such a background, this article designed a simple taxi meter based on STC89C52 microcontroller. This taximeter uses a single-chip microcomputer as the core, combines display module, clock module, storage module, key module, and combines software and hardware to build the required taxi meter model. The taxi meter introduced in this article has the functions of displaying time, starting price, and real-time display of the cost according to vehicle mileage. Theoretically, the meter has higher accuracy and has a wider practical value in daily life.

Keywords: Taxi meter, single chip microcomputer, STC89C52

1. Introduction

With the continuous progress of society and the rapid development of science and technology, people's living standards are developing rapidly. As one of the modes of transportation, taxi develops rapidly. People have become more and more dependent on the convenient, fast and comfortable services provided by taxis. Taxis have gradually penetrated into our lives and gradually developed into an indispensable part of our social services. More and more people prefer taxis to travel.

The taximeter is one of the most important parts of taxis, which plays an important role. The most important feature of taxi meter is to record the mileage calculation cost. Only by ensuring that its core function is recognized by passengers, can the industry operate well. At the same time, the taximeter serves the taxi, and constantly improving its function to better serve passengers will be the constant pursuit of the whole taxi industry. This paper will use STC89C52 MCU to design a simple and accurate taximeter.

2. Overall Design Ideas

The taximeter designed in this paper is composed of STC89C52 single chip microcomputer, AT24C02 memory chip and LCD1602 LCD display, including the main program module, power module, data display part, clock part, storage unit design, key module, etc. The meter can control the charging status of taxis by pressing buttons.

Firstly, different mileage corresponds to different charging modes. Secondly, it has different charging prices for day and night. At the same time, it also adds a charging standard of waiting state during driving. LCD1602 also displays real-time data such as time, distance and price. In addition, the taximeter designed in this paper not only meets the charging function, but also increases the practical function of carpool and special car, so that the taximeter can meet more needs of consumers and has better practicability.

3. Hardware Design

3.1. STC89C52

STC89C52 is the core part of this taximeter. This microcontroller has a wide range of applications, such as a design of intelligent led lighting systems[1], and a design of laboratory fire alarm system[2]. And it can be also used in a gain controllable system in RF receiver[3]. STC89C52 is an 8-bit microcontroller with low power consumption and high performance produced by STC company. STC89C52 uses the classic MCS-51 core, on this basis, the chip has done a lot of improvement. In this paper, STC89C52 MCU is used to solve the mileage recording, billing calculation and other functions.

The main features are as follows:

- Working voltage: 3.4V-5.5V.
- Working frequency: 0-40MHz.
- The instruction code of traditional C51 MCU is completely compatible.
- 512BS data memory can be integrated.
- The chip contains E2PROM function.
- There are three 16 bits timers / counters compatible with MCS-51 single chip microcomputer timers.

3.2. LCD1602

In many electronic devices, LCD1602 will be used as a display module[4]. For example, the display interface of multimeter, calculator and so on uses liquid crystal display, which will present the required digital data through the LCD module. As shown in the Fig.1, it is the LCD display module in this paper.

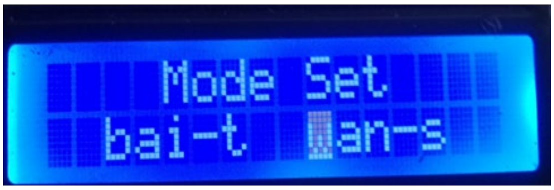


Fig.1 LCD module

In this paper, LCD1602 display module is used to display the mileage, price, time and other data in the process of driving. It is the display module of this design and undertakes the important function of displaying data.

As shown in the Fig.2, the circuit diagram of LCD display module.

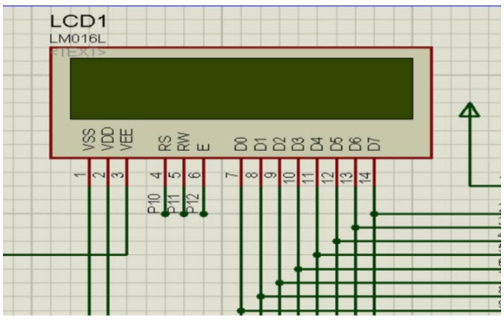


Fig.2 LCD circuit diagram

The main performance of LCD1602:

- The working voltage of the chip: 4.5-5.5V
- Capacity: 16 × 2 characters
- The best working voltage: 5.0V
- Working current: 2.0mA

3.3. DS1302

In this paper, DS1302 chip is used to realize the function of real-time display time. It has the characteristics of low power consumption and high performance. It is a kind of real-time clock circuit, which can display the date, time, minute and second accurately. In the daytime and at night, the charging standard of Taximeter is different, and time is the most basic way to distinguish day and night. It is very important to display the time in real time through DS1302 chip for the charging link in the process of driving. As shown in the Fig.3, it is DS1302 chip.

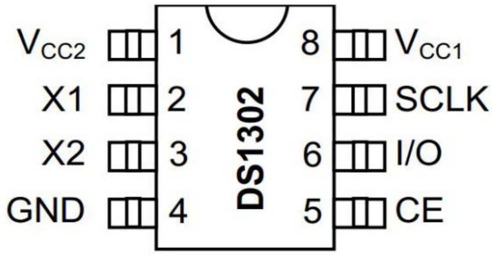


Fig.3 DS1302 pin diagram

Pins and functions of DS1302:

- Pin 1: VCC2 is the power supply pin.
- Pin 2: X1 is a 32.768KHz crystal oscillator pin.
- Pin 3: X2 is also a 32.768KHz crystal oscillator pin.
- Pin 4: GND.
- Pin 5: RST is the reset pin.
- Pin 6: I / O is the input / output pin of data.
- Pin 7: SCLK is the serial clock pin.
- Pin 8: VCC1 is the power supply pin.

3.4. Key design

For the taximeter, setting the charging standard is a very important link. This paper sets the charging standard by pressing the button. In the system designed in this paper, a total of six buttons are designed.

The first is the confirm key. By pressing this key, you can enter the menu to set the parameters. No matter which setting you want to change, you must press this key before entering the page to make corresponding changes.

The second is the add key, which can be used to add variables when setting parameters. For example, it can be used to accurately adjust the time and unit price. When the time is slow or the unit price is set low, you can use this key to speed up the time or increase the unit price.

The third key has two functions. It can be used as a minus key and as a start or pause key for managing waiting time. Waiting time refers to the time that passengers spend when they want to get off and do something in the process of driving. There is a charging standard for waiting time. The minus key can be used to adjust the time and unit price.

The fourth is exit key: Exit menu setting key. After each setting, you can press this key to realize an exit function.The fifth is the reset key, which is to reset it after the end of the previous trip, so as to start a new mileage charging.

The sixth key is used to simulate the increase of mileage, that is, to simulate and record the driving process of a taxi. Pressing this key is equivalent to a taxi driving 0.1km.

As shown in the Fig.4, it is the circuit diagram of the key part.

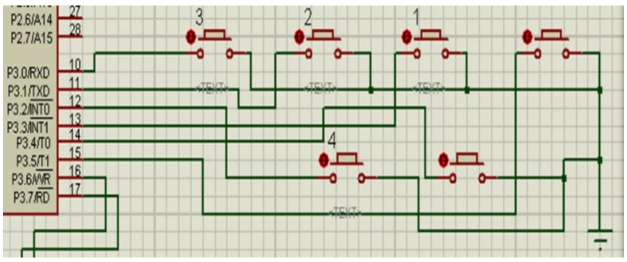


Fig.4 Key module circuit diagram

3.5. AT24C02 memory chip

The memory chip AT24C02 of ATMEL firm was chose as the memory cell for losing electricity[5]. AT24C02 memory chip is responsible for the realization of power-off protection in this design. It is an electrically erasable memory chip E2PROM with I^2C bus interface. This kind of chip can save data when power is off.

The I^2C bus can be used to communicate with single chip microcomputer and save the working data. In the process of driving, if there is power failure, there is no need to be nervous, because AT24C02 memory chip has saved the data of the previous section, once powered on, it can be recovered, which has great practical significance for the taximeter.

Pins and functions of AT24C02 are shown below:

- Pins 1-3: A0 -A2 are device address selection pins.
- Pin 4: GND, ground.
- Pin 5: SDA, serial data input and output port.
- Pin 6: SCL, serial shift clock control terminal.
- Pin 7: WP, hardware write protect control pin.
- Pin 8: VCC, connected to + 5V voltage.

As shown in the Fig.5, it is AT24C02.

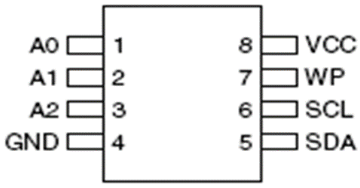


Fig.5 AT24C02 pin diagram

4. Simulation

In this paper, software Proteus is used for theoretical simulation. Proteus has a rich library of components, and can easily create components. When searching for components, we can accurately select the required devices according to little information. It has powerful function of schematic drawing. The MCU simulation and spice circuit simulation are combined. It supports the simulation of mainstream SCM system. It provides software debugging function. As shown in the Fig.6, it is a circuit diagram simulated by Proteus.

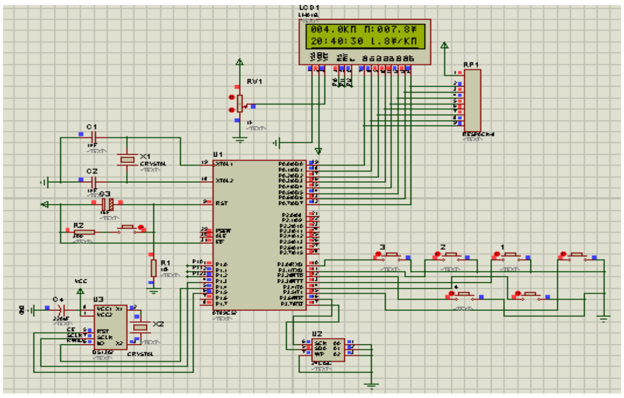


Fig.6 Simulation diagram

5. Software Design

In this paper, the design of software can be divided into several steps: main program design, timing interrupt service program, mileage counting interrupt service program, midway waiting interrupt service program, display subroutine service program, etc. these programs are combined with MCU, key circuit, LCD1602 display circuit and DS1302 clock circuit involved in hardware circuit. This paper uses the program to drive the circuit to achieve the design function of the taxi meter, and finally presents a simple but accurate taxi meter.

6. Conclusion

After the actual test, the function of the taximeter designed in this paper can be realized and the expected effect can be achieved. The simple taximeter designed in this paper can meet various scenes of daily life after actual test. In theory, the meter has high precision and wide application value.

References

1. Wang T, Chen T, Hu Y, et al. Design of intelligent LED lighting systems based on STC89C52 microcomputer. *Optik*, 2018, 158: pp.1095-1102.
2. Zeng X. Design of Laboratory Fire Alarm System Based on MCU. *International Core Journal of Engineering*, 2020, 6(6): pp. 269-272.
3. Dai F, Wen H. A Design of Controllable Gain Amplification System Based on Single Chip Computer.
4. Wen H, Dai F, Zhao J, et al. Design of a portable instrument for measuring heart rate and blood oxygen. *Proceedings of International Conference on Artificial Life and Robotics*, 2020, 25: pp.682-685.
5. Zhao Y R, Wang D S, Li J D. Design of the Function Units of Electronic Password Lock//*Applied Mechanics and Materials*. Trans Tech Publications Ltd, 2012, 220: pp. 2008-2011.

Design of a High Precision Digital Clock Based on Single Chip Microcomputer

Haokang Wen*, Qiang Zheng

Tianjin University of Science and Technology, China

*E-mail: * haokangwen@foxmail.com*

www.tust.edu.cn

Abstract

Compared with traditional mechanical clocks, digital clocks have higher accuracy and durability and are widely used in people's daily lives. This paper designs a digital clock which uses AT89S52 single-chip microcomputer as the main control chip. It can count hours, minutes, and seconds, and can calibrate the time. It can switch between 24-hour and 12-hour systems. As a smart clock, while displaying the time, it also adds the function of temperature display. The intelligent digital clock designed in this paper has stable performance in theory and has certain practical value.

Keywords: digital clock, DS18B20, single-chip microcomputer

1. Introduction

Digital clocks have become electronic necessities in people's daily life. For example, they have been widely used in personal family life, subway transportation stations, movie theaters, cultural offices and other public places, which greatly facilitate people's life, study, work and entertainment. Compared with the traditional mechanical clock, the accuracy and intuitiveness of the digital clock have been improved. Because it has no mechanical device, the service life of the digital clock is relatively long and has been widely promoted.

The general trend of digital clock development in the current era is gradually moving closer to higher accuracy, smaller size, more functions, and lower power consumption. In this context, the degree of digitization and precision of the clock has become the dominant design direction of the current clock development and progress. The hardware system includes 5 independent touch switch buttons and an LCD liquid crystal display, which can display necessary information. According to the needs of the user, the time can be further calibrated, selected time, and temperature display at any time. In addition, it can convert between 12-hour and 24-hour formats.

2. Overall Design

The digital clock designed in this article mainly includes two parts: hardware and software. The hardware part is the basis for the completion of the operation of the designed system, and the software part effectively and fully supports and utilizes the hardware of the system to meet the functional requirements of the designed system. As shown in the Fig.1, it is the schematic diagram.

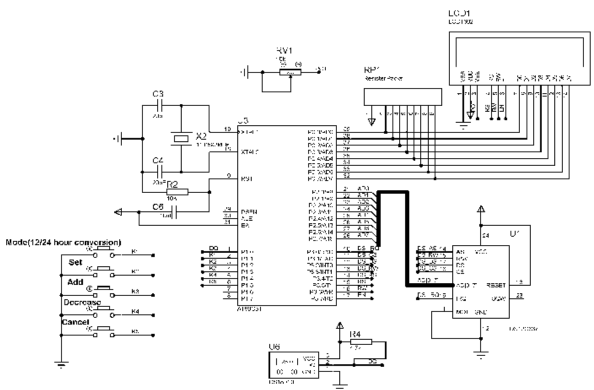


Fig.1 The schematic diagram

The digital clock designed in this paper uses AT89S52 single-chip microcomputer as the main control chip[1], the clock module uses DS12C887 as the clock chip, the temperature sensor DS18B20 chip as the temperature module, the LCD1602 as the display module, and the button control part uses the touch switch button circuit.

In theory, the electronic clock designed in this paper has high practicability and stable performance. The software part of this system runs under the Keil µVision4 software platform environment, and implements specific functions with a single-chip C language program.

3. Hardware

3.1. AT89S52

AT89S52 is a member of MCS - 51 family[2]. It is a low-voltage, high-performance CMOS 8-bit single-chip microcomputer produced by Atmel in the United States. It contains 8KB of program memory that can be written repeatedly and 256 bytes of data memory (RAM). The device is produced using Atmel's high-density, non-volatile storage technology, compatible with the standard MCS-51 instruction system, and is equipped with a general-purpose 8-bit central processing unit (CPU) and Flash storage unit on-chip. As shown in the Fig.2, it is the AT89S52 chip pin diagram.

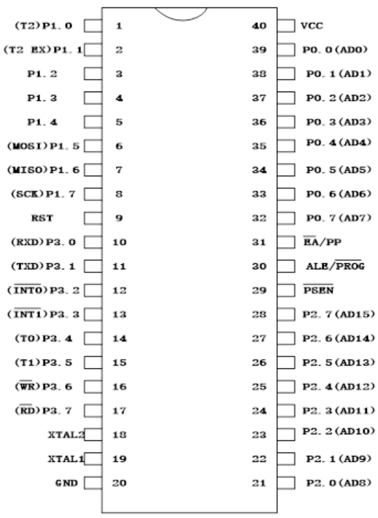


Fig.2 AT89S52 pin diagram

The powerful AT89S52 single-chip microcomputer can be flexibly applied to various control fields. AT89S52 single-chip microcomputer is an enhanced type of AT89C51 single-chip microcomputer, and has extremely high compatibility with Intel's 80C52 series chips in terms of pin arrangement, hardware composition, working characteristics and instruction system. In the system designed in this article, the internal clock mode and the button level reset circuit are selected to form the smallest circuit of the single-chip microcomputer.

The main operating characteristics of AT89S52 are as follows:

- 8k byte independent programmable Flash storage.
- Eight interrupt sources.
- Fully compatible with MCS-51 series single-chip products.
- Fully static command operation: 0Hz～33Hz.
- Three-level encrypted program memory.

3.2. DS12C887

The DS12C887 clock chip is a real-time clock/calendar chip launched by the Dallas Semiconductor Company of the United States[3]. It uses a parallel interface and is made of CMOS technology. It has a crystal oscillator and a lithium battery backup for the clock chip. It is compatible with the pins of MC146818 and DS12887, which are commonly used in daily computers, and can be directly replaced. The clock circuit designed with the DS12C887 clock chip does not require any peripheral circuits and devices, and has a good microcomputer interface. The DS12C887 clock chip has the advantages of low power consumption, relatively simple peripheral interface, extremely high accuracy of time travel, and stable and reliable operation. It is widely used in various real-time clock systems that require higher precision. As shown in the Fig.3, the connection circuit diagram of the DS12C887 clock chip and the microcontroller.

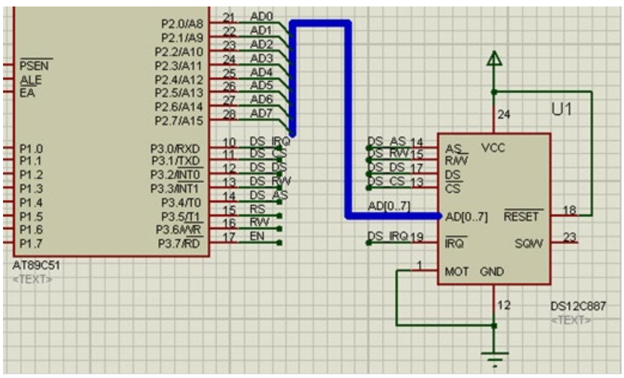


Fig.3 The DS12C887 connection circuit diagram

3.3. DS18B20

DS18B20 temperature sensor is an improved intelligent temperature sensor launched by DALLAS Semiconductor Corporation of the United States. Compared with the traditional thermistor and other components, the sensor can directly read the temperature of the environment where the sensor is currently located, and can achieve 9-12 digital value readings through simple programming according to actual performance requirements[4]. The temperature of the environment is directly used for data transmission in a one-wire bus, which greatly improves the anti-interference characteristics of the sensor itself. This chip is suitable for temperature measurement in harsh environments, such as temperature measurement consumer electronic products of various scales. As shown in the Fig.4, it is the interface circuit diagram of DS18B20 temperature sensor.

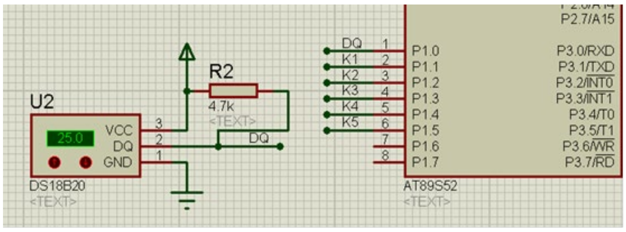


Fig.4 DS18B20 circuit diagram

The performance characteristics of this chip are as follows:

- The unique single-wire interface requires only one port pin for inter-chip communication.
- The temperature measurement range is -55℃ to 125℃, and the maximum resolution can reach 0.0625℃.
- The 3-wire system is connected to the single-chip microcomputer, which reduces the use of external hardware equipment.
- It can be powered by the data line, and the voltage range is 3.0V-5.5V.
- Negative voltage characteristics, when the polarity of the power supply is reversed, the thermometer will not burn out due to heat, but it will not work normally.

3.4. LCD1602

This LCD1602 character liquid crystal display module is a dot matrix liquid crystal module specially used to display letters, numbers, symbols, etc. It is composed of several 5×7 or 5×11 dot matrix character bits. Each dot matrix character bit can display a character. There is a dot pitch between each bit, and there is also a gap between each line that plays the role of character spacing and line spacing, because of this, it cannot display graphics. The 1602LCD display means that the displayed content is 16×2, that is, it can display two lines, each line has 16 characters LCD modules.

The characteristics of LCD1602 are as follows:

- +5V voltage.
- Built-in reset circuit.
- Provide various control commands, such as: clearing the screen, blinking characters, blinking cursor, display shift and other functions.
- There are 80 bytes display data memory DDRAM.
- Built-in character generator CGROM with 192 5×7 dot matrix fonts.
- With user-definable 5×7 character generator CGRAM.

4. Software

The entire software system in this paper uses C language programming. The high-precision digital clock software designed adopts a modular structure and mainly realizes the following functions:

- LCD function realization.
- Time data collection.
- The realization of 12/24 hours conversion.
- Temperature collection.
- Key recognition processing.

5. Conclusion

This text design uses AT89S52 as the main control single-chip microcomputer, the clock module chooses DS12C887 as the clock chip, the temperature module chooses DS18B20 as the temperature sensor, LCD1602 as the display module, and the button circuit for the setting part.

The high-precision digital clock based on the single-chip microcomputer in this design realizes the counting of seconds, minutes and hours, and can switch between the 24-hour system and the 12-hour system. After testing, the system has achieved the goal of aligning

the minutes and hours at any time, and the timing error is less than 0.01s/day.

As shown in the Fig.5, it is the final test result.

Fig.5 Test result

The key functions are as follows:

- K1: Digital clock 12/24 hours conversion button.
- K2: Setting button, used to adjust the hours, minutes, and seconds of the time.
- K3: "+" button, the number increases by 1.
- K4: "-" button, the function of reducing the number by 1.
- K5: Cancel button, cancel the current operation, return to the main page (time display page).

The content displayed on the LCD1602 display includes precise hours, minutes, seconds, and current ambient temperature.

After actual testing, the digital clock designed in this article has high performance and certain usability.

References

1. Wu H, Tang M, Huang G. Design of multi-functional street light control system based on AT89S52 single-chip microcomputer//*2010 The 2nd International Conference on Industrial Mechatronics and Automation.* IEEE, 2010, 1: pp.134-137.
2. Munir M M, Suhendi A, Thaha H, et al. An AT89S52 microcontroller‐based single board computer for teaching an instrumentation system course. *Computer Applications in Engineering Education*, 2007, 15(2): pp.166-173.
3. Ding Y, Yao Y. Illuminance Test System Based on MSP430//*2010 2nd International Conference on Information Engineering and Computer Science*. IEEE, 2010: pp.1-3.
4. Wen H, Dai F, Zhao J, et al. Design of a portable instrument for measuring heart rate and blood oxygen. *Proceedings of International Conference on Artificial Life and Robotics*, 2020, 25: pp.682-685.

Design of Manchu Image Acquisition System Based on STM32

HongBo Hao [1,] *, Fulin Zheng [2]

[1] Tianjin University of Science and Technology, China;

[2]Dalian Minzu University, Liaoning,China

E-mail: 2844011096@qq.com

www.tust.edu.cn

Abstract

This paper designs a Manchu image acquisition system based on *STM32* to protect and utilize Manchu ancient books. The system uses *STM32F407* as the core controller, with the help of image sensor, *LCD* screen, *and SD* storage module to complete the function of image acquisition, storage and upload of Manchu ancient books. This paper describes in detail the hardware circuit design, software programming and overall function debugging of the image acquisition system. In order to achieve the purpose of digital protection of Manchu ancient books, we can collect, display, save and upload images in different situations.

Keywords: Manchu, *STM32F407*, *SD* storage module, image acquisition system

1. Introduction

Manchu is also used to write Manchu characters. Manchu not only created Manchu characters, but also used Manchu characters to record a large number of customs historical stories. Therefore, Manchu has a very important position in the history of Chinese characters. Manchu ancient books also have important historical and cultural value. With the change of time, the preservation of Manchu ancient books is becoming more and more difficult and the researchers' research is more and more inconvenient. In order to facilitate the research of researchers, at the same time better preservation of Manchu ancient books, with the development of time, we need to use more advanced technology to protect ancient books. Considering the present situation of science and technology development, design cost and people's use mode, the first choice of literature protection method is to digitize ancient books. In this paper, we choose the Manchu image acquisition system based on *STM32* to carry out the research. That is to use the ancient books digital technology to digitize the Manchu ancient books, so as to achieve the purpose of protecting and using the ancient books [1].

For the protection and utilization of Manchu ancient books to make a certain contribution. The research method of this topic can be applied to more ancient books protection work. After the digital processing of ancient books, it can save a lot of paper resources, reduce the generation of waste and reduce the felling of trees. It plays a certain role in protecting the environment. At the same time, digital preservation can save a lot of human resources, and the transmission speed is faster, realizing the coordinated development of economy, environment and society. Based on the embedded processor platform, using *STM32* core controller and *CMOS* camera, this paper designs an image acquisition device to realize the acquisition, storage and display of Manchu images in ancient books.

2. Overall scheme design

In this paper, *STM32F407* chip is used as the core controller of Manchu image acquisition system based on *STM32*. Image acquisition module, image saving module, image display module and key module are controlled by different *IO* ports to realize image acquisition, storage display. At the same time, the data transmission module completes the connection with the upper computer.

According to the designed system scheme, the camera image acquisition module, image storage module, key control module and TFT liquid crystal display module are designed and built. After completing the circuit design, according to the relevant theory of image acquisition and processing and the camera control method, the system software design and overall debugging are completed. The function of collecting, displaying, saving and uploading Manchu images in ancient books is realized through the key control module [2].

3. Hardware circuit design

3.1. Design of core controller

In this paper, the core controller is used to receive the images collected by the image sensor and transmit them to *TFTLCD* for display. According to the research requirements, *STM32* series chips are selected. *STM32F407* is selected as the core controller to complete the design. *STM32F407* has a great improvement compared with *STM32F1* series. First of all, the speed of *FSMC* is very fast. The screen swiping speed of *F407* can reach 3000W pixels per second, while *F103* is only about *500W*. When using the image sensor with *200W* pixels, *F103* has a stuck phenomenon, and *F407* can be perfectly processed.

3.2. Touch the capacitor button

This design uses the capacitive touch screen to complete, so we plan to use a capacitive touch button as the return key of the system. When the system enters a function module, touch the button to return to the upper level. The circuit diagram is shown in the Fig.1 below.

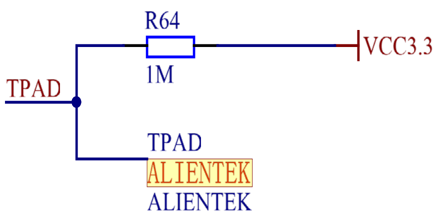


Fig.1. Capacitive touch button path diagram

3.3. Image acquisition module

In this paper, through the analysis and comparison of the performance of a variety of image acquisition sensors, for *CCD* sensor, the photoelectric conversion of *CMOS* sensor is very direct and the efficiency is higher. With the development of microelectronic technology, the quality of *CMOS* imaging is getting better and better, and the cost is lower than *CCD*. To sum up, this design uses *CMOS* camera as the image sensor of this study.Through the study and research of *CMOS* image sensor, the circuit diagram design of image acquisition module is completed according to the function of different interfaces, as is shown in Fig.2 below.

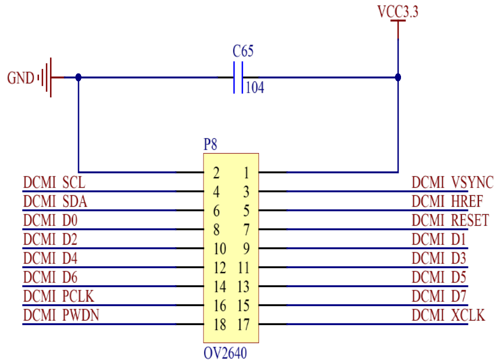


Fig.2. Image acquisition module circuit

In the interface of capacitive touch screen, *CS* is the chip selection signal pin, which is effective at low level. *RS* is the data command signal pin, *0* is the data and *1* is the command. *WR* is the write enable signal, which is valid at low level. *Rd* can read the pin effectively. *RST* is a reset signal, which can control the reset of capacitive touch screen. *CS* is the reset signal of capacitive touch screen. According to the function of each pin, the capacitor touch screen can be controlled to complete the required operation [3].

3.4. Storage module

The research content of this paper is Manchu image acquisition, so the preservation of collected image is very important. After comparison and research, this paper uses an *8G SD* card as the storage module of Manchu image acquisition system. The circuit diagram is shown in Fig.3.

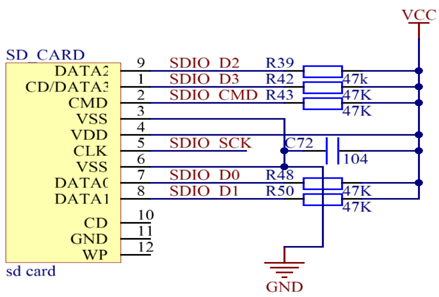


Fig.3. Circuit diagram of image storage module

As shown in the Fig.3, the *SD* card has *6* communication lines and *3* power lines. The *CMD* is a command line and a two-way signal line. *DAT0, DAT1, DAT2* and *DAT3* are four data lines, which are also bidirectional signal lines. The transmission rate of each data line is *25 Mbit* per second. The maximum is that four data lines read data at the same time, and the maximum can reach *100 Mbit*.

3.5. Data transmission module

The data transmission of this design is to transmit the collected image to the upper computer for saving or browsing. After comparing the speed and stability of serial communication, wireless communication and *USB* communication, *USB* communication is selected as the data transmission mode between Manchu image acquisition system and upper computer. *USB* transmission speed is faster and more stable, when transferring large files such as pictures, it is faster than serial transmission and Bluetooth transmission. *WiFi* transmission needs to install *WiFi* module and specific host computer receiving. *USB* transmission does not need a specific host computer software. It can be convenient for data transmission with most computers. The standard of *USB* is four buses, *VCC* and *GND*, and *D* + and *D* -. Because *USB* cannot be directly connected with *STM32*, the circuit diagram of this design is shown in Fig.4 below [4].

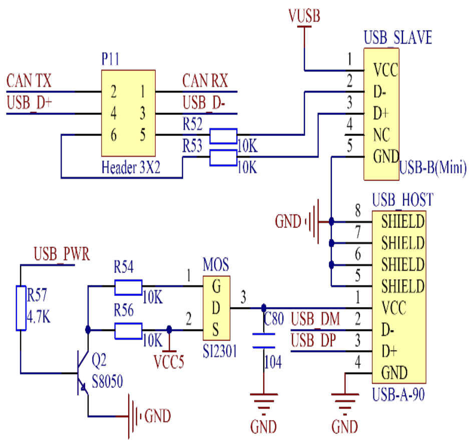


Fig.4. *USB* transmission circuit diagram

4. Software design

4.1. Main program flow chart

The program design of this chapter is based on Keil software and C language. Finally, the program design of Manchu image acquisition system is completed. In the main. C main program, the system initialization is to create a system_ Int function [5]. The flow chart of the whole is shown in Fig.5 below.

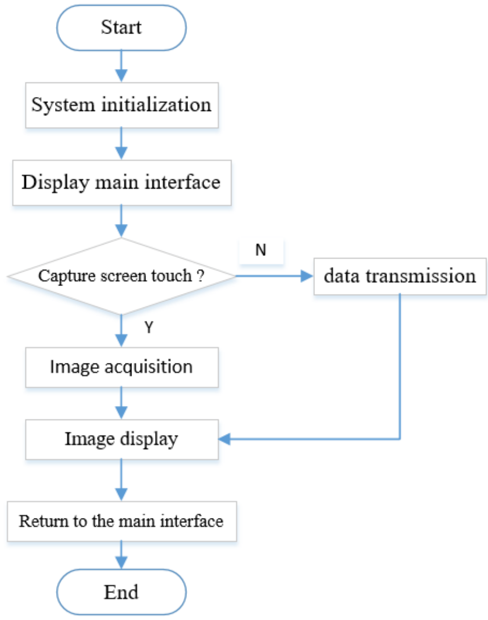


Fig.5. Main program flow chart

4.2 Main program flow chart

After the initialization, the initial interface is displayed. The initial interface mainly shows the English name of the research content in this paper and the detection results of whether each module is normal. It includes whether the icon image of *APP* exists, whether the screen is normal. The main interface is designed to display three main function icons, which are image acquisition, image display and *USB* transmission. Then the system detects whether there are touch screen events. If the three icons click events, it will enter different modules according to the click events.

5. Conclusion

5.1. System

There are many problems in the testing process of image acquisition. In the first acquisition, *OV7725* image acquisition sensor is used, which is a *30W* pixel sensor. It is found that the collected images cannot be displayed clearly. After the teacher's help, the *OV2640* image sensor was re selected to complete the image acquisition.

It is the acquisition interface of Manchu image acquisition system, as is shown in the Fig.6 below.

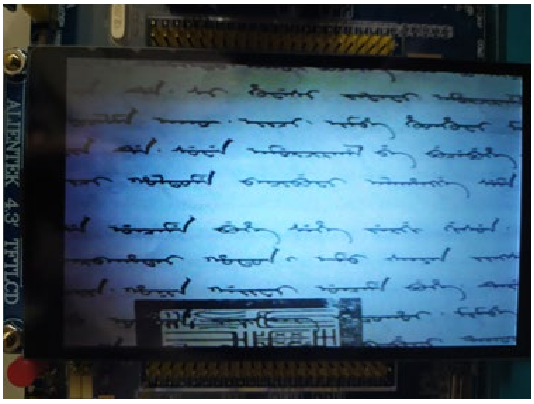


Fig.6. Manchu image acquisition and testing

After image acquisition, the image display module needs to be tested. First, the main interface of image display is shown in the Fig. below. After entering the main interface, select the image you want to view and click. Click backward to enter the mode of automatic play. Automatically play the first picture you click on. When playing automatically, click the middle area of the screen to pause the play, and click again to resume the play. Click the top and bottom of the screen to control the up and down movement [6]. BMP format image display is shown in the Fig.7 below.

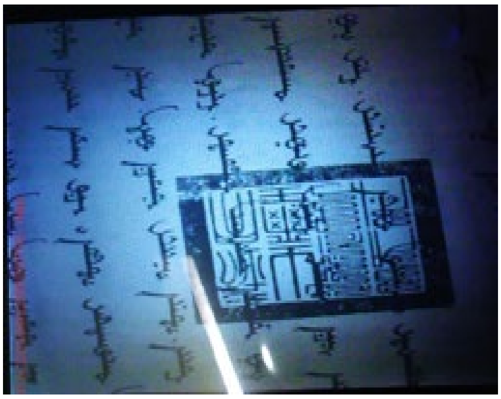


Fig.7. BMP format image display

This paper compares the advantages and disadvantages of various data transmission methods, and uses *USB* data cable to connect with the upper computer. According to the *USB* communication protocol, a kind of analog USB flash disk is realized. Click the data transmission icon, the system will enter the state of analog *U* disk and communicate with the upper computer. Complete the system function of data transmission.

References

1. Shuang Xu, Min Li, Rui-Rui Zheng, et al. *Manchu character segmentation and recognition method*. 2017, 20(1): 43-53.
2. M. LI, R. ZHENG, S. XU, F. WANG. Design and Implementation of Embedded Handwritten Manchu Letter Recognition System, *2018 International Conference on Information Systems and Computer Aided Education* (ICISCAE), Changchun, China, 2018, pp. 7-11.
3. R. Zheng, M. Li, J. He, J. Bi, B. Wu. Segmentation-Free Multi-Font Printed Manchu Word Recognition Using Deep Convolutional Features and Data Augmentation, *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics* (CISP-BMEI), Beijing, China, 2018.
4. Guang-yuan Zhang, Jing-jiao Li, Rong-wei He, Ai-xia Wang. An offline recognition method of handwritten primitive Manchu characters based on strokes. *Ninth International Workshop on Frontiers in Handwriting Recognition*, Kokubunji, Tokyo, Japan, 2004, pp. 432-437.
5. M. Li, R. Zheng, S. Xu, Y. Fu, D. Huang. Manchu Word Recognition Based on Convolutional Neural Network with Spatial Pyramid Pooling, *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics* (CISP-BMEI), Beijing, China, 2018.
6. Choi Minseok, Lee Choong Ho. A Study on the Preprocessing for Manchu-Character Recognition. *Journal of the Institute of Convergence Signal Processing*, 2013, 14(2).

Design of Timing Socket Based on Single Chip Microcomputer

HongBo Hao[1] *, Shuailin Chen [2]
[1] *Tianjin University of Science and Technology, China;*
[2]*Qingdao University of Science and Technology, Shandong, China*
E-mail: 2844011096@qq.com
www.tust.edu.cn

Abstract

In this paper, the design of timing socket is mainly used for smart timing control of household appliances, and the working time of timing is set by pressing the key. In the working period, the external connection device and the power supply are in the conduction state. Outside the set time period, the external connector is disconnected from the power supply. This design uses *STC89C52* as the driving chip. *DS1302* clock module provides accurate timing. *LCD1602 LCD* module displays time and operation interface. The relay module controls the connection between the socket and the power supply. *DC* power supply module is converted to DC power supply module.

Keywords: Smart timing control, *STC89C52, DS1302* clock module, *DC* power, Socket

1. Introduction

Nowadays, with the improvement of living standards and the acceleration of life rhythm, the sockets on the market cannot meet our needs because of their single function. For example, electric vehicle charging is usually about 8 hours. If it is too long, it is easy to lose the battery. If it is too full, it is not enough to charge. If it is charged at night, you often don't want to go out from home to the garage to charge the car. In the daytime, they forget to turn on the power; the water tower at home forgets to pump water, which causes temporary inconvenience to life. The school's broadcast of getting up is delayed because the staff on duty overslept. The fish tank in the home has not been supplied with oxygen for too long, which causes the fish to die of lack of oxygen. Many fans or news enthusiasts miss the start-up time and miss the wonderful ball games, so it is not convenient to use it at night. Considering the above reasons, we urgently need a smart socket. This socket can supply power to electrical appliances regularly and cut off the power supply of electrical appliances outside working hours. In this way, the standby loss of electrical appliances can be solved, the purpose of saving electricity can be achieved, and the hidden danger of safety can be eliminated [1].

At present, in the market, the technology of time controller is relatively advanced and complete. Time controller is widely used in all kinds of electrical appliances. In the field of smart socket, however, it is still in the stage of development. The smart socket is only limited to the knob timer, or the time control socket with larger volume designed by counter chip. This kind of socket timing time is single, the function is single, the timing accuracy is low and it is difficult to really meet the needs of our daily life. Now we need a powerful time control socket to meet the market demand and improve our life. With the development of modern electronic science and technology, low power consumption and other characteristics. The price of finished products has decreased, and has been widely used. Accurate display of real-time time and timing time, it has the memory function and the data will not be lost after power off. It not only achieves the function and reduces the cost, but also it let our daily life have a profound impact.

2. Overall scheme design

This chapter mainly discusses the overall scheme of the system and the design scheme of each module, including the control core, switch, display module, clock circuit and keyboard circuit. In this paper, *STC89C52* is selected as the core component of this design. The basic timing control function is realized by using the flexible programming design and rich IO port of the single chip microcomputer and the accuracy of its control. The overall design of the system mainly realizes the following functions:

(1) Human computer interaction interface: through independent buttons to establish a complete set of human-computer interaction interface.

(2) Infrared remote control: the system can set two timing time through the infrared remote control, and It can be forced on / off to confirm the start timing function

(3) Buzzer prompt: when the power is on and off, there will be a buzzer prompt.

2.1 Power supply scheme

The design power supply is + 5V DC power supply. In this paper, USB power cord is used for power supply, one end is connected with DC interface, the other end is connected to + 5V power supply. The power supply mode can be computer USB, mobile phone charger, power bank, etc. The power supply is supplied by 220 VAC to 5 VDC regulated power supply.

2.2. Relay circuit scheme

Relay is a kind of electrical switch, which is actually a switch device with small current to control large current. It can not only protect the circuit, but also realize the function of safe conversion circuit. There are many kinds of relay, including electromagnetic relay, time relay, and temperature relay. This paper uses ***SRD-05VDC-SL-C*** relay [2].

3. Hardware circuit design

3.1. Minimum system of single chip microcomputer

The minimum system consists of reset circuit, clock circuit and *STC89C2*. Minimum system circuit as is shown in the Fig.1 below. This system uses *STC89C52* control chip. The single chip microcomputer not only has strong function and low power consumption, but also has strong anti-interference ability and high cost performance ratio [3].

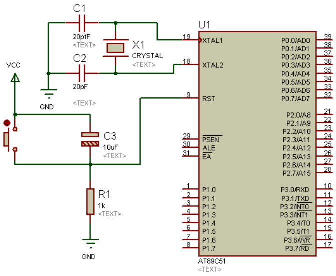


Fig.1. Minimum system circuit

3.2. Clock circuit module

The system needs to calculate the exact time, so the module uses an 11.0592MHz crystal oscillator. The clock circuit is used to generate the clock signal needed by MCU. In order to ensure the realization of synchronous working mode, the circuit should work in strict accordance with the time sequence under the control of the only clock signal. The clock circuit diagram is shown in Fig.2.

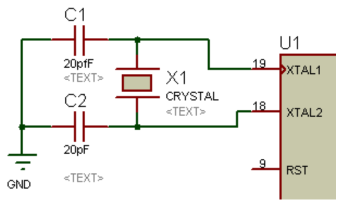


Fig.2. Clock circuit diagram

3.3. Reset circuit

Reset is the initialization operation of single-chip microcomputer. When the single-chip microcomputer starts and runs, it must be reset. The reset circuit includes two parts: power on reset and manual reset. The reset signal of MCU is high level reset. The reset circuit diagram is shown in Fig.3.

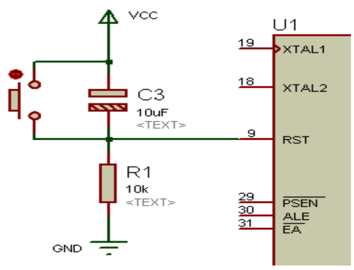


Fig.3. Reset circuit diagram

3.4. Socket power circuit

The power on and off is not frequent, so this design uses the relatively low price relay as the control circuit device of socket power supply. An I / O port of the single chip microcomputer controls the on-off of the relay through a triode, so as to control the power on and off of the socket. The control circuit of socket power supply is shown in Fig.4.

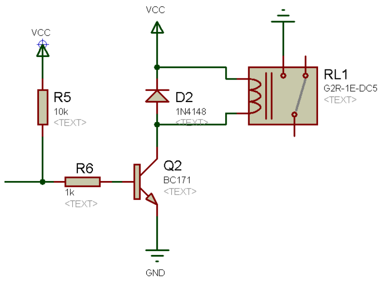


Fig.4. Socket power circuit

3.5. Independent key module

The key K1 is used to set the subtraction operation of the time value. The key K2 is used to set the addition operation of the time value. Key K3 can be set separately, and then select manual mode or automatic mode. The default mode of MCU is automatic mode. It can also work with the K4 button to set the position of the variable. Schematic diagram of key module is shown in Fig.5 [4].

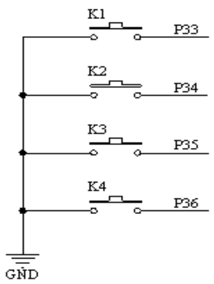
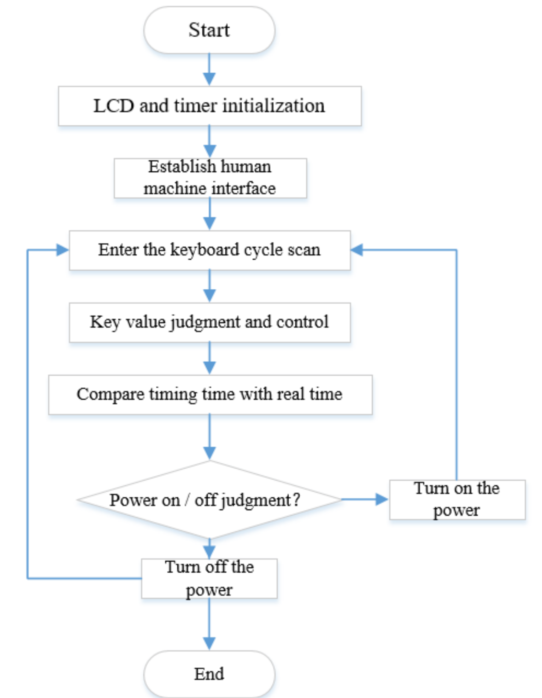


Fig.5. Schematic diagram of key module

4. Software design

4.1. Main program flow chart

In this design, the program is mainly under the control of single-chip microcomputer. In this process, the MCU first initializes, including setting the direction of each port, the initialization of each variable, the initialization of LCD, the disconnection of relay and the calibration of oscillation frequency.



The flow chart of the whole system software design is shown in Fig.6.

Fig.6. Software design flow chart

4.2. Design of setting time function

The flow chart of setting timing is shown in Fig.7 below. After power on, enter the time timing program, and then judge whether the key is valid. If it is valid, set the start and end time of timing through four independent keys. If it is judged that the key is invalid, it will return to wait for the key program until the end of the setting. Set timing time program design: set function void *key_ with( )* [5].

5. Conclusion

5.1. System simulation and testing

According to the designed hardware schematic diagram, the simulation circuit diagram is drawn in Proteus. Then the designed program is compiled and saved in. After running, the LCD can display the calendar time. The display page shows the year, month, day, 2019-05-17 in the down line, 15:30:51 in the uplink and w: 5 in the week. System simulation circuit diagram is shown in Fig.8 below [6].

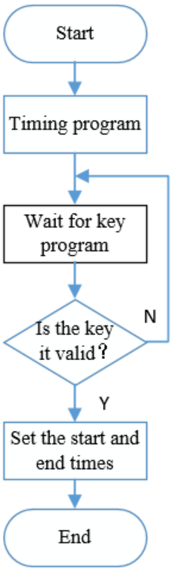


Fig.7. Schematic diagram of key module

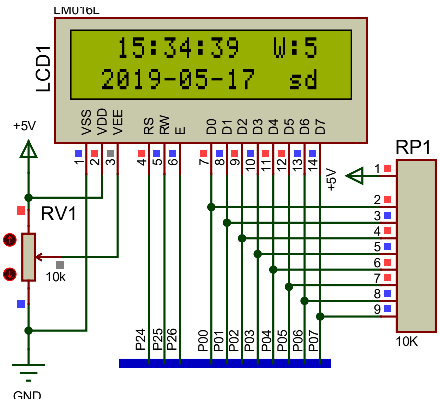


Fig.8. System simulation circuit diagram

5.2. Power simulation and testing

Multisim 13.0 is selected for the simulation and testing of DC voltage stabilizing circuit, which has rich simulation and analysis ability. This design first according to the hardware design circuit diagram, draws the circuit diagram in the Simulation Software Multisim, and then it carries on the operation simulation to the circuit. The waveform after running the simulation is shown in Fig.9.

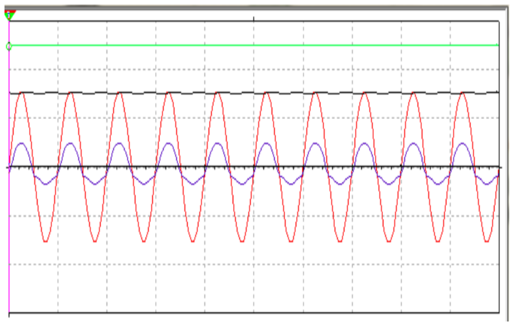


Fig.9. System simulation waveform

References

1. K. Tsai, F. Leu, I. You, Residence Energy Control System Based on Wireless Smart Socket and IoT, *IEEE Access*, vol. 4, pp. 2885-2894.
2. C. E. Shannon, A symbolic analysis of relay and switching circuits, *Electrical Engineering*, 57(12): pp. 713-723.
3. M. Wu, X. Wu, Design of the single-chip microcomputer experiment teaching system for application, *2011 International Conference on Electrical and Control Engineering*, Yichang, 2011, pp. 649-652.
4. Kong, Fan Rong, Wang, Wenxin, Zhu, Jiajun. A Smart Home Control System Based on Controller Area Network Bus. 2013, 11(9):1667-1671(5).
5. Sandy Suryo Prayogo, Farid Al Rafi, Yulisdin Mukhlis. Design and Built IoT Home Panic Button for Smart City. 2019, 1175(1).
6. Hamid Mahmoudi, Mohsen Aleenejad, Reza Ahmadi. Reconfigurable rapid prototyping platform for power electronic circuits and systems for research and educational purposes. 2018, 11(7): 1314-1320.

Study on the Effect of Physical Fitness Training on Children's Cognitive Ability

Jianhua Deng*, Lei Ning
Sports Training Speciality, Yanshan University, China
*E-mail: * 245615583@qq.com*
www.tust.edu.cn

Abstract

With the word physical fitness getting more and more attention in our country, physical fitness has been widely developed in our country. Therefore, in recent years, the students' physique has received attention from all aspects of society, and improving the students' physique has also become a new topic for physical education workers. I try to discuss the effect of physical fitness training on children's cognitive ability from the perspective of physical fitness training. To provide theoretical basis and reference for intervention of children's cognitive ability training methods. So as to promote the improvement of children's cognitive ability and provide guarantee. This paper studies the influence of physical fitness training on children's cognitive ability by using the methods of literature, observation and experiment. The following conclusions are drawn: 1. Before the experiment, 8 children from Qinhuangdao Aikushao Physical Fitness Center were tested for attention cognitive ability and memory cognitive ability. Among them, 8 children were deficient in cognitive ability to varying degrees. Compared before and after the experiment, physical fitness training has an effect on children's attention cognitive ability.3. Compared before and after the experiment, physical fitness training has an effect on the improvement of children's memory cognitive ability.

Keywords: Physical fitness of children; Children's cognitive ability; Experimental study

1. Introduction

Children are the motherland's "bud and future ". As early as 1988, at the 10th World Congress of the World Future Research Federation, experts from various countries discussed the results of the report on "the end of the children ". [1] It is generally believed that" children are the basic human resources and the creation of a world conducive to their development should be recognized as limited regional and global measures ".

However, due to the influence of the one-sided pursuit of the transition rate, the society and the school have the tendency to pay more attention to intellectual education and light sports, the students' schoolwork burden is too heavy, the rest and exercise time is seriously insufficient, and the students' interest in sports gradually weakens or even disappears [2]. The physical decline of young students seriously affects the healthy growth of children and even the future of the country and nation.

Children are in the golden development stage of life, their cognitive ability will directly affect the level of learning in the future. Therefore, according to the needs of children's own development and the requirements of improving children's basic sports ability, this study makes an experimental study on whether children's physical fitness can improve their cognitive ability [3].The word physical fitness comes from English Physical Fitness, first appeared half a century ago by the American Association for Health, Sports, and Leisure.

After many years of development, health fitness has become an important standard to measure human physical health [4]. Scientific research and related scientific theoretical research on health fitness play an important role in improving human physical health and improving human living standard [5].

Physical fitness exercises have a positive effect on the improvement of children's health and physical fitness level. These interventions aim to increase students' physical fitness level. However, most of the current studies on students' physical fitness are carried out outside the classroom, and there are few studies linking children's physical fitness with cognitive ability.

2. Object and Methods of Study

In this study, 7-8-year-old children in Qinhuangdao Aiku physical Fitness Center were selected to conduct physical fitness experiments. Eight 7-8-year-old children in Qinhuangdao Aiku physical Fitness Center were tested for attention cognitive ability and memory cognitive ability before the experiment. After 2 months of targeted physical fitness training, attention cognitive ability and memory cognitive ability of 8 7-8-year-old children in Qinhuangdao Aiku physical Fitness Center were tested again.

3. Findings and analysis

3.1 Analysis of Children's Attention Cognitive Ability before Physical Fitness Training Intervention

In Qinhuangdao Aiku Children's physical Fitness Center, 8 children aged 7-8 were randomly selected for the experiment. All the children participating in the experiment needed two months of targeted training to improve their cognitive ability. In the later stage of this experiment, the analysis of physical fitness training on children's attention cognitive ability and children's memory cognitive ability was analyzed. Before the experiment, in order to ensure the contrast of the experiment, according to the parents' attention ability test table and the children's memory ability test table as is shown in the Tab.1, the following table was obtained, and the attention ability and memory ability of 7-8 years old children were analyzed.

Tab.1. Cognitive status of attention and memory

Name	A	B	C	D	E	F	G	H
Attention	18	12	10	7	13	10	15	14
Memory	23	30	45	65	58	40	49	52

According to the scoring rules of children's attention ability test table, the attention cognitive level of 8 experimental children in Aiku Children's physical Fitness Center of Qinhuangdao City was obtained in Tab.2.

Tab.2. The attention cognitive ability before the experiment

Name	Number	Percentage
Well noted	0	0%
mild	4	50%
medium	3	37.5%
severe	1	12.5%

Through the analysis of the attention cognitive ability of 7-8-year-old children in Qinhuangdao Children's Physical Fitness Center, You know, There are 0 children with good attention skills, 0% of the total; There are four children with a slight lack of attention, 50 per cent of the total population; There are three children with moderate capacity deficits, 37.5 per cent of the total; There is one child with a severe lack of attention, 12.5% of the total number.

3.2 Interventional training arrangements for improving children's attention awareness

By understanding the training methods of expert physical fitness and the knowledge of physical fitness I have learned, the training contents of children's attention cognitive ability are screened out. According to the stage principle of children's physical fitness training, the training will last for 2 months

Practice is arranged in three stages of training in an advanced manner to prepare for the accuracy of the experiment. The experiment is shown in the Tab.3

Tab.3.The Content Stage of Attention Cognitive Training

Phase	Foundation	Improvement	Strengthening
Time	1-3 weeks	4-6 weeks	7-8 weeks
Training	Warm up Stretch Dumu Bridge S Run Relax	Warm up Stretch Porter Color identification Relax	Warm up Stretch Canoe Bridge S running colors Relax
Number of groups	6-8 groups	6-8 groups	6-8 groups

3.3 Analysis of Children's Attention Cognitive Ability after Physical Fitness Training Intervention

After two months of experiments, An analysis of attention cognitive ability of 8 7-8 year old children in Qinhuangdao Aikukuo Physical Fitness Center, You know, One child with good attention skills, 12.5 per cent of the total; There are five children with a slight lack of attention, 62.5 per cent of the total; There are 2 children with moderate capacity deficits, 25 per cent of the total population; There are 0 children with severe disability, It accounts for 0% of the total as shown in Tab.4.

Tab.4. The attention cognitive ability after the experiment

	Number of persons	Percentage
Well noted	1	12.50%
mild	5	62.50%
medium	2	25%
severe	0	0%

3.4 Comparative Analysis of Children's Attention Cognitive Ability

Through the comparison before and after the experiment as shown in Fig.1, 8 children with good memory ability before and after the experiment in Qinhuangdao Aikuo Children's physical ability Center had 1 person ,3 children with mild memory deficiency and 6 people after the experiment; There were 3 children with moderate memory deficiency and 1 children with severe memory deficiency. To sum up, after the experiment, the whole children's memory cognitive ability level comparison experiment came to see, has improved.

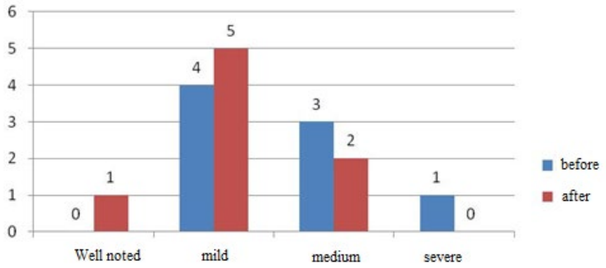


Fig. 1.Comparison of Attention Cognitive Ability of Children before and after the Experiment

4. Conclusions

Through the comparison of 7-8-year-old children in Qinhuangdao Aiku physical ability Center for 2 months, the physical fitness training is helpful to the improvement of children's attention cognitive ability.

References

1. Seeger JP; Thalain DH; Noor dam Physical fitness training improves physical fitness and vascular function in children with type 1diabetes. *Diabetes, Obesity & Metabolism*. 201.
2. Elbasan, Effects of chest physiotherapy and aerobic exercise training on physical fitness in young children with cystic fibrosis. *Italian Journal of Pediatrics*, 2012, 38.
3. Yang Taiji .On the concept and relationship between physical fitness and physical fitness. *Contemporary sports science and technology*, 2018, 8(16): 178-179.
4. Zhu JUANJUAN. *Research on the construction of physical education curriculum for large class children from the perspective of physical fitness*. Wuhan Institute of physical education, 2018.
5. Wang Zhen. *Experimental research on the influence of children's physical fitness training on 3-6-year-old children's sensory integration ability*. Wuhan sports College, 2018.

Effects of high heels on plantar stress in women

Jianhua Deng*, Lei Ning
Sports Training Speciality, Yanshan University, China
*E-mail: * 245615583@qq.com*
www.tust.edu.cn

Abstract

OBJECTIVE: to investigate the changes of Plantar pressure and gait in women wearing high-heeled shoes for different years. METHODS: from March 2019 to April 2019, a total of 30 female faculty members in Yanshan University were surveyed by questionnaires and plantar pressure was measured by static and dynamic tests. RESULT: 1. The peak value of plantar pressure of the arch of foot increased with the increase of wearing time (P & Lt; 0.05). 2. The plantar pressure and impulse of the second metatarsal and the fourth metatarsal increased with the increase of wearing time (P & Lt; 0.05). CONCLUSION: 1. Long-term wearing high-heeled shoes will cause foot discomfort, easy to cause flat foot and Hallux Valgus Lesions; 2. The distribution of plantar pressure and the change of arch shape and gait may be the main reasons for wearing high-heel shoes.

Keywords: high heels; women; questionnaire; plantar stress test

1. Introduction

The human foot is composed of 26 bones, 33 joints and a reticular layered structure composed of 126 ligaments, muscles and nerves. Its basic function is to bear weight, buffer, absorb impact force, produce forward driving force, and coordinate and maintain the balance of the human body. When the human body walks, the foot bears the reaction of the ground to the human body, up to 1.5 times its weight.

The study of the characteristics of plantar pressure distribution can reflect the structure and function of gait, plantar pressure and the movement of the whole body posture during walking [1]. The maximum peak pressure on the plantar was mainly concentrated on the medial and lateral sides of the heel [2]. With the increase of heel height, the position of the maximum peak pressure changed, mainly concentrated in the second metatarsal, the third metatarsal and the medial side of the heel. With the increase of heel height, the peak pressure in the medial and lateral areas of the heel gradually decreased, while the peak pressure in the second metatarsal and third metatarsal regions increased with the increase of heel height [3]. Metatarsal pressure is dense, long-term heavy load stimulation is easy to cause damage, so women should choose the appropriate height of shoes.

2. Research subjects and methods

The subjects were female workers and teachers of Yanshan University, and 30 were selected by random sampling, of which 17 were administrative posts and 13 were teachers. Finally, 15 people each completed the plantar pressure test as required. All the subjects had no abnormal gait and medical history. By using body composition analyzer and PODOTEK HD series pressure plate, the body composition results are printed and the plantar pressure distribution map is obtained [4]. Body composition analyzer was used to observe the height, weight and BMI of human body, and the mean peak force, plantar pressure, impulse and pressure of each area of plantar were observed by plantar pressure plate tester. The characteristics of plantar pressure were evaluated [5].

As shown in Fig.1, the observable sites 1 and 2 are heel (RF); 3,4 and 5 are arch (MF); The 5th metatarsal (MH5),4th metatarsal (MH4),3rd metatarsal (MH3),2nd metatarsal (MH2) and 1st metatarsal (MH1); the sites 11,12,13,14,15 were followed by the fifth phalanges (T5), the fourth phalanges (T4), the third phalanges (T3), the second phalanges (T2), and the first phalanges (T1).

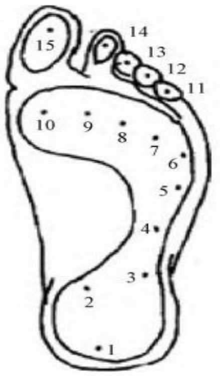


Fig.1. Plane planer

3. Research findings and analysis

3.1 Physical fundamentals and differences by group

From Table 1, we can see that there is no significant difference in height, weight and BMI in each group, and there is a statistical difference in age group. Because age differences are inherent differences, the impact on the study can be excluded. The statistical table shows that each group has the consistency of height, weight and BMI value, which can exclude the influence of height, weight and BMI on the test results of each group.

Tab.1. Basic information on groups (x±s)

	<2-year	2-5 years	6-10 years	11-20 years
Age	32.6±1.4	32.2±4.8	37.4±1.6	44.4±10.6
Height	160.5±12.5	158.9±5.4	162.3±10.3	161.3±11.8
Weight	54.3±8.7	54.4±4.7	61.8±9.2	61.6±12.3
BMI	19.6±3.0	20.0±2.1	21.3±4.5	20.9±5.1

3.2 Pressure in the foot region

After the completion of the test, the computer software automatically generates the pressure distribution map of each area of the foot as shown in Fig.2.

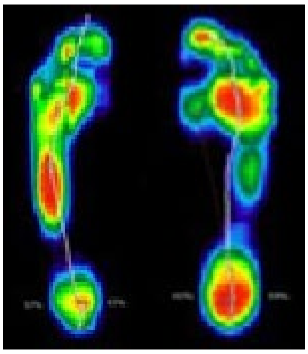


Fig.2. Distribution of pressure in various regions of the foot

The pressure peak data of each part are used to draw the following statistical as shown in Tab.2 and the change line diagram as shown in Fig.3.

Tab.2. Statistical tables of mean plantar pressure peak in left foot groups (x±s)

	<2-years	2-5 years	6-10 years	11-20 years
T1	40.88±2.95	53.16±10.36	56.96±31.47	56.49±10.04
T2-5	9.58±0.43	14.79±5.43	22.06±8.03	19.23±15.52
MH1	43.07±6.78	51.69±20.89	40.39±12.43	47.83±18.38
MH2	49.13±12.48	50.96±18.69	40.61±10.13	53.94±16.78
MH3	55.94±19.79	59.47±13.04	40.72±14.96	52.97±24
MH4	60.11±26.48	49.24±15.7	39.69±17.43	46.23±14.34
MH5	18.33±2.87	32.3±12.56	20.81±10.1	25.9±5.24
MFM	4.87±4.43	9.35±3.79	13.84±6.98	15.64±13.08
MFL	15.13±16.71	29.38±18.44	21.17±14.19	44.69±14.29
RFM	85.16±41.3	78.8±28.99	95.87±30.69	100.85±12.34
RFL	71.82±12.93	64.26±5.96	72.18±8.2	73.4±13.84

Fig .3 Breakline diagram of plantar peak changes in left foot groups

3.3 Analysis of Pressure and Impulse Test Results in Subsole Regions

The distribution of plantar pressure peak showed that the maximum stress position of plantar was medial heel > lateral heel > second metatarsal > third metatarsal > lateral arch, first phalanx > fourth metatarsal and fifth metatarsal medial arch >2-5 phalanges

4. Conclusions

Wearing high heels for a long time can lead to a variety of foot discomfort, the main part of which is located in the medial side of the anterior foot (60%) and the arch of the foot (25%), the main type of discomfort is intolerance to long walking and pain.

Wearing high heels for a long time can change the characteristics of plantar pressure distribution, and show different trends under different wearing years.

Wearing high heels for a long time can cause changes in the arch of the foot, and in the 2-5 year period, the arch changes significantly

High trends. In the period of 6-10 years of wearing, the arch of the foot has an obvious trend of low level, which is easy to cause the occurrence of flat foot.

Compared with lower heels, the second metatarsal pressure of higher heels was significantly higher than that of lower heels. The second metatarsal pressure of heel height in 5-7 cm compared with 3-5 cm was significantly higher than that of heel height in 3-5 cm.

References

1. A study on biomechanical characteristics of gait and plantar pressure in young women with subgloss. Wearing different heights of heel walking Shandong Institute of physical Education, 2014.
2. Wang Haijun, Dong Lishu, Chen Qinghe, Li Rongwei. Effects of different heel height on plantar pressure distribution in obese young women with natural walking gait. *Journal of Hebei normal University of Science and Technology*, 26(03): 63-67.
3. Branthwaite H, Chockalingam N, Grogan S, et al. Footwear choices made by young women and their potential impact on foot health. *Journal of Health Psychology*, 2013, 18(11): 1422-1431.
4. Morris PH, White J, Morrison ER, et al. High heels as supernormal stimuli: How wearing high heels affects judgements of female attractiveness. *Evolution and Human Behavior*, 2013, 34(3): 176-181.
5. Menz HB, Morris ME. Footwear characteristics and foot problems in older people. *Gerontology*, 2005, 51(5): 346-351.

A Research on Intelligent Classification of Urban Trash Bins Based on Machine Learning

Longyu Gao *, Zilong Liu, Luqi Shen, Songyun Shi, Yongzheng Lv

Tianjin University of Science and Technology, China;

*E-mail: * ggafcc@163.com*
www.tust.edu.cn

Abstract

Aiming at the problems of inaccurate, insensitive and general performance of the current intelligent garbage sorting bins, an intelligent garbage sorting bin based on machine vision is proposed. The trash can is mainly divided into five modules: main control module, machine vision module, classification module, overflow reminder module, and Wi-Fi Internet of Things module. The trash can uses convolutional neural networks to build an intelligent garbage classification model and classification algorithm to achieve rapid and accurate garbage classification. This experiment will be based on the identification of waste bottles, analyze the recognition characteristics of machine vision, and then propose methods to improve the accuracy of recognition.

Keywords: machine vision, convolutional neural network, intelligent classification

1. Introduction

At present, our country mainly relies on manual garbage classification, that is, placing several trash cans with different types of marks. Pedestrians need to judge which type of garbage they are in, and then put them in the correct trash can. Or hand over the trash to the staff nearby, and let them sort the trash and put it out. Although manual classification has achieved a certain effect, it is still not a long-term solution. This not only consumes a lot of time, manpower and material resources, but also loses the biggest advantage of the trash can itself, that is, convenience and time saving.

Therefore, the intelligent sorting trash can came into being. Although smart trash cans have long been proposed and designed, they are still not widely used in public places. The main reason is that the cost is too high and the garbage classification function is basically not available; and its main intelligent function is to use the infrared ranging module to sense whether the pedestrian is close to the trash can, and once it is close, the lid of the trash can is automatically opened for pedestrians to put Garbage; In addition, it also has voice prompts and driving functions. Obviously, this kind of smart trash can is mainly designed for family living, and many functions are not applicable in public relations situations.

For example, once a pedestrian just passes by the trash can and has no intention of throwing trash, the trash can will identify errors, consume unnecessary energy, and greatly increase costs. In addition, although there are some smart sorting trash cans on the market that can achieve recyclable and non-recyclable sorting functions, the use of sensor technology results in low classification accuracy and not widespread popularity.

2. Overall design scheme of the system

This article discusses an intelligent garbage sorting bin based on machine learning. The convolutional neural network is built through the artificial intelligence learning architecture, the training set is established in the early stage, the collected image data and image attribute tags are input, and the recognition model is obtained through multiple convolution operations; during image recognition, the garbage images collected by the camera are uploaded, After simple preprocessing, the identification model determines the attributes and categories of garbage [1].

That is, according to recognition methods such as contour recognition, feature recognition, color recognition, and material recognition, find the common points between the two, and finally realize the judgment and classification. Finally, the corresponding signal is transmitted back to the single-chip microcomputer, and different steering gear angles are output according to the type of the returned signal, and then the flip board and the push rod are controlled to sort the garbage into the correct group. The network classification flowchart is shown in Figure 1.

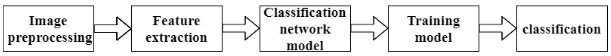


Fig.1. Classification process

The machine learning technology is applied to the system, and self-learning is carried out through the unsupervised learning task of clustering in machine learning. The specific work is: the system recognizes the characteristics of a large amount of garbage, and then presupposes. The unique characteristics of the four types of garbage are divided into four categories, each of which collects a large amount of data and saves the data in a database. When new garbage is encountered, the system extracts the data and matches it with the data in the database. Find out which type of data is the same or similar to the system, and classify it into one category. If there is no matching data in the system, the system automatically infers the new data from the existing data, and the system judges this by itself. Which category the new data belongs to, which makes the classification more accurate.

The main functions of the smart garbage sorting bin outlined in this article are: taking pictures of garbage thrown by pedestrians, intelligently sorting garbage, sensing whether the garbage bin is full, etc. It is mainly divided into five modules: main control module, machine vision module, mechanical module, sensor overflow module, Wi-Fi Internet of Things module. The main flow chart of the system operation is shown in Figure 2, and the structure design of the trash can is shown in Figure 3.

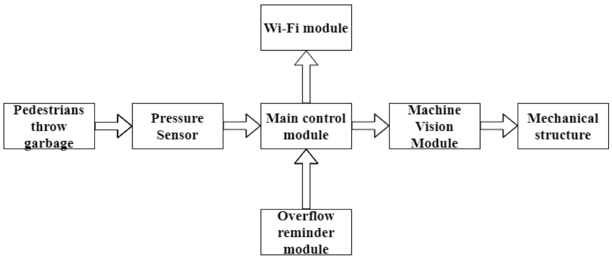


Fig.2. flow chart

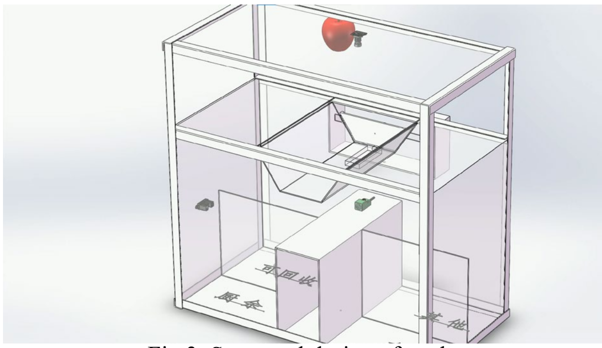


Fig.3. Structural design of trash can

In this experiment, a common beverage bottle is taken as an example. The beverage bottle data set comes from online and real-life photos. There are 4000 pictures in total, of which 60% of the training samples are 2400; the number of verification samples is 20%, which is 800; testing The sample accounts for 20%, which is 800 sheets. After completing the training of the garbage classification model, the model is tested for its recognition function and classification. According to the recognition results, the garbage classification model is optimized, and the collected garbage pictures are used as the input of the model, and the test is performed 800 times. Comparing the model output results with the real results, it is found that the recognition rate reaches 96%. It can be seen that all bottles tested can be accurately identified, accurately and reliably. Later, based on this classification model, other garbage data training will be carried out.

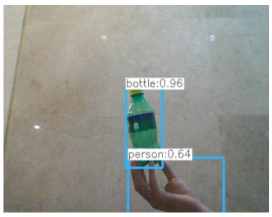


Fig.4. Recognition accuracy

3. Design of hardware control system

3.1. Main control module

The main control module is mainly composed of STM32 controller. It is the brain of the entire system. When the pressure sensor senses that there is garbage being dropped, the STM32 controller controls the operation of the machine vision module. After the image recognition module recognizes the image category and feeds it back to the main controller, the STM controller will control the mechanical module to turn the corresponding angle to realize the intelligent classification and placement of garbage. The STM32 controller is shown in Figure 5.

Fig.5. STM32F103ZET6

3.2. Machine Vision Module

The OpenMV module is installed at the garbage disposal port, and it captures images of garbage on the disposal platform by taking pictures [3]. It is mainly composed of a camera. The camera adopts a USB drive-free 1080P high-definition camera, which greatly improves the recognition resolution; at the same time, it also has a 160-degree wide angle, which greatly increases the shooting range; in addition, it also has a flash function to ensure that it takes pictures in a dim environment. OpenMV is shown in Figure 6.

Fig.6. OpenMV

The machine vision module mainly analyzes and processes the photos taken by the machine vision module. The photos taken by the camera can be regarded as a large two-dimensional array. The elements of the array are called pixels, and their values are called gray values. The captured photos are converted into grayscale images, binary images and true color RGB images according to the size of the color and grayscale value, and the images are matched with the garbage training atlas pre-input to the convolutional neural network. According to the contour recognition, Recognition methods such as color recognition and material recognition seek common features between the two, and then identify the type of garbage that is thrown in according to the image algorithm written.

This module also adds the function of machine learning. First, it compares the various characteristics of the identified garbage with the existing characteristics of the system. If they are the same or similar, they are classified into one category and the data is stored in the system at the same time. Enlarge the system's database to make the next recognition more accurate and faster.

Compared with the current smart trash can, if there is nothing in common with the data stored in the system, the system will have problems that cannot be recognized.

For the system designed in this article, due to the addition of machine learning functions, when encountering such problems, the system will analyze its own existing data, and then self-judge the characteristics of the new data collected, thereby Perform accurate classification.

3.3. Mechanical module

After the image recognition module recognizes the type of garbage and feeds it back to the main control module, the main control module will control the steering gear to turn the corresponding angle according to the programmed algorithm.

Control the forward and backward movement of the front and rear push rods and the left and right flips of the left and right flaps to realize the movement of garbage in two dimensions. Finally realize the garbage classification to the corresponding trash can. The sorting mechanical structure is shown in Figure 7.

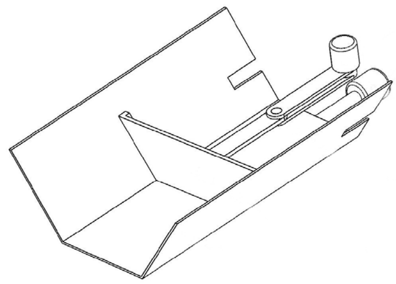


Fig. 7. Sorting mechanical structure

3.4. Sensor overflow module

This module is mainly used to detect whether the garbage in the trash can is full. In order to make the induction more accurate, it is divided into two small modules: weighing module and ultrasonic distance measuring module. The weighing module is used to sense whether the weight of the trash can exceeds the theoretical weight value. The weight value will be set according to the maximum allowable weight of different types of trash cans through investigation.

In order to prevent some special garbage thrown by pedestrians from being too heavy, causing the weight of the garbage can to exceed the theoretical weight value, but it is still not full, we also designed an ultrasonic ranging module at the mouth of the inner wall of the garbage can.

The working principle is that the ultrasonic transmitter emits ultrasonic waves in a certain direction, and starts timing at the same time as the transmitting time. The ultrasonic waves propagate in the air and return immediately when encountering obstacles on the way. The ultrasonic receiver stops timing immediately after receiving the reflected waves. Record the time t, then use the formula $S=vt/2$ to calculate the distance between the ultrasonic source and the obstacle. When the garbage is not full, the actual distance measured is the inner wall diameter of the bucket; when the garbage is full, the actual distance measured is the distance from the ultrasonic transmitter to the garbage, so as to determine that the garbage can is full. The ultrasonic ranging sensor is shown in Figure 8.

Fig.8. Ultrasonic distance sensor

Conclusion

Through the design of software and hardware, this paper designs an intelligent sorting bin based on machine vision. It mainly uses image recognition methods to classify garbage. In addition, the dual combination method of weight sensing and ultrasonic ranging is used to determine whether the trash can is saturated, which greatly improves the accuracy of the judgment. Compared with the existing smart garbage sorting bin, because of the use of machine learning functions, the recognition is more accurate and sensitive, and the practicability is higher. The recovery rate of recyclable garbage is greatly improved, and the loss of manpower, material resources and financial resources is reduced. , Has a certain use value. In general, the innovation of the system lies in simplifying the types of garbage classification and adding machine learning functions to improve the problem of garbage classification to the greatest benefit.

References

1. Golnabi, H., and A. Asadpour. Design and application of industrial machine vision systems. *Robotics and Computer-Integrated Manufacturing*, 2007, 23.6: 630-637.
2. Liu, Jianghai, Yadong Jiang. Design of intelligent trash can be based on machine vision. *2020 International Conference on Image, Video Processing and Artificial Intelligence*. Vol. 11584. International Society for Optics and Photonics, 2020.
3. Vernon, David. Machine vision-Automated visual inspection and robot vision. STIA 92 (1991): 40499.

A design on intelligent public trash can based on machine vision and auxiliary sensors

Longyu Gao*, Leixin Han, Jiangyu Wu, Mingfei Liu, Ruming Kang

Tianjin University of Science and Technology, China;

*E-mail: * ggafcc@163.com*
www.tust.edu.cn

Abstract

In order to improve the correct rate of front-end recognition in the garbage classification process, the automatic garbage classification system designed based on machine vision technology has a significant improvement in recognition accuracy compared to traditional smart garbage cans. But in the case of identifying irregular garbage, the recognition accuracy is greatly reduced. In order to solve this kind of problem, four types of auxiliary sensors are added to the trash can, through the mutual cooperation between the sensors, combined with the results of machine vision recognition, comprehensive judgment, greatly improving the recognition accuracy of irregular garbage.

Keywords: machine vision, auxiliary sensors, garbage sorting

1. Introduction

With the rapid development of society, the level of consumption and production has greatly improved. At the same time, waste production and environmental degradation have also increased. According to statistics, the world produces approximately 10 billion tons of waste each year. 400 million tons, if garbage cannot be classified correctly, the land resources occupied by long-term stacking will exceed 25 million square meters, so the task of garbage classification is very difficult.

Garbage sorting can be divided into two parts: front-end sorting collection and back-end sorting and recycling. At present, the world's main front-end method of garbage classification still relies on manual work, and mainly relies on citizens' consciousness and staff review. This method is very inefficient, difficult to solve the problem of large amounts of untreated garbage, and cannot guarantee the correctness rate. Therefore, in the past two years, the concept of a smart trash can was proposed.

The current smart trash cans are mostly based on infrared sensors, capacitive sensors or LCD screens for identification and classification. Some smart trash cans have introduced machine vision and machine learning related technologies. However, due to the wide variety of garbage and the same type of style and Abundant shape changes make recognition difficult, and current methods cannot meet the needs of daily intelligent recognition and classification [1]. Although some related studies can achieve a stable distinction between recyclable and non-recyclable garbage, they have not yet been able to meet China's need to distinguish four types of garbage.

In order to meet China's garbage classification requirements and improve the accuracy of front-end recognition in the garbage classification process, this design adds four types of sensors to the trash can, through the mutual cooperation between the sensors, combined with the results of machine vision recognition. Recognition and judgment have successfully achieved a substantial increase in the recognition accuracy of irregular garbage.

2. Overall design scheme of the system

The intelligent waste classification system is mainly composed of machine vision module, auxiliary sensor module and mechanical module that realizes automatic classification control. When the garbage is put into the system, the machine vision module starts the camera to obtain the garbage image, and then runs the convolutional neural network to obtain the garbage category information. For garbage that cannot be identified, auxiliary sensors are used for identification. Finally, according to the classification of the garbage type, the mechanical structure is controlled to make the garbage fall into the corresponding bin [2]. In the auxiliary sensor part, we have selected 4 auxiliary sensors for the commonality of different types of garbage, namely VOC sensors, metal sensors, capacitive proximity sensors, and humidity sensors. The result of machine vision recognition is shown in Figure 1.

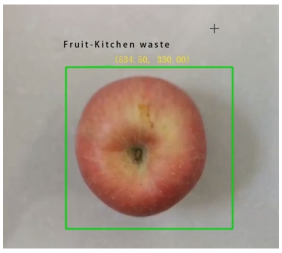


Fig.1. Machine vision recognition results

2.1. VOC sensor

The detection of hazardous waste that cannot be recognized by machine vision is mainly realized by the ZP16 digital VOC air quality sensor module. The air quality module uses advanced chip thick film semiconductor gas sensors. The gas sensors are resistant to formaldehyde, benzene, carbon monoxide, and ammonia. , Hydrogen, alcohol, cigarette smoke, flavors and other organic volatile gases have extremely high sensitivity. After aging, debugging, calibration and calibration, the module has good consistency and extremely high sensitivity. Extremely high sensitivity, excellent long-term stability, factory calibration, convenient and quick to use, with sensor fault self-diagnosis function, low power consumption and long life. The VOC sensor is shown in Figure 2.

Fig.2. VOC sensor

2.2. Metal sensor

The metal sensor is used to identify metals and realize the function of material identification. This system uses LJ18A3-8-Z/BX type metal sensor. The sensor has the advantages of fast response speed, strong anti-interference ability, waterproof and corrosion resistance, and very stable performance. It is widely used in automated assembly lines, intelligent hardware and security systems. The operating voltage of the sensor is 6-36vdc, and the switch contains a high-frequency oscillation circuit, a detection circuit, an amplifying circuit, a de-transmitting circuit and an output circuit. When power is supplied to the switch, the oscillator in the high-frequency oscillation circuit generates an alternating electromagnetic field on the detection surface of the switch. When a metal approaches the detection surface of the switch, the eddy current inside the metal absorbs the energy of the alternating electromagnetic field in the oscillator, so that the oscillator weakens or stops. The two states of the energy change of the oscillator are converted into a level signal by the detection circuit, the level signal is amplified by the amplifying circuit, and then the trigger circuit triggers the output transistor circuit to work to generate a switching signal. Thereby detecting the presence or absence of metal to achieve the purpose of detecting metal. The metal sensor is shown in Figure 3.

Fig.3. Metal sensor

2.3. Capacitive proximity sensor

The second part of the material identification is mainly realized by the capacitive proximity switch M8M12. It can detect any dielectric substance, including conductors, semiconductors, insulators, and can even be used to detect liquid level and powdered materials. For non-metallic substances, the movement distance depends on the dielectric constant of the material. The greater the dielectric constant of the material, the farther the sensing distance is. Cooperating with motor control, changing the distance between it and the tray can realize the classification of different materials. The capacitive proximity sensor is shown in Figure 4.

Fig.4. Capacitive proximity sensor

2.4. Humidity Sensor

Use SHT3x temperature and humidity sensor module to communicate with STM32, and turn it on when machine vision can't recognize it, so as to realize the identification of wet garbage. The module applies digital module collection technology and temperature and humidity sensing technology, and the product has high reliability and long-term stability. Built-in humidity and temperature sensor elements, analog to digital converters, signal processing, calibration data and I2C host interface. The technology of using low-k polymer dielectric for humidity sensing creates a low-power, monolithic MoS sensor module with low deviation and hysteresis and long-term stability. It is suitable for measuring humidity, dew point and temperature. The module is small in size and the accuracy can reach ±0.4°C. It can be embedded in many highly integrated environments. With millisecond measurement conversion time, there is no need to wait between starting the measurement and reading the data. The 2.54mm spacing interface is convenient for integration testing. The humidity sensor is shown in Figure 5.

Fig.5. Humidity Sensor

3. Auxiliary sensor recognition

First, the VOC sensor detects the harmful gas produced by the corrupted hazardous garbage, and if it is detected, the garbage is marked as hazardous garbage. If no harmful gas is detected, the metal sensor is used to realize the function of material identification and distinguish between recyclable metals and others. If the metal is not recyclable, the capacitive proximity sensor is used to determine whether it is paper or plastic, and the paper and plastic are marked as recyclable. Finally, the humidity sensor is used to distinguish food waste from other wastes based on the characteristics of food waste containing more water, and mark them. The auxiliary sensor recognition process is shown in Figure 6.

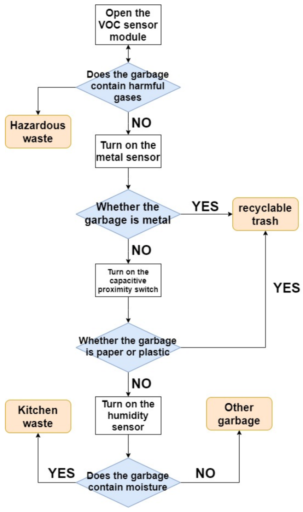


Fig.6. Auxiliary sensor recognition flowchart

In the early stage, a large amount of data training is carried out on the machine recognition system, so that it can independently realize the correctness of a certain amount of garbage and recommend the database based on this, and then use the visual camera to capture and recognize the image of the discharged garbage. Compare and analyze with the data in the database to determine the type of garbage, and then use its own mechanical structure to sort and put it into the corresponding garbage bin [3].

If there is garbage that the garbage can cannot identify, the four auxiliary sensors are used to determine the type of garbage in real time, and the garbage is placed in the buffer area of the garbage can for manual processing to avoid misplacement. Mark the judgment result, take a photo and upload it, send it to the staff for manual review, and finally upload the result to the database to serve the next identification and judgment. When faced with this kind of garbage next time, the trash can will be able to directly identify and classify through machine vision.

As the database shares the data between all smart trash cans online, the learning speed of smart trash cans will increase exponentially. The overall flow chart is shown in Figure 7.

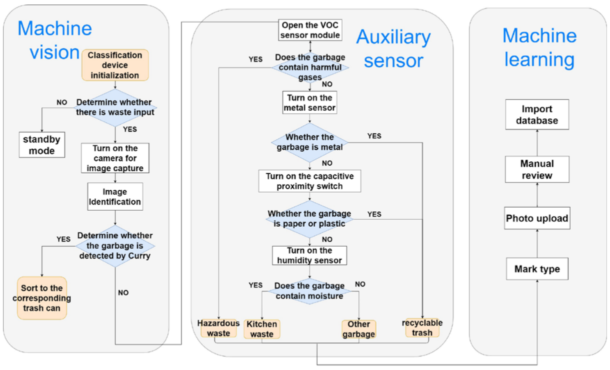


Fig.7. Overall flow chart

4. Prospect analysis

Through the analysis of various existing smart trash cans, the problem of insufficient machine vision recognition accuracy in the front-end collection part of the garbage classification process and cannot meet the needs of daily intelligent recognition and classification is summarized and studied. This article adopts four types of sensors. The organic combination of "new garbage" classified by auxiliary sensors is marked and then manually reviewed and entered into the database, which greatly improves the accuracy of garbage identification.

At the same time, it combines the needs of Chinese garbage classification, and expands the scope of classification to four types: recyclable, hazardous, food waste and other garbage. This classification method is based on whether it is degradable, harmful to the human body, and reusable. It has universal applicability to the garbage classification and treatment methods of various countries. Therefore, the design has been adjusted to some extent and can be applied to the global garbage classification.

In addition, placing the "new garbage" mark classified by the auxiliary sensor into the buffer area for manual processing ensures the accuracy of the classified garbage and ensures that the back-end classification and recycling can be carried out directly after the front-end collection part is completed. This saves nearly 80% of human resources, and greatly improves the efficiency of waste sorting and treatment, reducing land resources wasted due to long-term garbage stacking, and alleviating the problem of global waste sorting and treatment. Therefore, this design has excellent application prospects.

Conclusion

Aiming at the problem of insufficient accuracy of smart garbage classification, this paper proposes the use of four types of sensors to assist machine vision and machine learning recognition to improve the recognition accuracy and the correct rate of intelligent classification. Realize the classification and release of four kinds of garbage with high accuracy, greatly improve the accuracy of garbage identification, reach the standard of actual use, save a lot of waste of manpower and natural resources, and the design has the advantages of low cost and easy iteration. , Provides a new reference for the front-end collection method of garbage classification.

References

1. Golnabi, H., and A. Asadpour. Design and application of industrial machine vision systems. *Robotics and Computer-Integrated Manufacturing*, 2007, 23.6: 630-637.
2. Liu, Jianghai, Yadong Jiang. Design of intelligent trash can be based on machine vision. *2020 International Conference on Image, Video Processing and Artificial Intelligence*. Vol. 11584. International Society for Optics and Photonics, 2020.
3. Vernon, David. Machine vision-Automated visual inspection and robot vision. STIA 92 (1991): 40499.

A Research on Front-End Garbage Classification Based on Machine Vision

Longyu Gao*, Zhiqing Xiao, Junlong Hao, Luqi Shen, Manqian Hu

Tianjin University of Science and Technology, China;

*E-mail: * ggafcc@163.com*
www.tust.edu.cn

Abstract

Adding a machine vision recognition module to the traditional smart trash can can effectively improve the efficiency of trash recognition. The intelligent garbage classification model constructed by the convolutional neural network can accurately identify the types of garbage, with an average accuracy rate of 0.87. Deploy the trained model on openMV and test it on the produced physical trash can. After the system is stable, the average time to complete a sorting and recovery is 2s. Experiments show that the system can effectively identify the types of garbage and complete garbage classification and recycling.

Keywords: machine vision, intelligent classification, convolutional neural network

1. Introduction

With the rapid economic development and the continuous improvement of people's living standards, the amount of garbage has increased sharply in the cycle of consumption and production. According to the research report, by 2050, the global waste volume will increase by 70%, from 2.01 billion tons in 2016 to 3.4 billion tons in 2050. The task of sorting waste is very difficult, and it is already very difficult to study an effective method for sorting waste. urgent. Scholars at home and abroad have done a lot of analysis on waste classification, but most of the proposals put forward are innovations in terminal recycling methods. Front-end collection still relies on people's consciousness, and the accuracy of garbage classification is very low. Research on an effective front-end collection method is of great significance to the status quo of garbage classification.

For the front-end collection of garbage, smart trash cans have begun to appear in foreign markets, such as Nastar in the United States, ccko (cckoing) in Germany, Zarton and somatosensory smart trash cans in Japan, which are mostly designed based on infrared sensors and LCD screens. Domestic research has also been carried out to make trash cans have functions such as automatic classification, alarming and monitoring of available capacity, but it has not yet been able to meet the needs of automatic identification and classification.

With the rapid development of artificial intelligence technology, machine vision technology has gradually matured, and image classification methods based on deep learning have gradually diversified. This provides a new direction for the research on garbage recognition in this article. Scholars have made a lot of efforts in garbage recognition and image classification. Research and achieved certain results. For example, using the YOLOv4 network for garbage identification, a detection accuracy of 94.8% for dry and wet garbage can be achieved, and an accuracy rate of over 90.0% can be achieved in a complex detection environment.

This paper proposes an automatic garbage classification system based on machine vision. The system can meet the classification requirements of recyclables, kitchen waste, other waste and hazardous waste. The intelligent garbage classification model constructed by the convolutional neural network can accurately identify the types of garbage. With mechanical structures such as the front and rear push rods and left and right flaps, the entire garbage classification process can be completed.

2. Overall system design

Using Solidworks software to model the trash can, the overall structure is shown in Figure 1. The trash can includes a box body, the box body has a top plate, a camera is installed on the lower surface of the top plate, and the box body is provided with a plurality of sorting compartments for collecting different types of garbage. The sorting compartments are arranged oppositely to form two Row [1]; the upper middle of the sorting compartment is provided with a garbage flipping plate that can be driven left and right to flip, the garbage flipping plate is a trough-shaped structure that can accept garbage, and the trough-shaped structure is provided with the garbage to be accepted Push the push plate toward the front. The structural design of the trash can is shown in Figure 1.

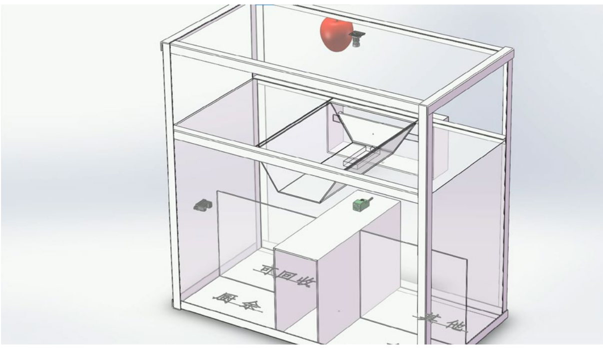


Fig.1. Structural design of trash can

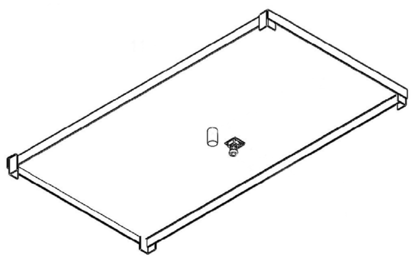


Fig.2. Roof structure

A light sensor and a light source module are also installed in the identification barrel, which are located in the top cover part, which is shown in FIG. 2. To provide a light source for machine vision at night, the light sensor and the light source module are respectively connected to the central processing module, and the light sensor is used for sensing and recognizing light changes in the barrel; the light source module is used for providing the light source for the machine vision module.

The inner wall of the trash can is equipped with an infrared sensor. When the trash is thrown in, the infrared sensor emits and receives the infrared rays reflected by the object, which can detect the filling height of the trash.

2.1. Master chip

This text selects STM32f407zet6 as the core controller of the hardware control system, which is mainly used to realize the functions of data acquisition and analysis and execution of component motion control. The STM32f407zet6 processor is an Arm Cortex-M4 high-performance core with a maximum operating frequency of 168mHz and a package form of 1qfp144. Its peripheral configuration is powerful, supporting communication interfaces such as SP interface, I2C interface, USB interface, USART interface, and peripherals such as ADC and timer. The development board used is connected to OpenMV, and OpenMV obtains garbage pictures, processes the obtained picture information, and recognizes the types of them. Finally, return the identified garbage type information to the STM32f407zet6 development board [2]. The development board controls the movement of the steering gear to realize the movement of the front and rear push rods and the left and right flaps, and then the garbage is dropped into the correct box. STM32F103ZET6 microcontroller is shown in Figure 3.

Fig.3. STM32F103ZET6

2.2. OpenMV

The vision part uses OpenMV single-chip microcomputer to communicate with STM32. The OpenMV camera is a small, low-power, low-cost circuit board that can be scripted by a high-level language Python (to be precise, MicroPython). You can use an external terminal to trigger shooting or execute an algorithm, or use the result of the algorithm to control the IO pin [3].

OpenMV uses STM32H743II ARM Cortex M7 processor, 480 MHz, 1MB RAM, 2 MB flash. All I/O pins output 3.3V and are 5V tolerant. The OpenMV module is shown in Figure 4.

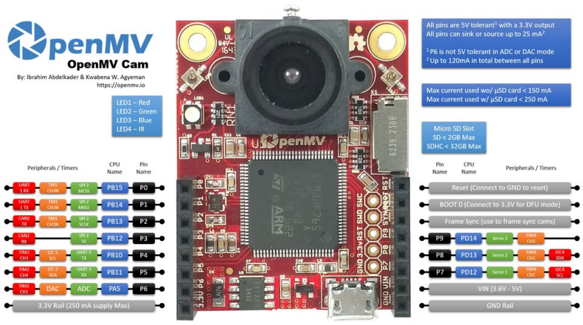


Fig.4. OpenMV module

2.3. DS5160 Servo

The smart trash can passes through the DS5160 servo. The working voltage of the steering gear is 6V-8.4V, and the controllable angle is 180°. The DS5160 actuator works fast and has stable performance. Its control method is PWM control, pulse width range: 500-2500us, control precision: 3us. The actuator can quickly respond to the motion instructions of the system and sort different types of garbage into the corresponding bins. The DS5160 servo is shown in Figure 5.

Fig.5. DS5160 Servo

2.4. Classification agency

This intelligent sorting trash can realizes garbage classification based on machine vision by using the camera. Through the left-right flip of the garbage dumping plate and the front and back movement of the push plate, it can realize the movement of the garbage in two dimensions, and then realize the movement of the garbage position, and finally realize The correct classification of garbage is sorted and placed in the classification cabin below to complete the automatic classification of garbage. The classification mechanism module is shown in Figure 6.

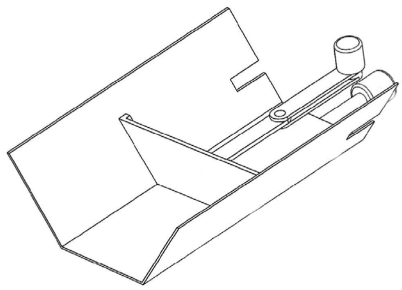


Fig.6. Classification mechanism module

2.5. Overflow reminder module

In order to facilitate the frequent overflow of garbage in the trash bin, the system is designed to detect the full bin. By installing an ultrasonic module just above each sorting bin, periodically detecting the distance L1 between the garbage in the bin and the module, and making the difference with the distance L2 from the module to the bottom of the sorting bin, you can get the current height H of the garbage in the sorting bin, that is, H = L2 -L1. If H exceeds 90% of the height of the sorting bin, the mark is that the sorting bin is full. When the bin is full for processing, a reminder will be sent to the corresponding APP through the WiFi module to remind relevant personnel to clean up the garbage in the bin as soon as possible. On the other hand, in order to avoid increasing the burden, the Raspberry Pi should disable the recognition function, that is, put it aside for processing before the full bucket state is removed, and no longer run the neural network to identify the thrown garbage until the full bucket state is removed.

Fig.4. Capacitive proximity sensor

2.6. Analysis of results

This verification experiment conducted experiments on 6 commonly used types of garbage. For each type of garbage, 10 representatives with obvious differences in appearance were selected, and 100 throws were made at different angles, sizes, and different degrees of damage. The result data of mechanical structure classification was recorded. The specific data records are shown in Table 1.

It can be seen from Table 1 that this system has an accuracy rate of over 87% for the identification of common 6 types of garbage in actual use, and from the appearance model of the trash can in this experiment, it can be seen that the size of the trash can is small and meets the expected design requirements. It is also possible to increase the types of garbage identification by expanding the training set. The trash can realizes the two-dimensional movement of trash through the movement of the front and rear push rods and the left and right flaps, thereby realizing the correct classification of trash. The average time for each classification is about 2 seconds, which meets the experimental expectations. The system identification data table is shown in Table 1.

Table 1. System identification data table

Type of garbage	Recognition correct times/time	Average recognition rate
Can	87	87%
plastic bottle	92	92%
Milk carton	90	90%
Paper cup	94	94%
Paper ball	93	93%
battery	95	95%

Conclusion

This article mainly studies the intelligent classification of urban trash bins based on machine vision, uses convolutional neural networks to build an intelligent trash classification model, and then realizes real-time monitoring of the filling status of smart trash bins through the Internet of Things technology to achieve intelligent control of trash bins .

The convolutional neural network is applied to garbage classification, the design scheme is simple to implement, and the research results achieve practical results. At present, there are very few trash cans that are classified in the process of centralized garbage collection on the market. The innovation of this implementation scheme lies in the combination of new convolutional neural network technology, which can transform ordinary trash cans into smart devices through low-cost embedded devices. Trash can to realize the function of identifying and automatically sorting garbage.

References

1.Golnabi, H., A. Asadpour. Design and application of industrial machine vision systems. *Robotics and Computer-Integrated Manufacturing*, 2007, 23(6): pp. 630-637.

2.Liu, Jianghai, Yadong Jiang. Design of intelligent trash can be based on machine vision. *2020 International Conference on Image, Video Processing and Artificial Intelligence*, Vol. 11584. International Society for Optics and Photonics, 2020.

3.Vernon, David. Machine vision-Automated visual inspection and robot vision. STIA 92 (1991): 40499.

NeuroPhyllotaxis: An Interactive Application for Generative Art Based on EEG Data

Chien-Tung Lin, R.P.C Janaka Rajapakse
Graduate Institute of Studies in Documentary & Film Archiving, Graduate Institute of Animation and Film Art, Tainan National University of the Arts, No. 66, Daqi, Guantian Dist., Tainan City 72045, Taiwan

Yoshimasa Tokuyama
Department of Media and Image Technology, Tokyo Polytechnic University, 1583, Iiyama, Kanagawa 243-0297, Japan
E-mail: 109130101@tnnua.edu.tw, janakaraja@gmail.com, tokuyama@mega.t-kougei.ac.jp
www.tnnua.edu.tw and ‡www.t-kougei.a.cjp

Abstract

Generative art is produced by procedural techniques. It has obtained a lot of attention since the beginning of computer graphics. Many works of art are inspired by nature, among which phyllotaxis is as well. It is a combination of mathematics and the beauty of nature. Not only can it be seen everywhere in nature, but also often appear in man-made objects, becoming part of culture or religion. This paper presents the development of an interactive generative art application that is created from a phyllotaxis pattern by using the user's EEG data. When people are using it, it will allow them to more easily relax and achieve the function of art therapy. We tried to use EEG data to make an interactive installation art which creates phyllotaxis patterns that are projected on the wall. Everyone has a different state, the generated patterns are also different from person to person, which creates interesting interactive contents. In addition, sound can also be changed by EEG to become dynamic and real-time contents.

Keywords: EEG, phyllotaxis, generative art, interactive art, installation art

1. Introduction

With the development of microcomputers and single chips, the way people interact with computers has changed. It has come a long way from the early keyboard and mouse to the development of the multi-touch. Developers always want to find better ways to let people interact with apps in a more natural way. Such as gestures (Leap Motion[1]), wii[2], voice input, etc. However, these are all motor functions, that is, different people using the same operation method will get the same result. Compared with electroencephalogram (EEG), because people can't see how the user moves, it just exists in the brain's operation, so no one can copy the actions of another person and produce a personal style. Therefore, we aim at the state of concentration and meditation in the brain waves as an interactive input method to bring a new perspective.

In addition, installation art is becoming more and more popular as a form of modern art. Compared with other art forms, installation art allows people to be more involved, use touch, and even play in it. Therefore, interactive methods are also added to allow people to enjoy, participate in, communicate, and interact with art, to become interactive installation art. However, most of the art content is fixed, arranged according to the creator. So there is not much change, resulting in people not wanting to repeat, come, and participate.

Generative art is produced by computer programs. It has the characteristics of non-repetition and endless evolution. It has received a lot of attention since the beginning of computer graphics. Many works of art are inspired by nature, among which phyllotaxis is included. It is a combination of mathematics and the beauty of nature. Not only can it be seen everywhere in nature, but also often appear in man-made objects, becoming part of culture or religion. For example, Madala Kolam is not only the expression of art, but also healing people's body and mind. And Chakra's representative totem is also like phyllotaxis. When people are using it, it is easier to relax the mood and achieve the function of art therapy. Therefore, we try to use EEG to make interactive installation art, and use EEG input to create phyllotaxis patterns to project it on the wall. Because everyone has a different state, the graphics drawn are also different from person to person, creating interesting interactive works. In addition, the sound can also be changed by EEG, becoming a dynamic, real time content. More importantly, when the user looks at the graphics generated by his own brain, this can bring about communication with the heart and then achieve the effect of art therapy[3].

2. Background: EEG Control art

2.1. *The use of EEG in art*

Since 1965, the use of EEG in artistic creation has been documented. Most of the content has an established pattern. Authors or users use their brains or mental states as artistic creations, such as focusing or relaxing to control sound and images[4].

2.2. *Passive control and aesthetics*

Another way to use EGG as artistic creation is more passive. It uses the mental and emotional state of the author or participant to produce changes in sound and images, especially music that affects the emotional changes of the user. The changing screen synchronized with the music[4].

However, this application goes beyond this way of expression, and focuses more on the communication between the self and the heart. Through the displayed graphics and sounds to perceive and explore one's mental state, open up internal communication and dialogue to achieve relaxation and meditation. This allows for self-reflection, and ultimately the opportunity to achieve the goal of art therapy.

3. Development and application

3.1. *system structure*

The development environment is Intel I7 PC on windows 10 system, with the NVIDIA GeForce RTX 2080TI graphics card. The programming language C# has been used to develop the application on the Unity SDK. The EEG headset is NeuroSky MindWave Mobile 2.

Figure 1 and Figure 2 show the EEG headset is connected to the computer via Bluetooth, and uses the intermediary software ThinkGear Connector[5] to send the received data to Unity[6] via UDP network protocol.

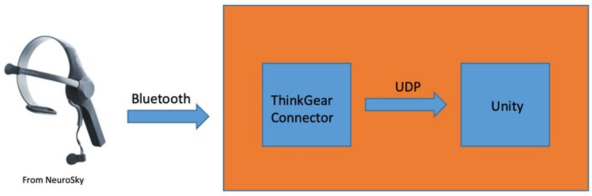


Fig. 1. EEG connecting to Unity

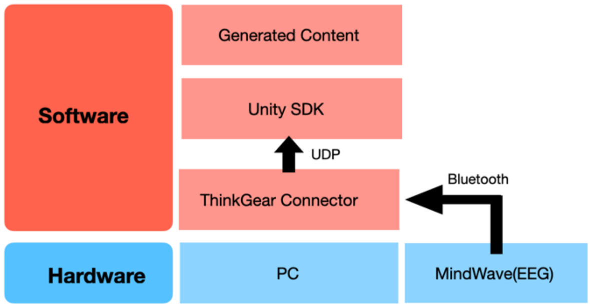


Fig. 2. System Structure

3.2. *Content generation*

The simple model for the florets of sunflower was Vogel 's formula. As the book "The Algorithmic Beauty of Plants"[7] descript below:

$$\phi = n * 137.5°, \qquad r = c\sqrt{n}, \tag{1}$$

• n is the ordering index number of a floret, from the center to outer.

• φ is the angle of the nth floret in a polar coordinate system.

• r is the distance from the center of the capitulum to the nth floret.

• c is a constant scaling.

In Figure 3, it can be found that changing different angles will produce different pattern changes, so the angle is used as a variable in the program to generate changeable patterns.

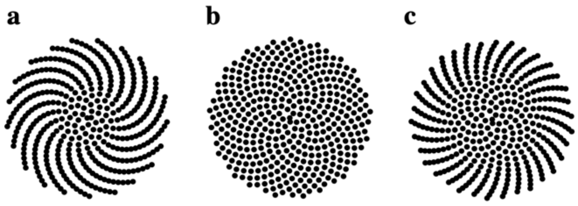


Fig. 3. Generating phyllotactic patterns from the Vogel's formula. (a) α =137.3°, (b) α =137.5°, (c) α =137.6°

3.3. *EEG Data*

Different waveform data can be obtained from the signal from the brain waves, such as alpha wave, beta wave, etc., as well as data on concentration and meditation. We used the level of concentration to generate graphics, and use the level of meditation as a reference value for color and sound.

The degree of concentration can be divided into a value of 0-100. We divide it into 7 ranges. The value of concentration is low, the angle of the graph is small; and, the higher the value is, the larger the angle as shown in the Table 1.

Table 1. Use value of attention for angle φ.

Range of Attention	Angle φ
0-14	41
15-29	59
30-44	73
45-59	97
60-74	109
75-89	137.5
90-100	157

Similarly, the degree of meditation divides the value from 0-100 into 7 equal ranges, which are represented by 7 colors of red, orange, yellow, green, blue, indigo, and violet. Divided into 7 ranges, the number and color of the chakras are the same[8], and they also form the basic color of white light. This correspondence is natural and conforms to the human body.

Table 2. Use value of meditation for color.

Range of Meditation	Color
0-14	RED
15-29	ORANGE
30-44	YELLOW
45-59	GREEN
60-74	BLUE
75-89	INDIGO
90-100	VIOLET

3.4. *Installation*

Figure 4 shows the installation for the application which used a projector to project the generated graphics onto the neutral wall, allowing users to sit in front of the projection screen and watch. This allows the user to relax and wear the EEG headset.

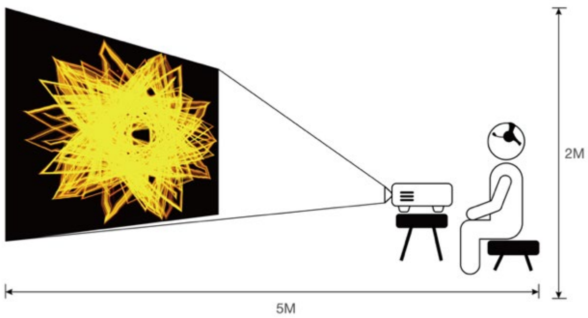


Fig. 4. Installation for the application.

4. User feedback

A user study was conducted over 50 university students, between the ages of 16 to 25. The experiment apparent and the projector was used to project the patterns to the neutral wall. Each subject used the applications for 2 to 5 minims. After that, they were asked to take an online survey. The questions included what they thought while using the application and how they felt when used the applications. Figure 5 shows some patterns made by different users.

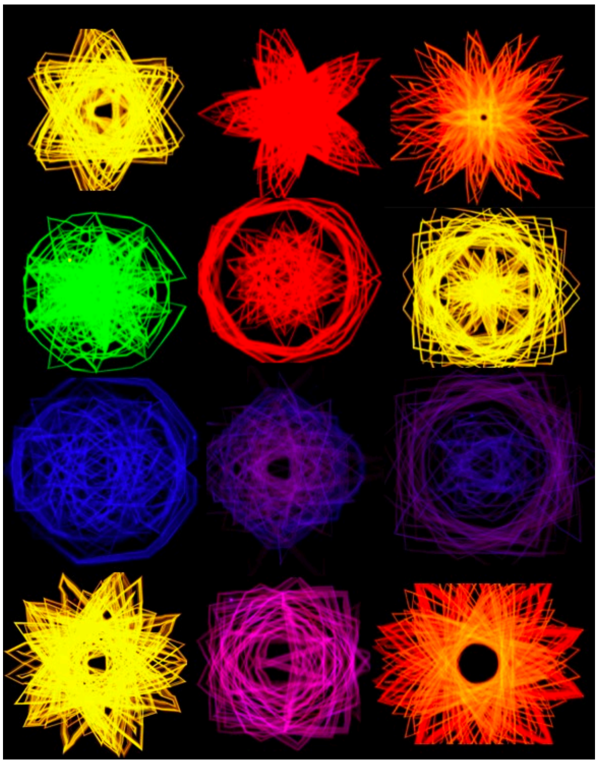


Fig. 5. Pattern generated by people.

Figure 6 shows when people used the application, 22/50 tried to think about different things to see the changes in graphics, colors, and sounds. 4/50 tried using different emotions to see different changes. 21/50 were in the state of emptying meditation. 3/50 were thinking about the meaning of these graphics and how did such things make.

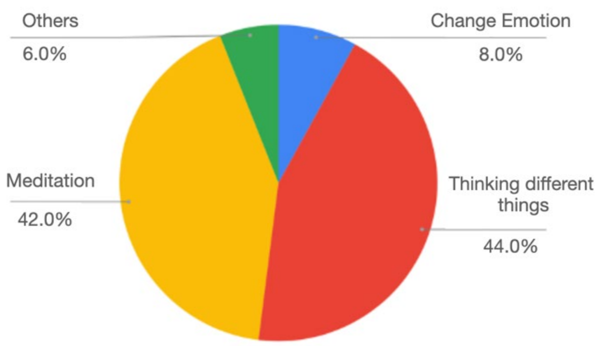


Fig. 6. How people thinking while using the device.

Among them, 42/50 found it interesting and novel. In addition to the artistry brought about by the graphics and color changes, almost everyone wanted to know the meaning of these graphics and whether they can understand their emotions or psychological state.

5. Conclusion

The use of EEG to generate phyllotaxis is a very novel experience for most people. As an installation art, the generated graphics have different results according to the conditions of different people, making the installation art show a variety of effects.

Through some applications, users can also recognize their own state. Through different thinking patterns, imagination, and memories, people were able to produce graphics of different colors. Watching the visual and sound effects of these graphics can trigger psychological states, such as novelty, such as a sense of excitement and calmness, to achieve an increasingly relaxed state, and achieve the effect of meditation. It is like a dialogue and communication with the self, allowing the heart to settle, reflect, meditate, and think, and find the inner peace and answer.

After using this program, does it have a positive impact on the user? Can it improve people's concentration? Does it allow people to be more relaxed? Will it increase people's creativity and work efficiency? These can all become further research projects.

Acknowledgements

This research was supported in part by the Ministry of Science and Technology of Taiwan under the grant (IV) MOST 109-2218-E-002-038-.

References

1. Leap Motion, https://developer.leapmotion.com. Accessed 6 Dec 2020.
2. Wikipedia, Wii, https://en.wikipedia.org/wiki/Wii. Accessed 6 Dec 2020.
3. Wikipedia, Art therapy, https://en.wikipedia.org/wiki/Art_therapy. Accessed 7 Dec 2020.
4. M. Zhang, Y. Huang, S. U. Ahmed, M. Shidujaman, Mapping Between Mind Cybernetics and Aesthetic Structure in Real-Time EEG Art. DOI: 10.1007/978-3-030-60117-1_26
5. NeuroSky, ThinkGear Connector, http://developer.neurosky.com/docs/doku.php?id=thinkgear_connector_tgc. Accessed 7 Dec 2020.
6. Unity, https://unity.com. Accessed 7 Dec 2020.
7. P. Prusinkiewicz et al, The Algorithmic Beauty of Plants, Regina, Canada, 1990, pp. 99-109
8. Healthline, chakras, https://www.healthline.com/health/fitness-exercise/7-chakras. Accessed 8 Dec 2020.

ThoughtMix: Interactive Watercolor Generation and Mixing Based on EEG Data

R.P.C. Janaka Rajapakse
Graduate Institute of Animation and Film Art, Tainan National University of the Arts, No. 66, Daci, Guantian District, Tainan 72045, Taiwan

Yoshimasa Tokuyama
Department of Media and Image Technology, Tokyo Polytechnic University, 1583, Iiyama, Kanagawa 243-0297, Japan
E-mail: janakaraja@gmail.com, tokuyama@mega.t-kougei.ac.jp,
[†]www.tnnua.edu.tw and [†]www.t-kougei.a.cjp*

Abstract

In modern art, new types of installations and artworks use sensor-based inputs as an interactive method to create new forms of media art. Most of those installation artworks have static and pre-captured contents; they tend to be played as dynamic, and by some concerned as interactive creations. However, most installations are not able to capture the visitor's rich feedback which could trigger the contents. In order to identify the relation between the cybernetic-aesthetics and visitor's brain wave activities, we propose an electroencephalogram (EEG) data-driven approach to generate watercolor and mixing effects which visualizes the changes in brain wave activities in real-time. Our proposed application transforms the brain waves into real-time watercolor effects that are displayed by generating the data about emotion, attention, meditation, and neural mechanisms with EEG data. "ThoughtMix" creates watercolor effects as an immersive experience to influence creativity and art therapy by actually visualizing the colors the user's mind creates.

Keywords: EEG art, watercolor, thought colors, digital colors, interactive mix, interactive generation

1. Introduction

Currently, people living in a digitally connected era are in the space of reality and virtuality. Digital technology is commonly used for human-computer interaction and content creations [1]. Interactive art is a method that stimulates a variety of the user’s senses and activities inside digitally created contents with a computer, so that users may interactively experience environments and spaces that can be difficult to create in traditional techniques [2]. As computer graphics-based applied art and technology becomes advanced and as various sensor hardware interfaces are developed and spread, a variety of applications that enable the user to interactively encounter spatial and temporal experiences are being created [3], [4]. Additionally, there is a growing demand for art and technology that support these interactive applications [5].

It is important that digital art technology moves beyond the boundaries of the traditional interaction techniques to provide a creative space that allows the user to create and experience therapeutic sensations. The acceptance of brain-computer interface (BCI) leads many art researchers to explore the diversity of mind cybernetic methods and expression forms in electroencephalogram (EEG) art [6]. This paper presents an EEG data-driven approach to generate real-time watercolor and mixing effects that visualize the changes in brain wave activities in order to identify the relation between the cybernetic-aesthetics and user’s brain wave activities. The proposed application "ThoughtMix" transforms the brain waves into real-time watercolor effects that are displayed by generating the data about emotion, attention, meditation, and neural mechanisms with EEG data. This interactive application creates watercolor effects as an immersive experience to influence creativity and art therapy by actually visualizing the colors the user's mind creates.

The paper is organized as follows: section 2 presents related work. Section 3 describes our development framework. Section 4 is devoted to experimental results. Section 5 discusses the main points presented throughout the article, drawing conclusions on future research directions.

2. Background

There are several works in literature that cover interactive art in various forms. Most papers feature contents and comparison methods between different generative techniques, providing performance results of the interactive techniques, and often reports of users' preferences. Unfortunately, due to the variety of different art installations and different interfaces used for interaction and control, it is still difficult to derive visitor-generated contents. Therefore, further research is necessary to understand which are the best options for a specific immersive experience to influence creativity and therapeutic sensations.

2.1. *EEG in Art*

When compared to traditional interactive feedback techniques, using EEG as a biological control is considered relatively implicit. Mechanical operation and motion detection are widely used in traditional interactive art applications. In these traditional applications, the explicit macro-communication generated is based on the participant's movements and behaviors. However, EEG control directly uses electrodes to acquire the activity of neurons and takes thinking as a controller. EEG artwork is a new form of communication that cannot be observed through body behavior but the mind.

Prpa and Pasquier [6] defined a structured overview of the expanding field of BCI art, with utilized approaches and BCI devices, and proposed a systematic way of categorizing artworks based on their similarities in the taxonomy. This presented taxonomy encompasses sixty-one artworks that illustrate the nuances of the categories in the taxonomy. According to the distance between input and output control, Fishkin's [7] introduced a classification of the perceptive interface based on the level of self-containment. Wadeson et al. [8] provided different views on creative control, which are separated into four categories for participants: passive, selective, direct, and collaborative. These classification models described the spatial relationship between the existences of participants and artwork, which determine the behavioral characteristics of aesthetic construction. In this paper, we focus on passive inputs which does not require the participant to perform any particular task to change or influence their brain activity explicitly.

2.2. *Generative Art: Color and Mixing*

Computer-based generative art is a procedure to create artforms by executing a set of rules in a methodical collaboration between an artist and a machine [10]. Galanter [9] defined "*Generative art refers to any art practice where the artist uses a system, such as a set of natural language rules, a computer program, a machine, or other procedural invention, which is set into motion with some degree of autonomy contributing to or resulting in a completed work of art*". Shuhei Matsuyama [11] created an installation "dyebirth" that continually creates organic patterns through digitally controlled physical phenomena produced by a mixture of water, ink, and chemical substances.

Refik Anadol [12] presented an installation artwork named "*Melting Memories*" based on measured EEG data, which offered new insights into the representational possibilities emerging from the intersection of advanced technology and contemporary art. "*Thinking in Color*" was an interactive experience that used an EEG brainwave scanner to generate colors, influenced in real-time by the user's brainwave activity [13]. To generate watercolor effects, the proposed application in this paper, named "ThoughtMix" also focused on real-time EEG data and generative art techniques.

3. Development Framework

In the proposed development framework, the EEG-based interaction approach handles brainwave inputs without additional equipment. In order to acquire EEG data, MindWave Mobile 2 EEG headset [14] device was chosen because of its economic affordability, reliability in the delivery of data related to the user's mental states, ease of use, and the comfort of the user feel when interacting with the device. Figure 1 shows how the EEG headset integrates with the TouchDesigner [15] to create a development environment. MindWave Mobile 2 EEG headset was connected to the computer via Bluetooth. Real-time EEG data acquisition, signal processing, and classification were used in ThinkGear connector installed on the computer that sent the classification data through a user datagram protocol (UDP) over the local network to the computer running TouchDesigner software which generates contents.

3.1. *Contents creation*

To generate watercolor effects, TouchDesigner was used as a visual programming environment. The Noise texture operation (TOP) generates a variety of noise patterns

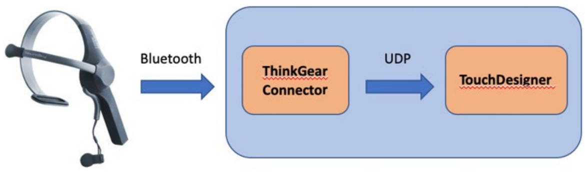


Fig. 1. Block Diagram of Brainwave Controlled

including Perlin, simplex, sparse, alligator, Hermite, and random which often used to procedurally generate organic surface textures and real-time effects. The Noise texture operation and the displacement texture operation nodes were used to create watercolor effects (Figure 2). In TouchDesigner texture operations, *Slope-TOP* used horizontal luminance and vertical luminance for Red and Blue channels while Green channel was kept neutral. *Blur-TOP* used as a filter with 5-pixel size, and *Level-TOP* used to quantization brightness and contrast.

3.2. *Mapping colors*

As literature described [6], the direct mapping technique was used. The classified eight EEG band power values are mapped to the colors as shown in Table 1. High Alpha and High Beta mapped into translate sampling parameters of Y-axis and Z-axis in Noise TOP (Simplex 3D GPU) which produce real-time dynamic noise patterns.

Table 1. EEG bands, their frequencies and colors

EEG band	Frequency range	Mapped Colors
Delta	1-3 Hz	Red
Theta	4-7 Hz	Orange
Low Alpha	8-9 Hz	Yellow
High Alpha	10-12 Hz	Green
Low Beta	13-17 Hz	Light Blue
High Beta	18-30 Hz	Blue
Low Gamma	31-40 Hz	Indigo
High Gamma	41-50 Hz	Violet

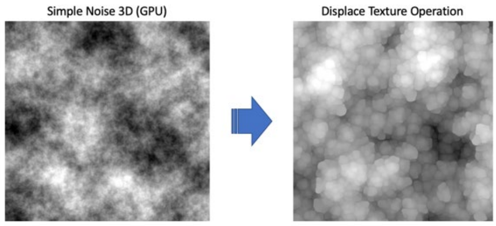


Fig. 2. The *Noise-TOP* and the *Displace-TOP* can be generated in TouchDesigner.

According to the ThinkGear Serial Stream SDK, the acquired eSenses Attention meter value is reported on a relative eSense scale of 1 to 100 which can be categorized into five different levels. However, this application used an indirect mapping technique which categorized a value between 0 to 40 at any given moment in time is considered "reduced". A value between 41 to 100 indicates "increased" levels of the eSense. The two states represent 0 and 1, which operates the reset parameter of the Feedback texture operation (*Feedback TOP* [15]) node. The *Feedback TOP* can be used to create a feedback effect that gives motion blur without cleaning the color buffer. When the "reduced" state, *Feedback TOP* passes through the image connected to input colors generated by EEG band values.

4. Experimental Results

As a pilot experiment, the ThoughtMix application was tested with graduate students from Tainan National University of the Arts. Figure 3 shows some examples of the participant's experiences and the effects created by their own brain waves, the left column shows dynamic attention and the right column shows stable and enhanced attention.

5. Conclusion

The main objective of this application is to provide an immersive experience to influence creativity and art therapy by visualizing the watercolor effects the user's mind creates. Despite the limited user study, the system successfully generated and incorporated EEG data which allows the user to identify the relation between the cybernetic-aesthetics and brain wave activities.

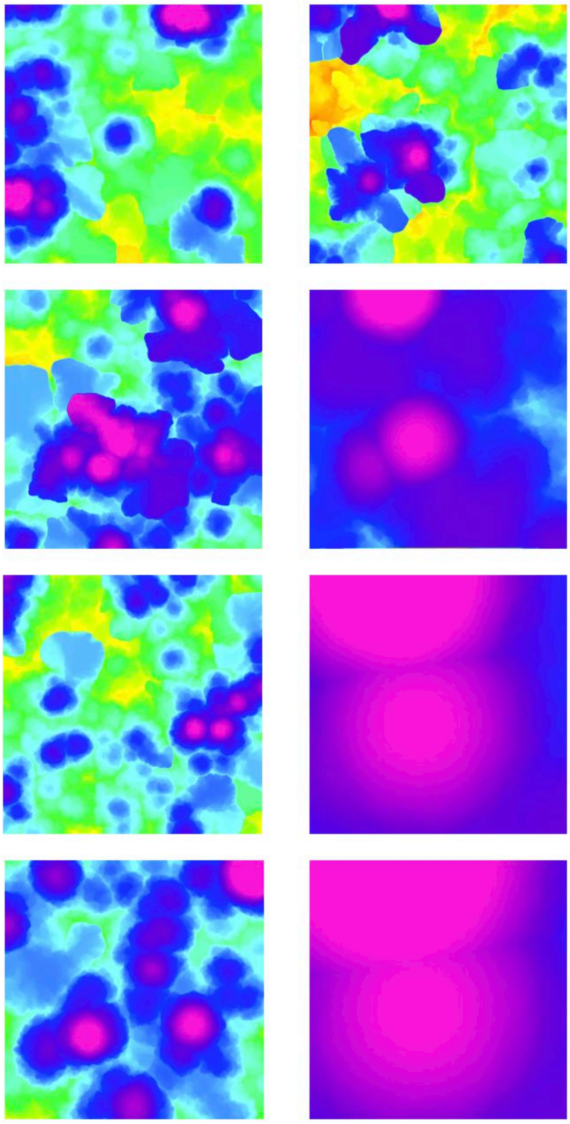


Fig. 3. Some examples of the participant's experiences and the watercolor effects created by their own brain waves.

References

1. S.U. Ahmed, A. Al Mahmud, K. Bergaust (2009) Aesthetics in Human-Computer Interaction: Views and Reviews. In: Jacko J.A. (eds) Human-Computer Interaction. New Trends. HCI 2009. Lecture Notes in Computer Science, vol 5610. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02574-7_63
2. E. Edmonds, G. Turner, L. Candy, Approaches to interactive art systems. In: Proceedings of the 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia, pp. 113–117, 2004.
3. Z. Wang, Theory of Digital Interactive Art Creation. Shanghai University, Shanghai (2011)
4. J. Strehovec, New media art as research: art-making beyond the autonomy of art and aesthetics. Technoetic Arts, Vol. 6, No. 3, pp. 233–250, (2009). https://doi.org/10.1386/tear.6.3.233_1
5. Suzanne Dikker, Matthias Oostrik, Peter Burr, Diederik Schoorl, and Matthew Patterson Curry. 2014. Mutual Wave Machine. Art. Retrieved May 3, 2019 from https://mai.art/content/mutual-wave-machine
6. M. Prpa, P. Pasquier, Brain-Computer Interfaces in Contemporary Art: A State of the Art and Taxonomy, Brain Art. Springer, (2019), https://doi.org/10.1007/978-3-030-14323-7_3
7. K. P. Fishkin, A taxonomy for and analysis of tangible interfaces. Personal and Ubiquitous Computing, Vol. 8, No. 5, pp. 347–358 (2004)
8. A. Wadeson, A. Nijholt, C. S. Nam, Artistic brain-computer interfaces: state-of-the-art control mechanisms. Brain-Comput. Interfaces Vol. 2(2–3), pp. 70–75, (2015).
9. P. Galanter, "What is generative art? Complexity theory as a context for art theory." In GA2003–6th Generative Art Conference. 2003
10. A. Dorin, J. McCabe, J. McCormack, G. Monro & M. Whitelaw (2012) A framework for understanding generative art, Digital Creativity, Vol. 23, No. 3-4, pp. 239-259, DOI: 10.1080/14626268.2012.709940
11. Shuhei Matsuyama, dyebirth: An installation work, https://nor.tokyo/dyebirth, Accessed on 10th Dec 2020.
12. Refik Anadol. 2018. Melting Memories. Installation Art, http://refikanadol.com/works/melting-memories/ Accessed on 10th Dec 2020.
13. Thinking in Color, https://www.lbbonline.com/news/this-experience-uses-brain-waves-to-create-colour, Accessed on 10th Dec 2020.
14. NeuroSky, ThinkGear Connector, http://developer.neurosky.com/docs/doku.php?id=thinkgear_connector_tgc. Accessed on 10th Dec 2020.
15. TouchDesigner, https://derivative.ca/, Accessed on 10th Dec 2020.

HaptWarp: Soft Printable and Motion Sensible Game Controller

Lee Jen Tun
School of Knowledge Science, Japan Advanced Institute of Science and Technology,
1 Chome-1 Asahidai Nomi, Ishikawa 923-1211, Japan[*]

R.P.C Janaka Rajapakse
†Graduate Institute of Animation and Film Art, Tainan National University of the Arts, No. 66, Daqi, Guantian Dist.,
Tainan City 72045, Taiwan [†]

Kazunori Miyata
School of Knowledge Science, Japan Advanced Institute of Science and Technology,
1 Chome-1 Asahidai Nomi, Ishikawa 923-1211, Japan[‡]

E-mail: [*]bmpsst511@gmail.com, [†]janakaraja@gmail.com, [‡]miyata@jaist.ac.jp

Abstract

Many interaction techniques have been introduced for controlling virtual reality contents and navigation. Traditional controllers need to map between two-dimensional inputs and three-dimensional actions while interactive surfaces enable more natural approaches, they still lack tactile feedback. We present real-virtual bend and twist, a tactile feedback technology that delivers the twist and bends effect into a virtual environment. The method uses two inertial measurement unit (IMU) sensors fixed on both sides on the 3D printed soft material bar to acquire object deformation to simulate the virtual deformation based on the changes in the angle exerted on a physical device. People can customize the 3D printed connection shape to apply to different virtual scenarios.

Keywords: Interaction, Game Controller, interactive bend, interactive twist, intuitive interaction

1. Introduction

During the development of haptic game control systems, careful consideration and devising are required to ensure system stability. Usually, efforts are focused on hardware and firmware improvement and as a result, an intuitive control method is of little concern. VR hand controllers are not dissimilar. Often a touchpad, trigger, or set of arrow keys are added seemingly as an afterthought. This may be sufficient in controlling systems for most of the games but for providing an intuitive experience and making it easier for users to understand how to interact in three dimensional virtual worlds, they often perform inadequately.

Over the past decade, the development of relatively three-dimensional haptic controllers from several companies has allowed experimentation in control with one important addition: inertia measurement unit (IMU). Our work has been applied to add a sensor system to a twistable and squeezable game controller through 3D printing, with the goal of giving users an intuitive operating experience. The software communicates with these systems and formulates a novel interaction between the player and the game.

Related work is presented in section 2. Section 3 explains the implementation of our system and the conclusion is stated in section 4 respectively.

2. Related Work

Research projects related to intuitive control and haptic applications aim to create an immersive experience, which can support the haptic feedback. We will not extract and review the entire area, but instead, discuss areas of the closest relevance to the work presented here.

Spatial user interaction plays an important role in Virtual Environments (VE). It allows the user to navigate the entire virtual environment to find a target or investigate a place or a virtual object in the world. In order to implement spatial interaction, this will often require 6 degrees of freedom (DOF). Inertial measurement unit (IMU) sensors play an important role to implement space manipulation [7]. Tseng et al [5] present the design of the EZ-Manipulator, a new 3D manipulation interface using smartphones that supports mobile, fast, and ambiguity-free interaction with 3D objects. Oh et al [6] propose an interaction method where users can conveniently manipulate a virtual object with touch interaction recognized from the inertial measurement unit (IMU) attached to the index finger's nail and head movements tracked by the IMU embedded in the HMD.

To enhance the virtual environment, game controllers are usually used in VR. Strasnick et al [3] present a novel design of linkage, electro-mechanically actuated physical connections capable of rendering variable stiffness between two commodity handheld virtual reality (VR) controllers. Cho et al [4] utilizes a spindle and wheel to present the interaction technique where users can twist or roll a button ball within his/her fingers simultaneously with other hand motions. For the haptic feedback, Heo et al [8] uses a single 6-DOF force sensor and a vibrotactile actuator to render grain vibrations to simulate the vibrations produced during object deformation based on the changes in force or torque exerted on a device. Tokuyama et al [9] proposed a torque display method for free-form deformation (FFD) with a 6-DOF haptic device which combines bending and twisting operations to extend the editing capabilities of the constructed shape deformation system by means of screw-motions. Tamaki et al [10] achieves this by placing the Photo-reflectors over specific muscle groups on the forearm to detect hand gestures as input. For output, electrode stimuli pads are placed over the same muscles to control the user's hand movements. Both sensors and electrode stimuli pads target main muscle groups responsible for controlling the hand.

While these approaches effectively provide specific spatial interaction and 3D object manipulation and interaction with people between real-world and virtual environment, most researches need many expensive peripherals or deliver the haptic feedback from outside stimuli actuator which unafforded to the user and doesn't integrate well with virtual reality. The aim of the research in this paper was to develop an inexpensive intuitive game controller that can simultaneously provide the intuitive manipulation for VR applications.

3. Implementation

The system structure is shown in Figure 1. In the physical environment, people usually do most things by their hand motions. The controller digitalizes the hand motions through wireless transmission protocol into a virtual environment. In the virtual environment, players are able to interact with the virtual object in an intuitive way by their daily hand motions.

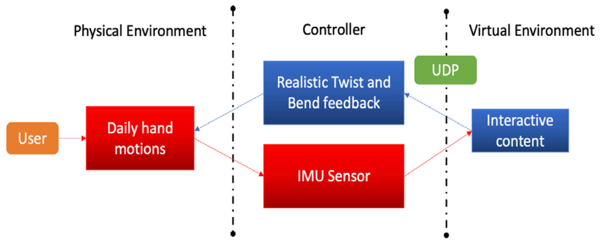


Fig. 1. This is the system structure of the research. Physical Environment include the daily hand motions. Controller is equipped with a 3D printed soft bar, IMU sensor and one MCU. Virtual Environment were the two created scenarios for the user to interact.

1.1. *Intuitive Control*

We explored how to integrate technology into the human experience by the development process of the game controller. It is interesting when a device becomes so natural that it is almost like a daily real activity in the virtual environment. Simultaneously, there are many advanced approaches in the user interfaces that aim to make the game controller more natural and intuitive, however, systems that achieve a truly integrated experience where the interface becomes part of a person's experience of themselves remains rare. At the forefront of this integration of technology and humanity, sensors and customized 3D printed components were designed specifically to integrate with the game controllers. A device like Tactical haptics [1] and Drag:on [2], transform the way a person can interact with the virtual environment. A common thread emerges when reviewing the game controller technologies that successfully integrate into the human experience, which is that it must provide an intuitive experience. In our project, it quickly became clear that it would be vital for our controller design to deliver a realistic and intuitive experience during gaming. As game controllers become wireless and more related to realistic manipulation, it becomes possible to embed them into a customized game application.

1.2. *Hardware Design*

The development of the game controller utilized a rapid prototyping methodology with various features trialled and refined to produce the final design. In particular, the prototyping involved considerable experimentation with different types of shape and mounting for the case and other electronic components. One of the early test prototypes is shown in Figure 2. The controller was printed by the 3D printer with soft material to provide customized joints for a variety of game scenarios. The IMU sensor was embed in the controller to deliver the bend and twist angle from the player to the Wemos D1 mini which is like an Arduino interface but provides WIFI with a small size all for a great price. Then Wemos D1 mini through the UDP protocol transmits the data to Unity virtual environment.

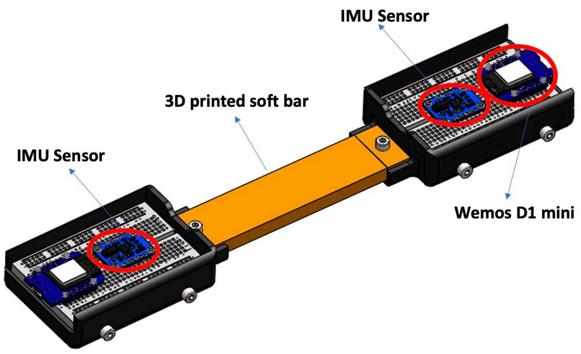


Fig. 2. This is prototype CAD of the HaptWarp controller.

1.3. *Wireless Data Transmission*

Wireless Communication is the fastest growing and most vibrant technological area especially integrated with game controllers. Wireless Communication is a method of transmitting information from one point to another, without using any connection like wires, cables, or any physical medium. Apart from mobility, wireless communication also offers flexibility and ease of use. We chose the UDP protocol since the situation in our research is a real-time data transmission case when the controller is delivering data that can be lost because newer data coming in will replace the previous data under a high frequency.

1.4. *Virtual Content*

When considering what application can be enhanced by Haptwarp controller, VR games came to our minds first. In this game, players can do daily activities from certain scene objects with the Haptwarp controller as shown in Figure 3 and Figure 4. While doing different natural activities, the player can perceive different tactile sensations from the controller. For example, while opening the door in a virtual environment, the user will receive the visual and haptic feedback of feeling the force from the 3D printed bar. When the user twists the controller, the screwdriver will lock the screw in the virtual environment. By combining those actions, the player needs to explore the virtual world by moving around himself, using the controller to trigger the game event, receiving the new clue, and figuring out how to escape from the chamber.

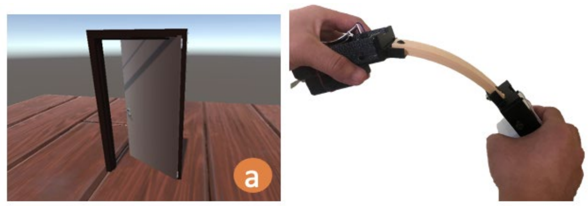


Fig. 3. Users can bend the controller to simulate the open-door activity in virtual environment.

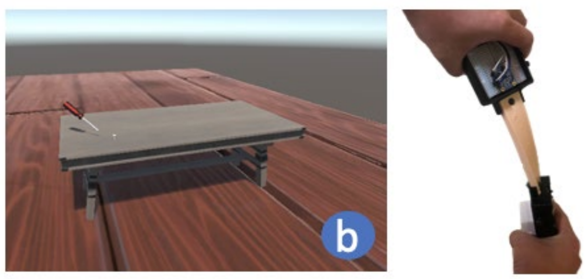


Fig. 4. Users can twist the controller to simulate the activity of locking a screw.

4. Conclusion

In summary, this paper has performed an intuitive game controller that can provide the user with the act of interacting and controlling virtual objects with sensors. This system approaches the unique capability of delivering the human hand motions and haptic sensations with wireless communication, as compared to conventional techniques. Furthermore, to validate the efficacy of this approach, we developed two interactive contents in the virtual environment for the user to experience the intuitive manipulation with the HaptWarp controller.

In our future work, we will optimize the controller. For the 3D printed soft bar, we focus on enhancing the immersive feeling by designing a variety of shapes of linkage able to bring more daily hand motions that can be applied to multiple events in the virtual environment. For the sensors, we focus on setup variety bio-sensor to acquire user body information which will be applied into virtual environments to activate different content to enhance the immersive experience. Finally, we will conduct further studies to investigate the complementary effect and the tactile illusions in VR for both partial and full body experiences.

1. Acknowledgements

The authors thank the Ministry of Science and Technology for financially and technically supporting this research in part under Grant MOST (IV) MOST 109-2218-E-002-038- and the project from National Taiwan University.

2. References

1. B. Lang, "Hands-on: Tactical haptics' vive demo is further proof that VR needs more than rumble," 2015. [Online]. Available:https://web.archive.org/web/20160312220656/http://www. roadtovr.com/tactical-haptics-htc-vive-demo-virtual-realityfeedback-rumble/
2. André Zenner and Antonio Krüger. 2019. Drag:on: A Virtual Reality Controller Providing Haptic Feedback Based on Drag and Weight Shift. *In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, Paper 211, 1–12. DOI: https://doi.org/10.1145/3290605.3300441
3. Evan Strasnick, Christian Holz, Eyal Ofek, Mike Sinclair, and Hrvoje Benko. 2018. Haptic Links: Bimanual Haptics for Virtual Reality Using Variable Stiffness Actuation. *In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems(CHI '18)*. Association for Computing Machinery, New York, NY, USA, Paper 644, 1–12. DOI: https://doi.org/10.1145/3173574.3174218
4. I. Cho and Z. Wartell, Evaluation of a bimanual simultaneous 7DOF interaction technique in virtual environments, *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, Arles, 2015, pp. 133-136, doi: 10.1109/3DUI.2015.7131738
5. Tseng, PH., Hung, SH., Chiang, PY. *et al.* EZ-Manipulator: Designing a mobile, fast, and ambiguity-free 3D manipulation interface using smartphones. *Computational Visual Media* **4,** 139–147 (2018). DOI: https://doi.org/10.1007/s41095-018-0105-0
6. Oh, Ju Y.; Park, Ji H.; Park, Jung-Min. 2019. "Virtual Object Manipulation by Combining Touch and Head Interactions for Mobile Augmented Reality" *Appl. Sci.* 9, no. 14: 2933.
7. Li Qian, Wu Hao, and Wang Peng. 2018. Manipulator Trajectory Planning and Control Method Based on IMU. *In Proceedings of the 2018 International Conference on Robotics, Control and Automation Engineering (RCAE 2018)*. Association for Computing Machinery, New York, NY, USA, 132–136. DOI: https://doi.org/10.1145/3303714.3303721
8. Seongkook Heo, Jaeyeon Lee, and Daniel Wigdor. 2019. PseudoBend: Producing Haptic Illusions of Stretching, Bending, and Twisting Using Grain Vibrations. *In Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. Association for Computing Machinery, New York, NY, USA, 803–813. DOI: https://doi.org/10.1145/3332165.3347941
9. Yoshimasa Tokuyama, R. P. C. Janaka Rajapakse, Yuji Nakazawa and Kouichi. Konno, Torque display method for free-form deformation with haptic device, In the Proceedings of *2009 ICCAS-SICE*, Fukuoka, 2009, pp. 3803-3808.
10. Emi Tamaki, Terence Chan, and Ken Iwasaki. 2016. UnlimitedHand: Input and Output Hand Gestures with Less Calibration Time. *In Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 163–165. DOI: https://doi.org/10.1145/2984751.2985743

Gray Level Co-Occurrence Matrix (GLCM) and Gabor Features Based No-Reference Image Quality Assessment for Wood Images

Heshalini Rajagopal[1,2], Norrima Mokhtar[1], Anis Salwa Mohd Khairuddin[1], Wan Khairunizam[3], Zuwairie Ibrahim[4], Asrul Bin Adam[4] and Wan Amirul Bin Wan Mohd Mahiyidin[1]

[1]Department of Electrical Engineering, Faculty of Engineering, University of Malaya, Malaysia

[2]Department of Electrical and Electronic Engineering, Manipal International University, Malaysia

[3] School of Mechatronic Engineering, University of Malaysia Perlis, Malaysia

[4] College of Engineering, University of Malaysia Pahang, Malaysia

[1]heshalini@gmail.com

Abstract

Image Quality Assessment (IQA) is an imperative element in improving the effectiveness of an automatic wood recognition system. There is a need to develop a No-Reference-IQA (NR-IQA) system as a distortion free wood images are impossible to be acquired in the dusty environment in timber factories. Therefore, a Gray Level Co-Occurrence Matrix (GLCM) and Gabor features-based NR-IQA, GGNR-IQA algorithm is proposed to evaluate the quality of wood images. The proposed GGNR-IQA algorithm is compared with a well-known NR-IQA, Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) and Full-Reference-IQA (FR-IQA) algorithms, Structural Similarity Index (SSIM), Multiscale SSIM (MS-SSIM), Feature SIMilarity (FSIM), Information Weighted SSIM (IW-SSIM) and Gradient Magnitude Similarity Deviation (GMSD). Results shows that the GGNR-IQA algorithm outperforms the NR-IQA and FR-IQAs. The GGNR-IQA algorithm is beneficial in wood industry as a distortion free reference image is not required to pre-process wood images.

Keywords: Wood images, GLCM, Gabor, GGNR-IQA, NR-IQA

.

1 INTRODUCTION

Wood is extensively used for furniture, building construction and paper production[1]. There are various types of wood and each of them has different attributes with regard to its formation, thickness, colour and texture[2]. These varying characteristics defines their ideal usages and economic values[3]. As the price and characteristics of every wood species differs, misclassification may cause financial losses. Hence, there is a need to identify different wood species accurately.

Conventionally, the recognition of wood species is performed manually by human subjects[4]. However, this practice is time and cost consuming to the lumber industry. Hence, several automatic wood species recognition systems have been developed[1,2,5,6]. The efficiency of automatic wood recognition systems can be improved by using superior quality microscopy images which are commonly enhanced to improve the rate of successful wood species recognition. Nevertheless, the image enhancement processes consume extra computational time, and could cause a checkerboard artefact to the wood images[7]. In addition, the dusty and dark environment in timber factories could degrade the quality of the image acquired[8]. Thus, an appropriate image quality assessment (IQA) algorithm is required to assess the acquired wood images prior to feeding it to any automatic wood recognition system.

IQA can be categorised into subjective and objective evaluations. Subjective evaluation is the scores given by human subjects based on their judgment on the image quality. While, objective evaluation is done based on numerical methods to determine the quality of the images. Even though, subjective evaluation is the benchmark of IQA, it is impracticable in an industrial environment as it is time and cost consuming. Therefore, it is necessary to develop an objective evaluation procedures that is capable to imitate subjective IQA evaluation[9].

Objective evaluation can be categorised into Full-Reference-IQA (FR-IQA), Reduced Reference-IQA (RR-IQA) and No-Reference/Blind IQA (NR-IQA)[10,11]. FR-IQA uses the reference image fully to assess the images whereas RR-IQA uses the reference images partially. In contrast, NR-IQA assesses an image without using a reference image. NR-IQA is the most appropriate algorithm to evaluate the quality of the wood images as it may be impossible to obtain high quality images in the dusty and dark setting of lumber factories. Therefore, we propose the Gray Level Co-Occurrence Matrix (GLCM) and Gabor features-based NR-IQA, GGNR-IQA algorithm to evaluate wood images.

The GLCM and Gabor features are widely used in wood species recognition system[5,12–14]. The proposed GGNR-IQA algorithm is compared with a commonly utilized NR-IQA, Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE), and FR-IQAs namely, Structural Similarity Index (SSIM)[15], Multiscale SSIM (MS-SSIM) [15], Feature Similarity (FSIM)[16], Information Weighted SSIM (IW-SSIM)[17] and Gradient Magnitude Similarity Deviation (GMSD)[18]. The performances of the GGNR-IQA, BRISQUE and FR-IQAs are evaluated by using the Pearson Linear Correlation Coefficient (PLCC) and Root Mean Squared Error (RMSE) computed between the human mean opinion scores (MOS) and the algorithms.

2 METHODS

2.1. Training and Testing Database

An SVR model is trained with the GLCM and Gabor features calculated for normalized wood images with the human MOS which are obtained from the subjective evaluation for wood images. The MOS, GLCM and Gabor features are utilized as the training and testing database to obtain an optimized SVR model

2.1.1 Wood Images

Ten wood images from various wood genus, as shown in **Fig. 1** were chosen. The images were acquired from a wood database: https://www.wood-database.com/ [19]. The ten reference images were distorted by Gaussian white noise and motion blur. These two types of distortions usually occur in the industrial setting. Generally, the wood images are exposed to Gaussian white noise due to the poor illumination and heat in the lumber mill while acquiring the wood images[8,20]. On the other hand, wood images are exposed to motion blur when there is a relative motion between the wood slice and camera[6].

These distortions degrade the quality of the wood images where the features of the pores on the wood texture may not be discerned. Hence, this may lead to misclassification of the wood genus as the feature extractor may not obtain distinctive features from the wood images efficiently[21].Nine modulations of Gaussian white noise with standard deviation, σ_{GN} and motion blur with standard deviation, σ_{MB} were added to the reference images, i.e.: σ_{GN} = 10, 20, 30, 40, 50, 60, 70, 80 and 90 for Gaussian white noise and σ_{MB} = 2, 4, 6, 8, 10, 12, 14, 16 and 18 for motion blur.

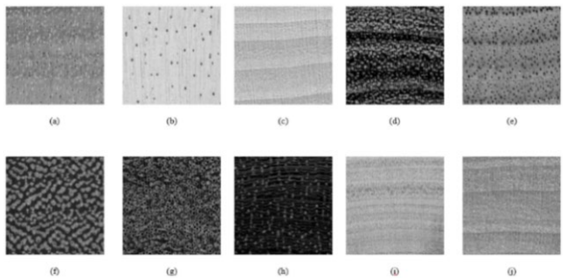


Fig 1. Ten wood images used as reference images (a) Turraeanthus africanus, (b) Ochroma pyramidale, (c) Tilia americana, (d) Cordia spp., (e) Juglans cinerea, (f) Vouacapoua americana, (g) Dipterocarpus spp., (h) Swartzia Cubensis, (i) Cordia spp., (j) Cornus florida

2.1.2 GLCM and Gabor Features

First, Mean Subtracted Contrast Normalized (MSCN), $\hat{I}(m,n)$ is calculated from the wood image, *I(m,n)* using Eq. (1) [22]:

$$\hat{I}(m,n) = \frac{I(m,n)-\mu(m,n)}{\sigma(m,n)+1} \tag{1}$$

where $\mu(m,n)$ and $\sigma(m,n)$ denote the mean and variance of wood image, *I(m,n)*, respectively $,m \in 1,2,\ldots,M, n \in 1,2,\ldots,N$ are spatial indices while M represents the height and N represents width of image, *I(m,n)*.

The mean, $\mu(m,n)$ and variance, $\sigma(m,n)$ of the wood image are computed using Eq. (2) and Eq. (3), respectively [22]:

$$\mu(m,n) = \sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l} I_{k,l}(m,n) \tag{2}$$

$$\sigma(m,n) = \sqrt{\sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l}\left(I_{k,l}(m,n)-\mu(m,n)\right)^2} \tag{3}$$

where $w = \{w_{k,l} | k = -K,\ldots,K,\ l = -L,\ldots,L\}$ is a 2-dimension (2D) circularly-symmetric Gaussian weighting function that is sampled out to three standard deviations and rescaled to unit volume, and K and L represent the window sizes.

The MSCN coefficients, $\hat{I}(m,n)$ highlights the main features of the wood images such as pores and grains, with few low-energy residual object boundaries[23]. Therefore, the MSCN is used to compute the GLCM and Gabor features instead of the image, *I(m,n)*. Next, two types of features namely, GLCM and Gabor features were incorporated in this study.

2.1.2.1 GLCM Features

The GLCM depicts second order statistical analysis of an image by analyzing how often the pairs of pixels which consist of specific values and spatial relationship take place in an image. The probability, $p(m,n)$ is computed using Eq. (4) [24]:

$$p(m,n) = \{C(m,n)|(d,\theta)\} \quad (4)$$

where d is the inter-pixels displacement distance,θ denotes orientation and $C(m,n)$ denotes the frequency of gray level occurrence in MSCN of the image, $\hat{I}(m,n)$. Four statistical textures such as contrast, correlation, energy, and homogeneity were extracted from the GLCM matrix.

Contrast calculates the local variations in the gray-level co-occurrence matrix and is defined as Eq. (5) [24]:

$$\text{Contrast} = \sum_{m,n} |m-n|^2 p(m,n) \quad (5)$$

Correlation computes the joint probability occurrence of the specified pixel pairs and is defined as Eq. (6) [24]:

$$\text{Correlation} = \sum_{m,n} \frac{(m-\mu m)(n-\mu n)p(m,n)}{\sigma_m \sigma_n} \quad (6)$$

Energy calculates the sum of squared components in the GLCM. It is also known as uniformity or the angular second moment. The energy parameter is computed as Eq. (7) [24]:

$$\text{Energy} = \sum_{m,n} p(m,n)^2 \quad (7)$$

Homogeneity calculates the closeness of the distribution of elements in the GLCM to the GLCM diagonal and is computed as Eq. (8) [24]:

$$\text{Homogeneity} = \sum_{m,n} \frac{p(m,n)}{1+|m-n|} \quad (8)$$

These 4 parameters were computed at four directions, 0°, 45°, 90° and 135° and this form 16 GLCM features.

2.1.2.2 Gabor Features

The 2D Gabor function which represents the spatial summation properties of simple cells in the visual cortex and it is defined as Eq. (9) [25]:

$$g(x,y;\lambda,\theta,\psi,\sigma,\gamma) = exp\left(-\frac{x\prime^2+\gamma^2 y\prime^2}{2\sigma^2}\right) cos\left(2\pi\frac{x\prime}{\lambda}+\psi\right) \quad (9)$$

Where

$$x' = x\,cos\,\theta + y\,sin\,\theta \quad (10)$$

$$y' = -x\,sin\,\theta + y\,cos\,\theta \quad (11)$$

λ denotes the wavelength of the sinusoidal factor, θ denotes the orientation of the normal to the parallel stripes of a Gabor function, ψ represents the phase offset, σ represents the standard deviation of the Gaussian envelope and γ represents the spatial aspect ratio.

The computational models of 2D Gabor filters are defined in Eq. (12) and (13) [25]:

$$h_e = g(x,y)cos\big(2\pi f(x\cos\theta + y\sin\theta)\big) \quad (12)$$

$$h_o = g(x,y)sin\big(2\pi f(x\cos\theta + y\sin\theta)\big) \quad (13)$$

Where h_e and h_o represents the even symmetric and odd symmetric Gabor filters, respectively and $g(x,y)$ represents the isotropic Gaussian function and is computed as Eq. (14) [25]:

$$g(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{x\prime^2+y\prime^2}{2\sigma^2}\right) \quad (14)$$

And the spatial frequency response of the Gabor functions, f is as shown in Eq. (15) [25]:

$$f = N/P \quad (15)$$

Where N denotes the size of the kernel and P denotes period in pixel.

In this study, wavelength, λ is in increasing powers of two starting from $4/\sqrt{2}$ up to the hypotenuse length of the input image[26] and this produces 7 Gabor features. The 7 Gabor features were then computed in four orientations, 0°, 45°, 90° and 135°, similar to the GLCM computations. This forms 28 features Gabor features. In total, the 16 GLCM and 28 Gabor features were combined and this forms 44 features. These 44 features were calculated using the MSCN coefficients, $\hat{I}(m,n)$ and are used to train SVR.

2.1.3 MOS

The MOS values were obtained from subjective evaluation participated by 10 students aged between 20-25 years from Manipal International University (MIU), Malaysia. The evaluation was carried out as per the procedures suggested in Rec. ITU-R BT.500-11[27] where it was performed in an office environment using a 21-inch LED computer screen.

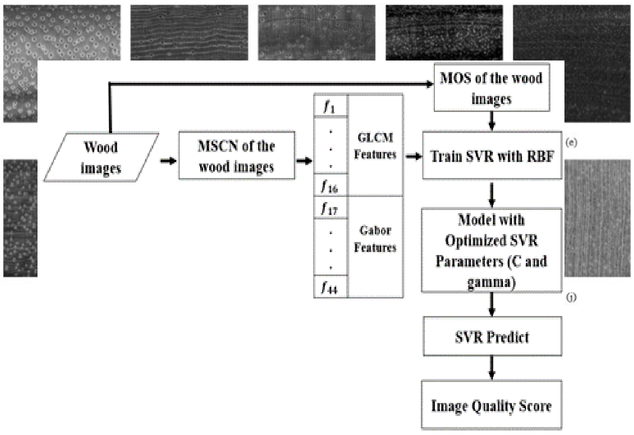
Simultaneous Double Stimulus for Continuous Evaluation (SDSCE) approach was used in this evaluation [27,28] where the reference and distorted images are shown side-by-side on the computer screen and each subject compares the quality of the images displayed on the right side with its reference image (left side) to evaluate the displayed image.

The score given by the human subjects are either Excellent (5), Good (4), Fair (3), Poor (2) or Bad (1) for each image displayed. The evaluation process takes 15 to 20 minutes for each subject. The scores obtained from the subjects were averaged to convert them to the MOS[29]. These MOS values are also used to train SVR.

2.1.4 Support Vector Machine Regression (SVR)

$\in -$SVR [30] is trained using MOS and 44 GLCM and Gabor features of wood images in this study. The 44 image features (GLCM and Gabor features) calculated for the wood images are mapped to the MOS values of the corresponding wood images.

The 44 GLCM and Gabor features and MOS of wood images were randomly split into training and testing sets where 80% of the 44 features and MOS values were used to



train the SVR model to obtain an SVR model with optimized parameters and 20% were utilized to evaluate the optimized SVR model.

The flow diagram of the proposed GGNR-IQA is shown in **Fig. 2**. The performance of GGNR-IQA was evaluated using PLCC[31] and RMSE[32] calculated between 1000 iterations were performed on the training and testing of the SVR model to obtain an optimized SVR model. The cost parameter, C, and width parameter, g, of the optimized SVR model are 32768 and 0.125, respectively.

Fig 2. Flow Diagram of the Proposed GGNR-IQA

2.2. Performance Evaluation

The proposed GGNR-IQA is compared with a well-known NR-IQA algorithm, BRISQUE and five FR-IQAs [21]: Structural Similarity Index (SSIM)[15], Multiscale SSIM (MS-SSIM)[15], Feature Similarity (FSIM)[16], Information Weighted SSIM (IW-SSIM)[17] and Gradient Magnitude Similarity Deviation (GMSD)[18].

The performance of the GGNR-IQA, BRISQUE and FR-IQAs is assessed using PLCC and RMSE[28] values calculated between these algorithms and MOS.

3 RESULTS AND DISCUSSIONS

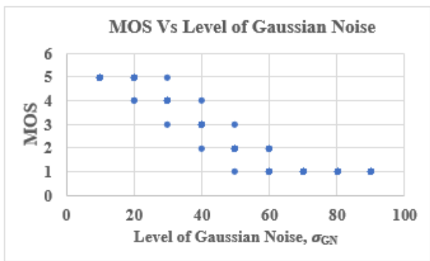
The efficiency of the GGNR-IQA was further assessed using a second dataset which was generated from the same wood image database [19]. This second dataset was produced using ten reference images acquired from ten various wood genus as shown in **Fig. 3**.

Fig 3. Reference wood images in the second dataset (a) Julbernardia pellegriniana,, (b) Dalbergia cultrate, (c) Dalbergia retusa, (d) Dalbergia cearensis, (e) Guaiacum officinale, (f) Swartzia spp., (g) Dalbergia spruceana, (h) Dalbergia sissoo, (i) Swartzia benthamiana and (j) Euxylophora paraensis

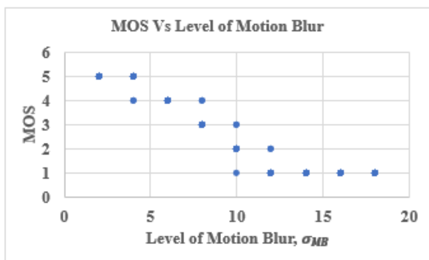
These reference images were added with the similar distortion type (Gaussian white noise and motion blur) and modulations as the training and testing database. This means that the second dataset includes 10 reference images and 180 distorted images.

3.1. Relationship Between MOS and Quality of Image with Different Distortion Modulations

Fig. 4 (a) and (b) shows the correlation between MOS and nine distortion modulations of Gaussian white noise and motion blur. Lower MOS values show lower image quality which is caused by higher distortion modulation. On the other hand, higher MOS values represent higher image quality which is generated by lower distortion modulation. Based on **Fig. 4** (a) and (b), the MOS decreases as the distortion modulation increases. This means that all the human subjects could discern the images distorted with the various modulations of Gaussian white noise and motion blur.



(a)



(b)

Fig 4. Scatter Plot of MOS versus nine distortion modulations of (a) Gaussian White Noise and (b) Motion Blur

3.2. Correlation between GGNR-IQA, BRISQUE and FR-IQAs Algorithms and MOS

The PLCC and RMSE values calculated between GGNR-IQA, BRISQUE, five FR-IQA algorithms: Structural Similarity Index (SSIM), Multiscale SSIM (MS-SSIM), Feature SIMilarity (FSIM), Information Weighted SSIM (IW-SSIM), Gradient Magnitude Similarity Deviation (GMSD) and MOS for Gaussian white noise, motion blur and overall images are shown in **Fig. 5** and **Fig. 6,** respectively. The most suitable IQA for wood images is expected to have the highest PLCC and lowest RMSE values.

Based on **Fig. 5** and **Fig. 6**, the GGNR-IQA has the highest PLCC and lowest RMSE values for Gaussian white noise, motion blur and the overall database compared to the BRISQUE and FR-IQAs. This shows that the GGNR-IQA algorithm outperforms BRISQUE and all the five FR-IQAs.

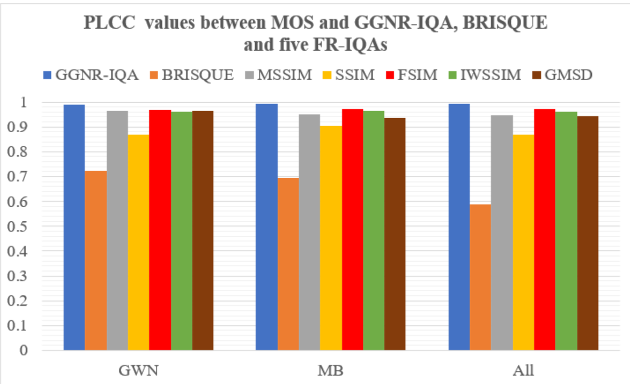


Fig. 5. PLCC values between GGNR-IQA, BRISQUE, FR-IQAs and MOS

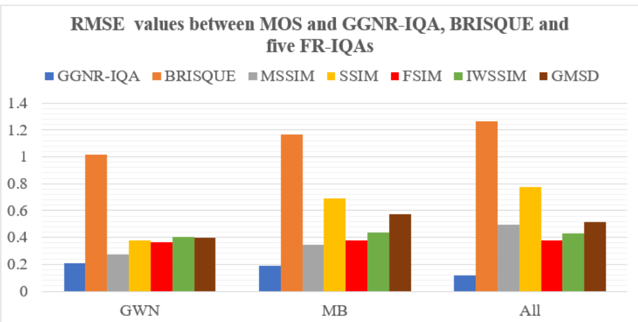


Fig. 6. RMSE values between GGNR-IQA, BRISQUE, FR-IQAs and MOS

4 CONCLUSIONS

A NR-IQA algorithm, GGNR-IQA was proposed to assess wood images prior to feeding the image to wood species classification and recognition system. The proposed GGNR-IQA algorithm was trained using GLCM, Gabor features and MOS obtained from wood images. The performance of the GGNR-IQA algorithm was assessed by comparing the PLCC and RMSE values calculated between GGNR-IQA, BRISQUE, five FR-IQA algorithms and MOS. PLCC and RMSE values showed that the GGNR-IQA algorithm outperforms BRISQUE and all the five FR-IQAs.

This shows that the GGNR-IQA algorithm could assess the quality of wood images accurately. In addition, the proposed GGNR-IQA algorithm would not require a distortion free reference image to determine the quality of the wood images. This is beneficial especially when it is impossible to obtain a distortion free reference image in the dusty environment of lumber mill.

REFERENCES

[1] Shivashankar S. Automatic Wood Classification using a Novel Color Texture Features. *Int J Comput Appl*. 2018;180(27):34-38.

[2] Barmpoutis P, Dimitropoulos K, Barboutis I, Grammalidis N. Wood species recognition through multidimensional texture analysis. *Comput Electron Agric*. 2018;144:241-248. doi:10.1016/j.compag.2017.12.011

[3] Zamri MIP, Cordova F, Khairuddin ASM, Mokhtar N, Yusof R. Tree species classification based on image analysis using Improved-Basic Gray Level Aura Matrix. *Comput Electron Agric*. 2018; (124):227-233.

[4] doi:https://doi.org/10.1016/j.compag.2016.04.004

[5] Gazo R, Wells L, Krs V, Benes B. Validation of automated hardwood lumber grading system. *Comput Electron Agric*. 2018;(June):0-1. doi:10.1016/j.compag.2018.06.041

[6] Venkatachalapathy K, Sudhakar P. An Intelligent Recognition System for Identification of Wood Species. *J Comput Sci*. 2014;10(7):1231-1237. doi:10.3844/jcssp.2014.1231.1237

[7] Guang-sheng C, Peng Z. Sensors and Actuators A : Physical Dynamic wood slice recognition using image blur information. *Sensors Actuators A Phys*. 2012;176:27-33. doi:10.1016/j.sna.2011.12.056

[8] Xiao B, Tang H, Jiang Y, Li W, Wang G. Brightness and contrast controllable image enhancement based on histogram specification. *Neurocomputing*. 2018;275:2798-2809. doi:10.1016/j. neucom. 2017.11.057

[9] Ratnasinga J, Ioras F, Swan TT, Yoon CY, Thanasegar G. Determinants of Occupational Accidents in the Woodworking Sector: The Case of the Malayasian Wooden Furniture

Industry. *J Appl Sci*. 2011;11(3):561-566. doi:10.3923/jas.2011.561.566

[10] Wang Z. Applications of objective image quality assessment methods. *Signal Process Mag IEEE*. 2011;28(6):137-142. doi:10.1109/MSP.2011.942295

[11] Gulame M, Joshi KR, Kamthe RS. A Full Reference Based Objective Image Quality Assessment. *Int J Adv Electr Electron Eng*. 2013;2(6):13-18.

[12] Chandler DM. Seven Challenges in Image Quality Assessment: Past, Present, and Future Research. *ISRN Signal Process*. 2013;2013:1-53. doi:10.1155/2013/905685

[13] Khalid M, Lee ELY, Yusof R, Nadaraj M. Design of an intelligent wood species recognition system. *Int J Simul Syst Sci Technol*. 2008;9(3):9-19.

[14] Tou JY, Tay YH, Lau PY. A Comparative Study for Texture Classification Techniques on Wood Species Recognition Problem. In: *2009 Fifth International Conference on Natural Computation*. IEEE; 2009:8-12. doi:10.1109/ICNC.2009.594

[15] Bremananth R. ; Nithya B. ; Saipriya R. Wood Species Recognition Using GLCM and Correlation. In: *2009 International Conference on Advances in Recent Technologies in Communication and Computing*. Vol 51. ; 2009:431-444. doi:https:/https:/10.1109/ARTCom.2009.10

[16] Wang Z, Simoncelli EP, Bovik AC. Multi-Scale Structural Similarity for Image Quality Assessment. In: *Asilomar Conference on Signals, Systems and Computers*. Vol 2. IEEE; 2003:1398–1402.

[17] Zhang L, Zhang D, Mou X. FSIM: a feature similarity index for image quality assessment. *Image Process IEEE Trans*. 2011;20(8):2378-2386.

[18] Wang Z, Li Q. Information content weighting for perceptual image quality assessment. *IEEE Trans Image Process*. 2011;20(5):1185-1198. doi:10.1109/TIP.2010.2092435

[19] Xue W, Zhang L, Mou X, Bovik AC. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *Image Process IEEE Trans*. 2014;23(2):684-695.

[20] Meier E. The Wood Database. https://www.wood-database.com/. Published 2007. Accessed August 16, 2018.

[21] Rahman T, Haque MR, Rozario LJ, Uddin MS. Gaussian Noise Reduction in Digital Images Using a Modified Fuzzy Filter. In: *International Conference on Computer and Information Technology (ICCIT)*. ; 2014:217-222.

[22] Rajagopal, H., Khairuddin, A.S.M., Mokhtar N et al. Application of image quality assessment module to motion-blurred wood images for wood species identification system. *Wood Sci Technol*. 2019:1-15.

[23] Mittal A, Moorthy AK, Bovik AC. No-Reference Image Quality Assessment in the Spatial Domain. IEEE Transactions on Image Processing. doi:10.1109/TIP.2012.2214050

[24] Heshalini Rajagopal, Norrima Mokhtar TFTMN, Izam WKWA. No-reference quality assessment for image- based assessment of economically important tropical woods. *PLoS One*. 2020:1-15. doi:10.1371/journal.pone.0233320

[25] Abd Latif, MH, Md. Yusof, H, Sidek, S.N, Rusli N. Implementation of GLCM Features in Thermal Imaging Affective State Detection. In: *IEEE International Symposium on Robotics and Intelligent Sensors*. Vol 76. ; 2015:308-315. doi:10.1016/j.procs.2015.12.298

[26] Rubiyah Yusof, Nenny Ruthfalydia Rosli MK. Using Gabor Filters as Image Multiplier for Tropical Wood Species Recognition System. In: *International Conference on, Computer Modeling and Simulation, International Conference*. ; 2010:289-294. doi:10.1109/UKSIM.2010.61

[27] Jain, Anil K. and FF. Unsupervised Texture Segmentation Using Gabor Filters. *Pattern Recognit*. 1991;24(12):1167-1186.

[28] Recommendation ITURBT (2002). 500-11. *Methodology for the Subjective Assessment of the Quality of Television Pictures*.; 2000.

[29] Chow LS, Rajagopal H, Paramesran R. Correlation between subjective and objective assessment of magnetic resonance (MR) images. *Magn Reson Imaging*. 2016;34(6):820-831. doi:10.1016/j.mri.2016.03.006

[30] Bindu K, Ganpati A, Sharma AK. A Comparative Study of Image Compression Algorithms. *Int J Res Comput Sci*. 2012;2(5):37-42. doi:10.7815/ijorcs.25.2012.046

[31] Li Sze Chow HR. Modified-BRISQUE as no reference image quality assessment for structural MR images. *Magn Reson Imaging*. 2017;43:74–87. doi:10.1016/j.mri.2017.07.016

[32] Chang, Chih-Chung and Lin C-J. LIBSVM: A library for support vector machines. *ACM Trans Intell Syst Technol*. 2011;2(3):27:1--27:27.

[33] Song X-K. *Correlated Data Analysis: Modeling, Analytics, and Applications*. Springer Science & Business Media; 2007.

[34] Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. *Geosci Model Dev*. 2014;7(3):1247-1250. doi:10.5194/gmd-7-1247-2014

Design and development of automated seeding and irrigation system

Anirban Kumar[1], Heshalini Rajagopal[1]
[1]*Department of Electrical and Electronic Engineering, Manipal International University, Malaysia*
[1]*anirban_kumar1993@yahoo.com*

Abstract

Malaysia, has a very conducive environment for agriculture. Six percent of the population are employed to some of the agriculture activities making agriculture the most curtail industry in Malaysia. Hence, there is a need to develop the agricultural facilities by incorporating latest technological advancements. Conventional seeding process is time consuming and requires additional labor cost. In this paper, an automated system is proposed for seeding as well as irrigation process in agriculture which reduces the labor cost. This system aims to increase the efficiency of the seeding process without affecting the nature of soil. The proposed system is equipped with Arduino MEGA and Arduino UNO which acts as the main control unit while ultrasonic and soil moisture sensors are used to detect the obstacles and soil moisture level, respectively. The robot consists of a funnel like arrangement to perform the seeding procedure. The field is equipped with moisture sensors placed at different areas that monitors the moisture level of the soil on a regular interval for irrigation purposes. The proposed system will be of great benefit to the future endeavor of agricultural business as well as it will be able to optimize the seeding and irrigation.

Keywords: Agriculture, Seeding, Irrigation, Arduino MEGA, Ultrasonic sensors, Soil moisture sensors.

1 Introduction

Agriculture is the oldest and yet an important economic activity for any given country. There are distinct available types of machines for harvesting, ploughing, pesticide control etc. Most of these machines are manually operated to execute the required action. Different machines are used to perform separate action which leads to lower profit returns [1]. Seeding is a tedious yet a crucial activity for a farmer. Seeding of a large farmland is even more tedious, which demands for big group of personnel for manual seeding. To minimize the personnel required for seeding, harvesting etc., agriculture machineries were developed. Planting of seeds manually, leads to low spacing efficiency and seed placement, serious health issues and tedious task [2]. Ideal solution to overcome the deficiencies discussed above is by developing a machine that carries out more than one action and by atomization of the developed machine, which will lead to intensification of the yield.

The primary aim of this paper is to design and develop an efficient seeding and irrigation system. This system consists of an Arduino MEGA, Arduino UNO, ultrasonic sensors, DC motor, servo motors and moisture sensors. The scissor lift is designed using Fusion 360 and 3D printed. Arduino MEGA and UNO holds all the commands, receive all the parameters from the sensors and send output signals accordingly. The ultrasonic sensors will be used for the object detection and moisture and DHT 22 sensor will continuously monitor the moisture of the soil, humidity and temperature of the surroundings on a regular interval and alerts the monitoring system.

2 Prior Work

M.Aravind Kumar et al. (2018) have proposed a machine for seeding and irrigation to limit the working expense and decrease the ideal opportunity for planting seed & irrigation operations by using solar energy to operate the system. There is a sharp pointed iron attached to the front of the robot to dig the soil for seed sowing, once the seed is planted the supporter at the back end of the robot covers the dug land with soil. Versatile computerized water system robot was built to irrigate the plants appropriately [3]. However, this study does not irrigate the planted seed immediately after the seeding process is completed.

Md. Didarul Islam Sujon et al. (2018) have proposed an autonomous prototype robot that will help farmers in the farmland. It is an Arduino controlled robot that carries out ploughing, sowing and watering process. The robot carries out farming using ultrasonic sensors in order to change its position from one farming strip to another. The robot aims contribute greatly in developing the farming strategies and reduce farmers cost [4]. However, this robot can carry out farming for land area of limited land area only.

K.A.Sunitha et al. (2017) have proposed a robotic system that localizes the path and can navigate itself without human action. For ploughing, this robot is equipped with tentacles that has saw blades. The locomotion is provided with wheels covered under conveyor belts. For each rotation every tooth on gear will take seeds and will drop them on field. Camera at the front end tracks the path for every fixed distance and at the minimum distance it takes the path pre-programmed [5]. However, this study has a drawback where the system uses camera and image processing for object detection which makes the system complex. In addition, the object detection can be inaccurate if the image captured is subjected to motion blur due to the dirty camera lens and camera movement.

S. Thawali et al. (2017) have proposed robot capable of performing operations like automatic ploughing, seed dispensing and pesticide spraying autonomously and also provides manual control when required. Initially the robot digs the entire field simultaneously dispensing seeds side by side. On the field the robot operates on automated mode. For manual control the robot uses the Remote controller as control device and helps in the navigation of the robot on the field [6]. However, there is no obstacle detection system and hence there are chances that this robot will runover planted plats if there is no human presence to deviate the robot from the predefined path.

Thorat Swapnil V et al. (2017) have proposed a sowing machine which is operated. This machine is capable of planting different types and different sizes of seeds and also the spacing between two seeds can be varied. Seed storage tank is connected at the top of the robot near rear wheels. Front sensor serves the function of guiding the robot. As any obstacle comes in front of robot it gives the signal to the robot and diverts the path of robot [7]. However, the system does not monitor the soil or plant conditions so the farmers still have to invest in systems that monitor those two variables.

3 Methodology

Fig 1 shows the block diagram for the motion, seeding, irrigation, obstacle detection and avoidance. The main control unit of this system is Arduino MEGA which is powered by a 9V battery. Ultrasonic sensors will give input to the Arduino MEGA regarding the obstacle detection. Arduino MEGA will analyse the received inputs and control the output to the lifting servos. DC motors are to be powered by a 7.7V Lithium Polymer battery; the motors are used for the movement of the robot. Arduino MEGA controls the motor direction and speed (RPM) using L298N module.

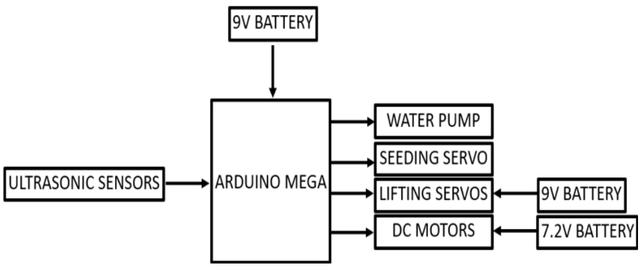


Fig. 1: Block Diagram for Seeding, Irrigation and Obstacle Avoidance

Fig. 2 shows the block diagram for the soil moisture, humidity and temperature monitoring system. Soil moisture and DHT 22 sensors will give input to the Arduino UNO regarding the soil moisture level, temperature and humidity of the surrounding respectively. The input information from the sensors will be sent to the Android Application and SMS via GSM module.

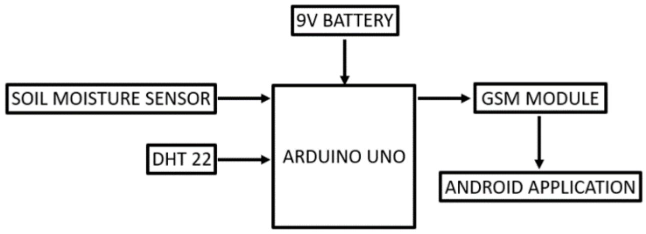


Fig. 2: Block Diagram for Soil Moisture, Humidity and Temperature Monitoring System

3.1. Circuit Diagram and Simulation

3.1.1. Arduino DC Motor Control

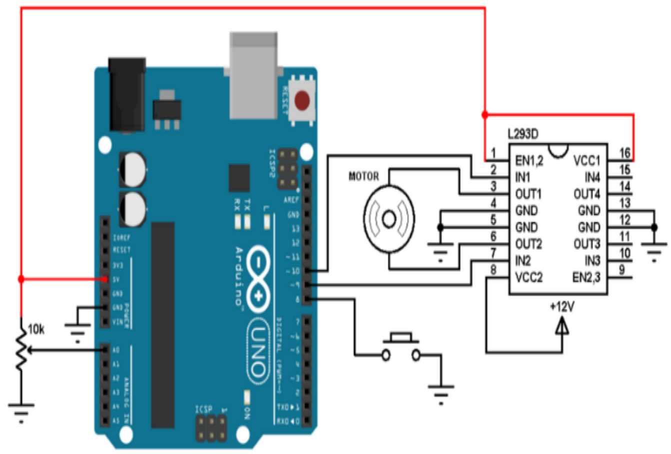


Fig. 1: Arduino DC Motor Control

Fig. 3 shows the circuit diagram of Arduino DC motor control. Only three Arduino pins are required, pin 8 is for the push button which controls the direction of motor rotation. Pins 9 and 10 are PWM signal outputs, at a given time there is only one active PWM, this allows for the control of direction as well as the speed by varying the duty cycle of the PWM signal. The active PWM pin decides the motor direction of rotation (one at a time, the other output is logic 0). The speed of the DC motor (both directions) is controlled with the 10kΩ potentiometer which is connected to analog channel 0 (A0). The L293D driver has 2 VCCs: VCC1 is +5V and VCC2 is +12V. Pins IN1 and IN2 are the control pins and their functions are shown in Table 1.

IN1	IN2	Function
L	H	Direction 1
H	L	Direction 2
L	L	Fast motor stop
H	H	Fast motor stop

Table 1: IN1 and IN2 function

3.1.2. Lifting System

Fig. 4 shows the circuit diagram for the lifting system that uses ultrasonic sensor, Arduino and motor.

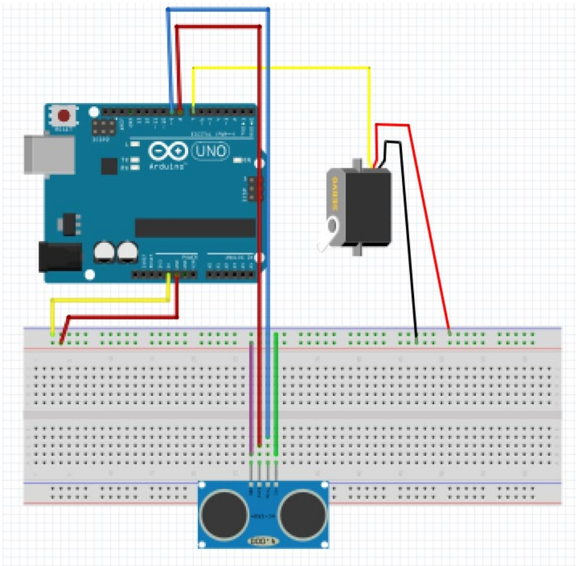


Fig 4: Lifting System

The lift like system will be used to propel the robot upwards once the ultrasonic sensor detects any obstacle, the system will be propelled upwards to a predefined height. Once the robot reaches the predefined height, the plant will be irrigated and then the robot will move forward to irrigate the plants ahead. When all the plants in that segment are irrigated the robot will be brought down.

3.2 Flow Chart

The various steps undertaken to develop the code to operate the proposed system is shown in Fig 5. Once the hardware design and assembly of the seeding and irrigation robot is completed, code is developed to carry out the desired action. During the coding process mainly loops like “while”, “if” and “for” will be used to achieve the desired movements.

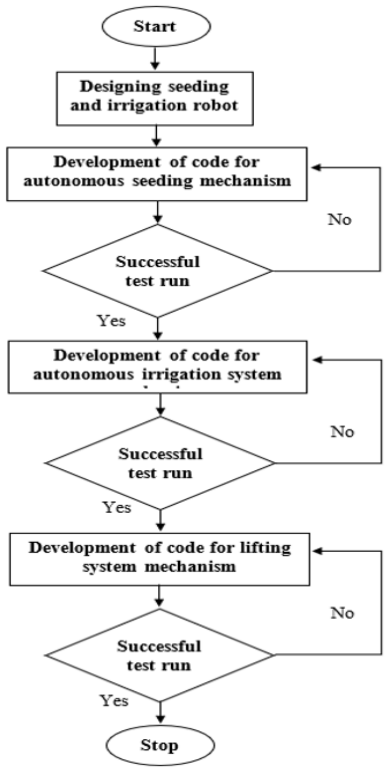


Fig. 5: Flowchart for Software Development

4 Results

4.1. Seeding Mechanism

The first objective of the system is to design an automated robot to perform seeding and irrigation procedure. The robot takes ten seconds to initialize. After initialization robot moves forward for two seconds at the speed of 200 rpm and then comes to a halt to facilitate the seeding and irrigation process, respectively. Once the seeding and irrigation process is completed, the robot moves forward for two more times before it turns to the next row. The robot is programmed to drops the seed at an interval of six seconds and 7 cm. Fig. 6 shows the seeding process of mung dhal

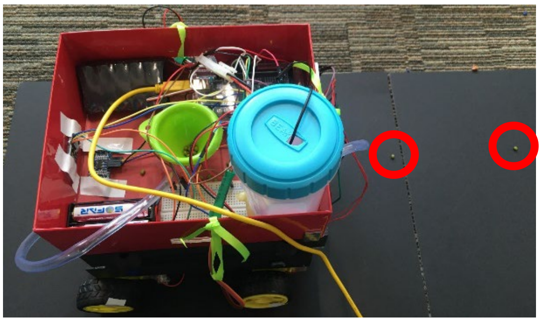


Fig. 6: Seeding Process of Mung Dhal

4.2. Irrigation Mechanism

Fig. 7 shows that the robot performs irrigation to the plant after the seeding process. For the design of water harvesting systems, it is necessary to assess the water requirement of the crop intended to be grown. For crops like mung dhal and peas initially 20ml to 25ml water is required for irrigating each seed. Table 2 shows the irrigation details of the robot prototype.

Fig. 7: Robot performing irrigation

Table 2: Irrigation Detail

Irrigation Tank Capacity: 500ml
Irrigation Pump Flow Rate: 1.2-1.6 L/min
Irrigation per seed: 20ml - 30ml
Irrigation type: Drip irrigation

4.3. Monitoring System

The second objective of the study is to develop a temperature and soil moisture monitoring system by using Android Application to enhance the user accessibility. The soil moisture, humidity and temperature monitoring system is in place for the landowner or the user to monitor the soil conditions of the separate segments using the Android App which was created using the MIT App Inventor. The app is designed to display the information regarding the soil moisture, humidity and temperature level of each land segment. The layout of the App is shown in Fig. 8.

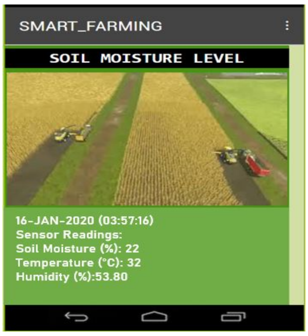


Fig. 8: Monitoring the Soil Condition Using Android App

The monitoring system is beneficial for the user, because it helps to keep track of the soil moisture, the surrounding environment and carry out on demand irrigation. It also helps the farmers to store the data for future weather prediction and planning on land management based on the data collected.

4.4. Lifting System

Thirdly, this paper aimed to design and develop a lifting system to avoid obstacles while maneuvering and assist in irrigation. The robot is equipped with two ultrasonic sensors to detect obstacles at the front and base of the robot. When front ultrasonic sensor detects an obstacle at distance of 10 cm or less, the system stops and activates the lifting system to avoid the obstacles. When the system has fully lift upwards, the base ultrasonic sensor activates, and the robot continues to move forward. The base ultrasonic sensor is placed in order to bring the lifted system back to its original level if distance between the lifted system and ground is more than 20 cm and if there is no obstacle in the front. Fig. 9 shows that the robot is lifted when there is an obstacle in front.

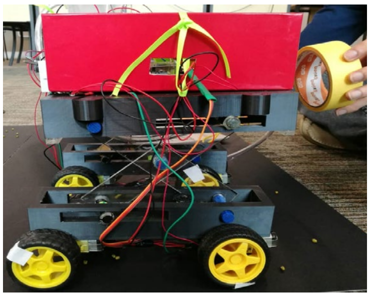


Fig. 9: Robot Lifts Up

The robot is capable to run for about 4km in a single battery charge.

5 Conclusions

The proposed system is capable to perform seeding followed by irrigation procedure. The robot is equipped with eight wheels to facilitate the movement of the robot. The robot is also equipped with two ultrasonic sensors of any obstacle at the front and base of the robot to detect obstacles in the front and under the robot. The robot lifts up before moving forward when it detected obstacles in the front and comes back down with the assistance of the base ultrasonic sensor when the obstacle has been passed. This also enables the system to water the tall plants. Soil moisture and DHT22 sensors are fixed on the land to sense the soil moisture, humidity and temperature. The readings of these parameters are sent to the user vis SMS every one minute. In addition, the user can view the readings of the parameters via Android App. The robot moves forward once and perform the seeding at one point, irrigate the seed and moves to the next point for the seeding process. The robot dispenses seeds automatically at a fixed distance (7 cm).

References

[1] A. A. M. M. Amrita Sneha.A, "Agricultural Robot for Automatic Ploughing and Seeding," *IEEE International Conference on Technological Innovations in ICT for Agriculture and Rural Development,* pp. 17 - 23, 2015.

[2] M. L. K. P. G. V. P. R. N. Thorat Swapnil V, "Design and Fabrication of Seed Sowing Machine," *International Research Journal of Engineering and Technology (IRJET),* vol. 04, no. 09, pp. 704 - 707, 2017.

[3] M. Kumar, A. S. Reddy and K. Sagadevan, "Automatic Seed Sowing & Irrigation Agribot using Arduino," *International Journal of Pure and Applied Mathematics,* vol. 119, no. 14, pp. 1089-1093, 2018.

[4] M. D. I. Sujon, R. Nasir, M. M. I. Habib, M. I. Nomaan, J. Baidya and M. R. Islam, "Agribot: Arduino Controlled Autonomous Multi-Purpose Farm Machinery Robot for Small to Medium Scale Cultivation," *International Conference on Intelligent Autonomous Systems,* pp. 155-159, 2018.

[5] K. A. Sunitha, G. S. G. S. Suraj, C. P. N. Sowrya, G. A. Sriram, D. Shreyas and T. Srinivas, "Agricultural Robot Designed for Seeding Mechanism," *IOP Conference Series: Materials Science and Engineering,* 2017.

[6] S. Thawali, B. Yadav, H. Rumde and R. S. Sewane, "Design and Manufacturing of Robot for Digging and Seeding in Agriculture," *International Conference on Ideas, Impact and Innovation in Mechanical Engineering,* vol. 5, no. 6, p. 1637 – 1644, 2017.

[7] M. L. K. P. G. V. P. R. N. Thorat Swapnil V, "Design and Fabrication of Seed Sowing Machine," *International Research Journal of Engineering and Technology,* vol. 4, no. 9, pp. 704-707, 2017.

Investigation of A Real-Time Driver Eye-Closeness for the Application of Drowsiness Detection

Muhammad Zubir bin Kamazlan[1], Wan Khairunizam[1], Abdul Hafiz Halin[1], M. Rudzuan M. Nor[1], Azian Azamimi Abdullah[2], Norrima Mokhtar[3]

[1]Depart. of Mechatronic Eng., Faculty of Electrical Engineering Technology, University Malaysia Perlis, Malaysia
[2]Depart. of Biomedical Elec. Eng., Faculty of Electronic Engineering Technology, University Malaysia Perlis, Malaysia
[3]Department of Electrical Engineering, Faculty of Engineering, University of Malaya, Malaysia
[1]khairunizam@unimap.edu.my

Abstract

The increase in accident and death rates due to drowsiness while driving raises concerns to the community. An efficient solution is vital to ensure the safety of all drivers on the road. Most previous studies have analyzed drowsiness using head tilt, yawning, and eye condition. Face detection applied in drowsiness detection from previous research not included distances between subject and camera. The features used for eye detection required large storage and long-term process which are not applicable in a real-time system. This study uses Haar algorithm and analysis is performed based on the size of the region of interest for face detection. Eye monitoring uses facial landmark features and the evaluation is dependent on the width of the eye. The percentage of eye closure is used to describe the eyes as closed. This study only takes into account the normal rate of blinking eyes while driving because of the long-time constraints required for a person to be in a drowsy state. In this research, the Raspberry Pi 3B+ and Pi cameras are used as processing and vision devices. The highest accuracy of face detection achieved based on the ROI area at a distance of 80 cm is 98.33%. The lowest difference between eye width and the intercanthal distance is 0.36%. The overall normal eye blink rate while driving is in the range of the normal eye blink rate which does not exceed 20 blinks/min as reported by the previous researcher.

Keywords: Face detection, eye detection, drowsiness detection

1 INTRODUCTION

One of the potential applications of smart vehicle systems is the drowsy driver detection system [1]. Thus, the development of this system vital in preventing traffic accidents. This system works to measure the blink rate of the drivers based on eye-closeness. The eye condition contains critical information, and it is possible to determine a driver's level of drowsiness if such visual activity can be assessed [2].

Face detection is a primary component that requires a Pi camera that acts as a video sensor. The location of the eye should be determined before the eye's status can be specified using suitable image processing.

It has been confirmed in a U.S. Department of Transportation publication: "Only PERCLOS that are highly correlated with PVT blunders in and between subjects even though almost every innovation demonstrated the ability to diagnose drowsiness by detecting lapses in at least one subject or a subset of subjects" [3]. Increasing subjective sleepiness, decreasing efficiency, and the number of lapses in a visual reaction time task closely associated with PERCLOS values [4]. The proportion of time the eyelids cover at least 80% of the pupils is measured by PERCLOS [5].

PERCLOS is more dependable across drivers than EEG, blinks, and head position in the research since its algorithm was more than 90% precise in identifying deteriorated performance during alertness tasks as validated by Dinges et al. (1998) [6]. The eye conditions could be divided into a few sections relying on PERCLOS. Heartbeat rate, yawning, and head tilting are other terms that can calculate the level of drowsiness. However, this research only focuses on eye-closeness to determine the normal eye blink rate while driving due to limitations that required a long period for a person to fall in a drowsy state.

2 METHODS

2.1. Research Flow Chart

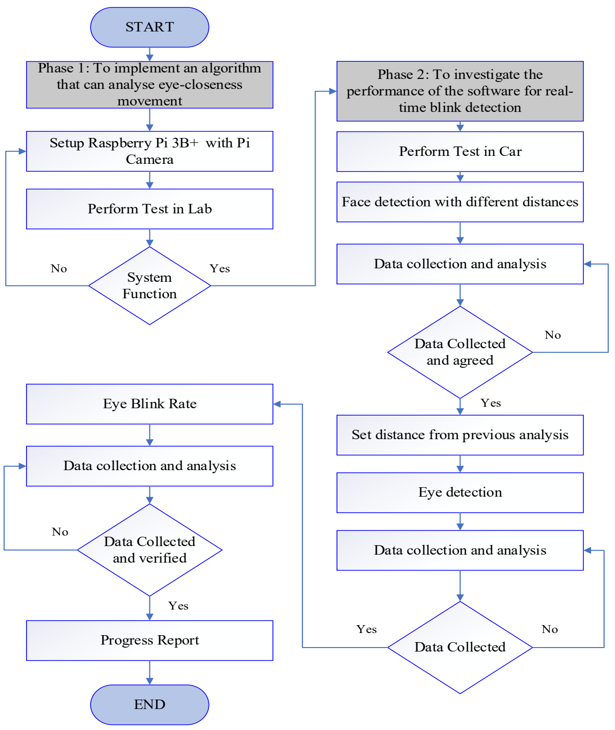


Fig. 1. Research flowchart

Fig. 1 shows the flow chart of the research. Phase #1 is to design an algorithm that analyze the eye-closeness movement. The setup of software and hardware is performed in this phase. The system functionality is tested before proceeding to phase #2. The aim of phase #2 is to investigate the performance of the software for real-time eye blinking. This phase is separated in three parts, which are face detection, eye detection and eye blink rate. The experiments are carried out part by part.

2.2. Software Setup

Raspberry Pi 3 Model B+ chosen as a computer base to execute the program. 5 MP Raspberry Pi camera also needed as a visual device and an ultrasonic sensor used to measure the distance between subject and pi camera. OpenCV-Python library is competent for this system to solve computer vision problems and obtain the required data. This library used the Haar Cascade Classifier (haarcascade_frontalface_default .xml) to detect the face in the frames. Dlib is a modern C++ open-source library that implements numerous machine learning models and supports functions such as threading and networking. This library enables the use of shape_predictor_68_face_landmarks.dat to extract facial features effectively. The system functionality tested in the laboratory first before tested the real driver in the car.

2.3. Experimental Setup

Eight (8) subjects from ages 23 to 25 involved in the face and eye detection tests while only four subjects with more than one-year driving experience being tested in normal blink rate evaluation. The experiments conducted from 10 am to 12 pm with sunny weather conditions. However, the dark-tinted car's window reduced the amount of light entering the car. Face and eye detection task conducted in a car at Universiti Malaysia Perlis. Normal Blink Rate experiment requires the subjects to drive on the highway from Pauh flyover to Universiti Malaysia Perlis as shown in Fig 2.

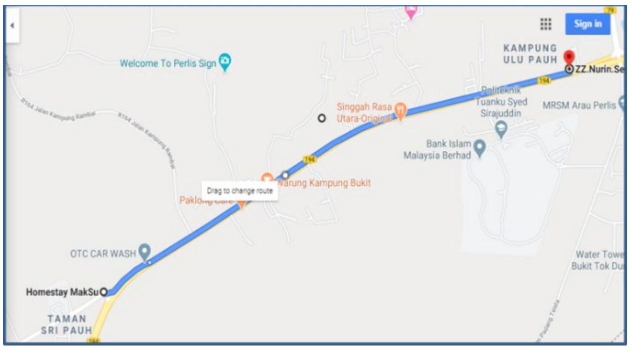


Fig. 2. Route used to measure blink rate while driving

2.4. Face Detection

The subjects placed in front of the camera as shown in Fig 3 with the value of x = 75 cm measured using the ultrasonic sensor in the car as the set distance between the seat and camera. The program executed to detect the face of the driver for one minute because the investigation of the system performance for the blink rate is based on every minute. The size of the rectangle as shown in Fig.4 on the driver's face was stored as a measured value. The process repeated with a value of x = 80 cm and x = 85 cm because the adjustable car seat has a 5 cm difference between each position. The program executed three times for each value of x to obtain the average value. The average and standard deviation of the face area of five subjects were measured and is set as the reference. Three subjects were tested and analyzed based on these reference value. The reference value will be used to determine the accuracy from the further data obtained. If the data are not within the standard deviation of the reference value, the data is not accurate or acceptable.

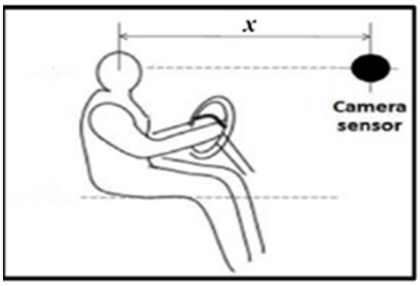


Fig. 3. Position of the subject in front of the camera

$$Area = w \times h \tag{1}$$

$$P = \frac{k}{m} \times 100$$

where:
P: Percentage Accuracy of the Area (%)
k: number of data (area) within standard deviation of reference value
m: total of data sample (60)

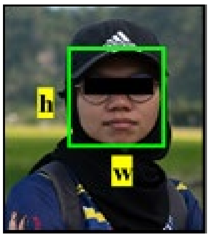


Fig 4. Face area based on ROI box

2.5. Eye Detection

Fig. 5. Intercanthal distance

$$Euclidean\ distance, d = \sqrt{(x2 - x1)^2 - (y2 - y1)^2} \tag{2}$$

$$intercanthal\ distance = width\ (right\ eye) = width(\ left\ eye) \tag{3}$$

$$(p39 \leftrightarrow p42) = (p36 \leftrightarrow p39) = (p42 \leftrightarrow p45) \tag{4}$$

The eye-detection proceeded using the 80 cm distance that has the most accuracy from face detection analysis. Facial landmarks were the process of tracking vital feature points on the face and had been used for distinguishing and portraying the distinctive regions of the face. The corners of each eye were located using the facial landmark predictor. The Euclidean distance between the points was determined. This task executed in one minute and repeated three times for each subject to determine the ability of the system to track the eye every minute. [7] stated that intercanthal distance in the average face is equal to the width of an eye. Thus, this task validates either the statement in [7] can be accepted or not.

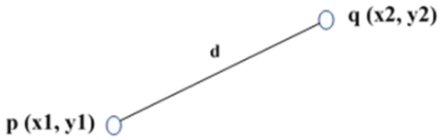


Fig. 6. Point axis

Fig. 7. Eye corner detection

2.6. Normal Blink Rate

This task performed within five minutes for each subject because the research analyses the blink rate for every minute. The Eye Aspect Ratio (EAR) from every subject while driving was recorded in this test. PERCLOS is the shortened form of eyelid closure over time [8]. PERCLOS (P80) applied to determine 80 % the percentage closure of the eyes in the preliminary test. The value of 80% of eye closure set as threshold. The purpose of this task was to validate the normal eye blink rate while driving as mentioned in [9] that the normal eye blink rate is 15 to 20 blinks/ min and the blink rate could decrease when the eyes focus on the work such as driving [10].

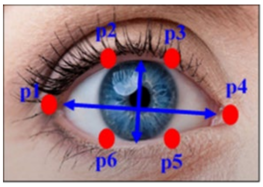


Fig. 8. Eye Aspect Ratio

$$EAR = \frac{||p2 - p6|| + ||p3 - p5||}{2||p1 - p4||} \quad (5)$$

$$EAR(both\ eye) = [EAR(left\ eye) + EAR(right\ eye)] \div 2 \quad (6)$$

$$Threshold = [\max(EAR) - \min(EAR)]\ x\ 0.2 + \min(EAR) \quad (7)$$

3 RESULTS AND DISCUSSIONS

3.1. Face Detection Analysis

Table 1. ROI box average area target on different distances for 5 subjects

Distance	Reference Target Area (mm^2)	Standard Deviation
75 cm	5116	480
80 cm	4376	217
85 cm	3932	216

Five (5) subjects are involved in assessing the average of the reference area of the ROI box for facial detection based on various distances. Three additional subjects are tested to assess the precision of the detection of the face. The data in this table is set as a reference for experiments of other subjects. If the data obtained in those tests are within the standard deviation, it is considered acceptable.

Table 2. ROI area for subject#1

Distance from subject to camera	Reference Area		Obtained		Percentage Accuracy of the Area (%)
	Area (mm^2)	Standard Deviation	Area (mm^2)	Standard Deviation	Note
75 cm	5116	480	4789	122	81.67
80 cm	4376	217	**4253**	**90**	**90.00**
85 cm	3932	216	3838	128	88.33

Subject #1 produces the average area of 4789 mm^2 for the distance 75 cm, 4253 mm^2 for the distance 80 cm, and 3838 mm^2 for the distance 85 cm. The experiments are run three times for each distance and only the average with the lowest standard deviation is chosen. Percentage accuracy for distance 75 cm which is 81.67 % measured using the equation 2.

Table 3. ROI area for subject#2

Distance from subject to camera	Reference Area		Obtained		Percentage Accuracy of the Area (%)
	Area (mm^2)	Standard Deviation	Area (mm^2)	Standard Deviation	
75 cm	5116	480	5437	134	81.67
80 cm	4376	217	**4393**	**81**	**98.33**
85 cm	3932	216	4071	109	83.33

Subject #2 produces the average area of 5437 mm^2 for distance 75 cm, 4393 mm^2 for distance 80 cm, and 4071 mm^2 for distance 85 cm. The experiments are run three times for each distance and only the average with the lowest standard deviation is chosen. Percentage accuracy for distance 75 cm which is 81.67 %.

Table 4. ROI area for subject#3

Distance from subject to camera	Reference Area		Obtained		Percentage Accuracy of the Area (%)
	Area (mm^2)	Standard Deviation	Area (mm^2)	Standard Deviation	
75 cm	5116	480	4840	168	80.00

80 cm	4376	217	**4413**	**89**	**96.67**
85 cm	3932	216	3743	91	86.67

Subject #3 produces the area of 4840 mm^2 for distance 75 cm, 4413 mm^2 for distance 80 cm, and 3743 mm^2 for distance 85 cm. The information gathered for the most recent subject is obtained value and the reference area is the aim value.

The highest percentage accuracy among three subjects is for the distance of 80 cm and the lowest is at 75 cm distance. The error in percentage accuracy resulted from the area of ROI that varies out of the reference target standard deviation. The errors in the value occurred due to the complex background [11] and the lighting condition during the test [12].The distance with highest accuracy is better to be applied in further test because the facial landmark point is depend on the ROI area. Thus, the highest accuracy will ensure that the position of facial landmark have the lowest variation.

3.2. Eye Detection Analysis

This task is performed with a distance of 80 cm from the camera to the subject because it has the highest accuracy of face detection. Table 5 records the eye width and intercanthal distance for each subject. Percentage of different measured the inequality between the average width of both eye and the intercanthal distance. As we can see, the lowest percentage difference is from subject 1 while others have a large percentage difference between eye width and intercanthal distance. Thus, only subject #1 fulfilled the statement in [7] that the eye width is equal to intercanthal distance on the average face. The other subject may have a broad range of measurements between eye width and intercanthal distances due to the low resolution of the vision device used.

Table 5. Eye width and intercanthal distance

Subject	Average eye width for both eye (mm)	Intercanthal Distance (mm)	Percentage of different
#1	9.945	9.909	0.36
#2	9.989	15.771	36.66
#3	9.048	15.614	36.05
#4	10.58	16.695	36.63
#5	10.85	16.968	36.05
#6	9.958	16.998	41.42
#7	10.069	17.362	42.01
#8	9.045	15.997	43.46

Fig 9 shows that only subject #1 has the smallest difference between eye width and intercanthal distance while the other has a large gap between both values. The eye width in the average face is not equal to the intercanthal distance as mentioned in [7] since seven over eight-person have a large percentage difference between those values.

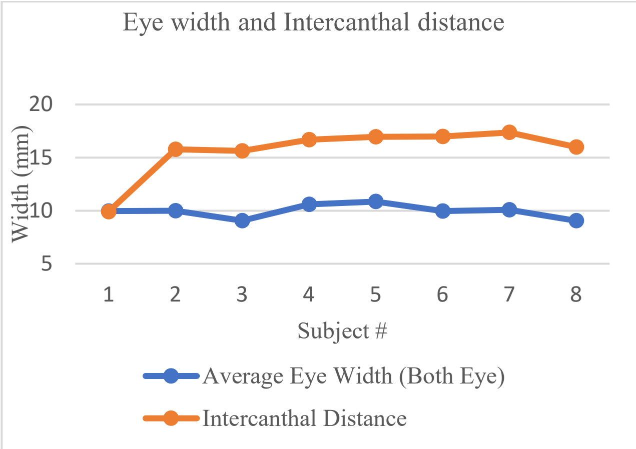


Fig. 9. Eye width and intercanthal distance

3.3. Eye Blink Rate Analysis

Table 6. Eye blink for subject #1

Time(minute)	Blink rate
1^{st}	4 blink/min
2^{nd}	5 blink/min
3^{rd}	6 blink/min
4^{th}	4 blink/min
5^{th}	5 blink/min

The normal eye blink rate for subject #1 for five minutes is listed in Table 6. Fig 10 shows the EAR plotted for subject #1. The threshold of the EAR for this subject 0.1814 determined by the minimum and maximum value of EAR which are 0.127 and 0.399. The value below the threshold is classified as a blink. The calculation of the threshold can be referred to equation 8.0.

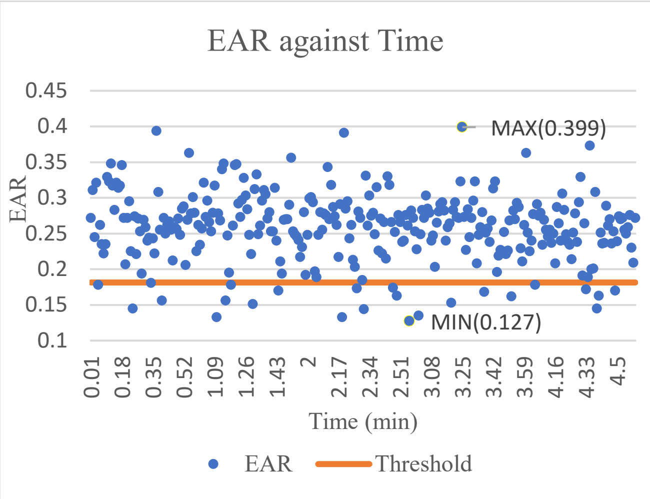


Fig. 10. The EAR for Subject #1

Table 7. Eye blink for subject #2

Time(minute)	Blink rate
1st	5 blink/min
2nd	4 blink/min
3rd	6 blink/min
4th	4 blink/min
5th	2 blink/min

Table 7 indicates five minutes of the normal eye blink rate for subject #2. The EAR for subject #2 is shown in Fig 11. The minimum and maximum values for EAR are 0.142 and 0.415 were used to determine the threshold value of 0.1966. The threshold is the value where 80% of the eyes are closed.

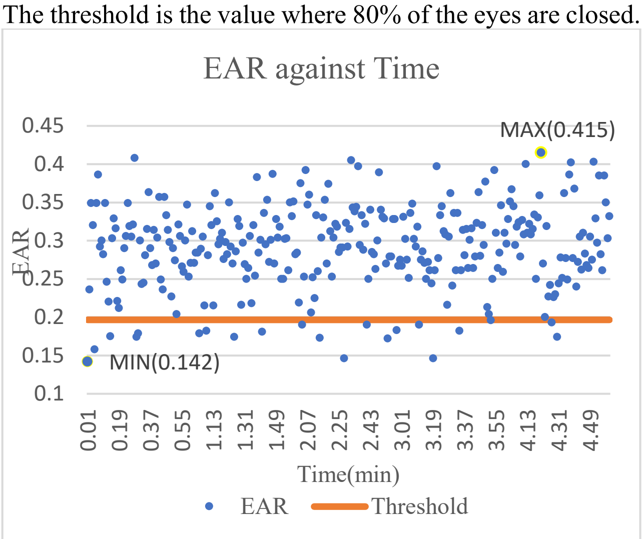


Fig. 11. The EAR for Subject #2

Table 8. Eye blink for subject #3

Time(minute)	Blink rate
1st	9 blink/min
2nd	5 blink/min
3rd	11 blink/min
4th	10 blink/min
5th	15 blink/min

The normal eye blink rate for subject 3 for five minutes is listed in Table 11. Fig 12 shows the EAR plotted for subject #3. The EAR threshold for this subject 0.2146 is determined by the minimum and maximum EAR values of 0.138 and 0.521. The eyes will be considered as close if the eyelids covered more than 80% or less than 0.2146 of the EAR.

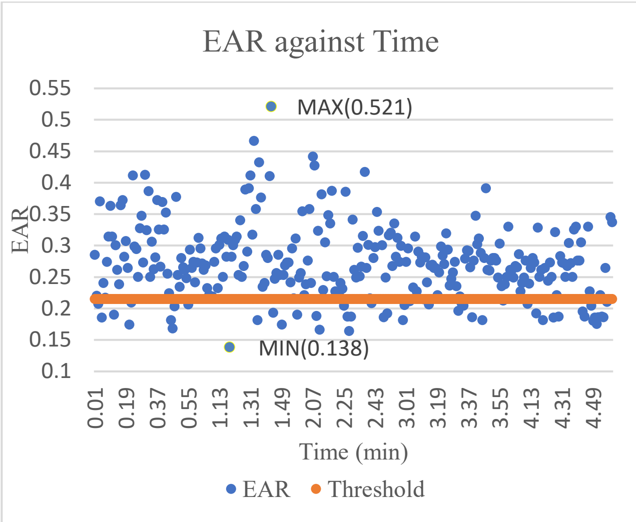


Fig 12. EAR for Subject #3

Overall eye blink rate for four subjects are not exceeding 20 blinks per minute. It is as mention in [9] and [10] that the normal eye blink rate for an average person is not more than 20 blinks per minute. There are differences in the blink rate between subject #3 and subject #1, subject #2 as well as subject #4. The lower blink rate of subject #1, subject #2, and subject #4 due to focus state while driving as referred to [9] and [10]. Subject #3 may have a higher blinking rate because of the cold environment due to the air conditioner in the car. The rise of the blink rate can increase eye temperature [13].

4 CONCLUSIONS

The Eye-Closeness system is a functional system to use for drowsiness detection applications. The features applied in this system is dependable to identify the drowsy condition. This research has proven that most accuracy of face detection is above 80% and the highest reading achieved is 98.33%. The error occurred only as a result of the complex background and unsteady lighting while performing the test. Seven over eight-persons have a broad range of measurements between eye width and intercanthal distances due to the low resolution of the hardware used. Throughout the whole experiment, the algorithm implemented is functional to analyse the eye-closeness movement. System performance is reliable because all four subjects tested for normal eye blink rate while driving is within the normal blink rate that is below 20 blinks/min as indicated in the previous research by Tereza Soukupov´ [9]. The safety of the driver and other road users should be concerned. Hence, the use of this system can prevent accidents caused by drowsiness.

REFERENCES

[1] E. Vural, M. Cetin, A. Ereil, G. Littlewort, M. Bartlett, and J. Movellan, "Drowsy driver detection through facial movement analysis," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4796 LNCS, no. November, pp. 6–18, 2007.

[2] P. R. Tabrizi and R. A. Zoroofi, "Drowsiness detection based on brightness and numeral features of eye image," *IIH-MSP 2009 - 2009 5th Int. Conf. Intell. Inf. Hiding Multimed. Signal Process.*, pp. 1310–1313, 2009.

[3] D. Sommer and D. Edwards, "PERCLOS : An Alertness Measure of the Past," no. October, 2017.

[4] G. Marquart, C. Cabrall, and J. de Winter, "Review of Eye-related Measures of Drivers' Mental Workload," *Procedia Manuf.*, vol. 3, no. Ahfe, pp. 2854–2861, 2015.

[5] S. Vitabile, A. De Paola, and F. Sorbello, "Bright pupil detection in an embedded, real-time drowsiness monitoring system," *Proc. - Int. Conf. Adv. Inf. Netw. Appl. AINA*, pp. 661–668, 2010.

[6] A. D. McDonald, J. D. Lee, C. Schwarz, and T. L. Brown, "A contextual and temporal algorithm for driver drowsiness detection," *Accid. Anal. Prev.*, vol. 113, no. January, pp. 25–37, 2018.

[7] J. van L. Gerhard Goos, Juris Hartmanis, *Audio- andVideo-Based Biometric Person Authentication*. 2001.

[8] J. J. Yan, H. H. Kuo, Y. F. Lin, and T. L. Liao, "Real-time driver drowsiness detection system based on PERCLOS and grayscale image processing," *Proc. - 2016 IEEE Int. Symp. Comput. Consum. Control. IS3C 2016*, pp. 243–246, 2016.

[9] M. Perception and C. Technical, "Eye-Blink Detection Using Facial Landmarks," 2016.

[10] Z. A. Haq, "Eye-Blink rate detection for fatigue determination," no. March, 2018.

[11] V. Singh, V. Shokeen, and B. Singh, "Face Detection By Haar Cascade Classifier With Simple And Complex Backgrounds Images Using Opencv Implementation," *Int. J. Adv. Technol. Eng. Sci.*, vol. 1, no. 12, pp. 33–38, 2013.

[12] M. Rezaei and R. Klette, *Driver Drowsiness Detection*. 2017.

[13] D. B. Gurung, K. C. Gokul, and P. R. Adhikary, "Mathematical model of thermal effects of blinking in human eye," *Int. J. Biomath.*, vol. 9, no. 1, 2016.

Towards Establishing Path Planning Strategies For Autonomous UAVs; A Brief Survey-Summary on Recent Techniques

Anees ul Husnain[1 2], Norrima Binti Mokhtar[1], Noraisyah Binti Mohamed Shah[1], Mahidzal Bin Dahari[1]

[1] Department of Electrical Engineering, University of Malaya, Kuala Lumpur, Malaysia
[2] Applied Controls and Robotics Research Laboratory, University of Malaya, Kuala Lumpur, Malaysia
(Tel: 60-10-245-5060)
norrimamokhtar@um.edu.my

Abstract
The extent of autonomy in path planning for a UAV primarily depends upon the capabilities of its algorithm. The diversity in UAV applications and an abundance of choices in autonomous path planning algorithms are swelling every day, so the selection of most appropriate algorithm gets baffling. The past two decades of research on UAVs revealed that seventy percent of it had been published in the previous three and a half years. Hence, a comprehensive survey study was proposed and conducted to obtain an overview of the recent developments in autonomous path planning applications and their respective algorithms. This article presents a summary of the survey and suggests most suitable path planning algorithms for a UAV application.

Keywords: Autonomous UAV, Survey, UAV Path Planning.

1 Introduction

UAVs can provide highly useful data, and their role becomes vital in information collection, aerial surveillance, industrial or agricultural monitoring, preventing disasters and so on. However, the most significant element that is common in most of these applications to pick out would be the access to real-time data. The role of UAV systems, due to their cost effectiveness and eased of access to deploy has influentially motivated scientists to opt them as a tremendous support to research Nevertheless, this would require a higher level of autonomous behavior among UAVs.

An autonomous flight correctly points towards an onboard intelligent sensors-based system that is expected to set and modify the intermediate flight paths, to meet the mission objectives. An Unmanned Aerial Vehicle (UAV) may either follow a pre-programmed flight trajectory, a remotely operated path by a human or an dynamic path calculation during the flight.[1] At all the times in flight, a UAV must maintain or observe the parameters that mainly include stable flight dynamics, energy consumptions, mission requirements or to restrain themselves within the areas of interest. This gives a very significant role to the aspect of path planning in UAV operations, as a global objective, alongside the local flight requirements.

Usually, a UAV has a global and local set of dynamics and parameters alongside a high- and low-level control strategies, to be generic. Path planning is a matter of higher-level UAV control strategy where the UAV considers global parameters to reach it optimality conditions.

From a broader perspective, path planning can be categorized into two groups, namely offline and online path planning. The path planning before the take-off is offline while modification of flight path during the flight, as required by the mission objectives or the environmental dynamics, is online path planning.

The extent of maximum autonomy that a UAV can achieve is an arbitrarily unachievable as it primarily depends upon degree of autonomous operations that a UAV may conduct or a group, swarm and flock of UAV may conduct. It can also be inferred with respect to current developments in this technology that the maximum autonomy of a UAV is a goal still unclear. However, a certainty in this regard so far is that a higher level of autonomy would inevitably require exploiting the onboard computational capabilities that demand computationally intelligent algorithms. Therefore, numerous approaches, criteria and innovations have been observed in effort to enhance the performance of autonomous flights. The graph depicted in fig.1. shows the published literature, indexed by the web of science, for the keyword "UAV" in the past 20 years. It shows that approximately 70% of the articles, published from 2016 to 2019.[2]

Another important factor to mention is the supporting role of UAVs in multidisciplinary research. This has contributed momentously in the spans of Geology, Chemistry, Environmental Sciences, Agriculture, Material Sciences, Physical Geography, Water Resources, Transportations, Mechanics, and many other research areas, elaborated in fig.2.[7]. Optimistically, this has brought scientists and engineers much closer to several aspects of their research.

2 Methodology

The selection of articles was based on two separate queries, *i.e.,* autonomous path planning for UAVs and autonomous UAVs swarm. The search was conducted on the *Web of Science Core Collection* database for the past five years. From the results, potential groups and subgroups were established. The research undertaken to improve the performance of existing algorithms or introducing innovative ideas for computing algorithms were grouped as developments in autonomous path planning algorithms. The rest of the part was considered as developments in UAV applications.

Algorithms were assigned labels relating to the category of their knowledge streams, as an instance, bio-inspired algorithms, graph search, *etc.* The algorithms in the second group were further labelled concerning the target application.

3 Needfulness

It was required to find the most suited algorithm for a specific target application and to identify a potential set of UAV applications for a particular algorithm. After discovering the top-notch algorithms and applications, an application to algorithms mapping was presented to complete the picture.

Some excellent surveys on path planning algorithms in recent years. In these surveys, the focus was relatively concise in most of them. A few studies about planning algorithms considered workspace dimensions,[3] and obstacle avoidance,[4] where the primarily concern was the nature of workspace and amount of information available to the robot. Another to mention is a technique-specific path computation,[5] moving obstacles cluttered environments.[6]. These dealt with the unexpected changes within the workspace and primarily deal with the lower level control of the UAV.

The considerations for this study were more towards practicality and direct in approach; first, the most recent developments in ongoing autonomous path planning applications and second, the algorithms associated with the most focused UAV applications.

This helped to develop an application to algorithms mapping out of the recent published literature. that offers the reader with appropriate choices of algorithms for a specific UAV application.

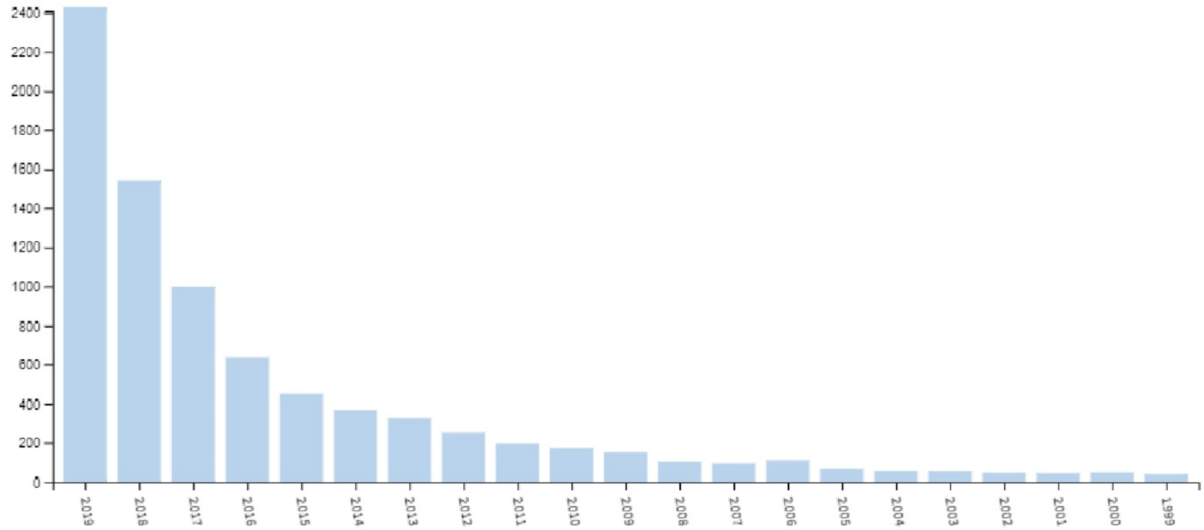


Fig. 1. Growth of research on UAVs in the past two decades (X: years, Y: publications)[2]

Fig. 2. Contribution of UAVs in Multi-Disciplinary Research Areas (2000-2020)[7]

4 Survey Findings

The conducted survey was categorized under two sections; one, most widely utilized algorithms and two, most significant UAV applications. The relation established between these two categories was found helpful in identifying the usefulness of algorithms and promising UAV applications. Hence, identifying the focus of UAV's path planning research and its directions. At here, only the survey findings from the algorithmic developments and most focused applications have been presented, respectively below.

It was found that the fusion of two or more algorithms was 25% of all the algorithms, and Evolutionary algorithms were widely chosen for the optimization process. However, the utilization of bio-inspired algorithms and modifications, if combined, these comprised 37% of the whole pie chart, depicted in Fig. 3.

From the survey, it was also observed that Cooperative Mission planning among UAVs has been the hottest and most focused research area in Autonomous Path Planning. Their reasons mainly include cost-effective, precision to conduct lower atmospheric aerial operations, ease of deployment, mobility, and accessibility of the required equipment. The second most attractive application was Sense & Avoid, followed by extended flight duration and coverage area, applications. New challenges have emerged in UAV research for communication infrastructure among UAVs, a unified identification of foes and friendly flights, secure communications, and UAV traffic regulatory systems.

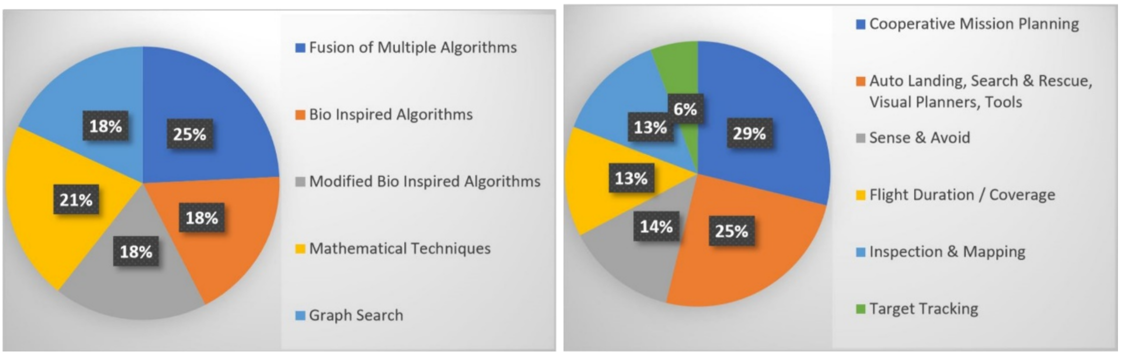


Fig.3. Breakdown of autonomous UAV path planning algorithms (left).and applications (right)

5 Reflections

Very high diversity in path planning applications and available algorithms engendered skepticism about the selection of a better or most suited algorithm for a specific UAV application. Therefore, autonomous path planning applications and the algorithms utilized for these have been mapped in Fig. 4. Some of the algorithms can be observed for serving different categories of applications, while others are dedicatedly developed to solve a particular path planning problem. It can also be inferred as the figure illustrates the most focused and demanded features of a UAV that provides a hint of futuristic models. The future of UAVs and developments in computational intelligence would look forward to the inclusion and selection of different data-process models to attain the highest possible intelligence capabilities, under the lights of techniques like reinforcements learning.

However, there have been some worth mentioning observations:

1. *It is quite disturbing to notice an absence of Machine Learning techniques*
2. *there can be observed a huge dominance of optimization algorithms.*
3. *The highest level of autonomy has not reached, and it stands as a goal not clear yet.*
4. *Any machine that is supposed to be intelligent must contain a set of specific smart or intelligent actions. The intelligent actions come from intelligent algorithms.*
5. *A criteria that can be foreseen at this point in time can be as follows: The selection of autonomous UAV's behavior from UAV itself, based on the randomness of its environment and mission objectives, ought to be rated as the highest degree of autonomy and hence may serve as the ultimate test of UAV's autonomy.*

Fig. 4. Applications to Algorithms Mapping for Autonomous Path Planning of UAVs

Acknowledgments:

The study was carried out as part of the research on "Autonomous Path Planning of UAVs for Environmental Monitoring" at the University of Malaya, Kuala Lumpur, Malaysia. The work is supported by University of Malaya, grant number: GPF062A-2018 for experimental equipment procurement. The Ph.D. scholar is sponsored by a funded scholarship from the Higher Education Commission (HEC) of Pakistan. The author would also like to acknowledge the support from the Islamia University of Bahawalpur, Pakistan.

References:

1 Cabreira, T.M., L.B. Brisolara, and P.R.F. Jr., Survey on Coverage Path Planning with Unmanned Aerial Vehicles. Drones, 2019. 3(1).

2 Analyzer, C.A.-S.R., 'UAV' search results analysis: Publication Years (1999-2019), in Clarivate Analytics - Web of Science, U.s.r.a.P.Y. (1999-2019), Editor. 2020, Clarivate Analytics: Online.

3 Yang, L., et al., A literature review of UAV 3D path planning, in Proceeding of the 11th World Congress on Intelligent Control and Automation. 2015, IEEE: Shenyang, China.

4 Mohammadreza Radmanesh, et al., Overview of Path-Planning and Obstacle Avoidance Algorithms for UAVs: A Comparative Study. Unmanned Systems, 2018. 6(2): p. 13.

5 Noreen, I., A. Khan, and Z. Habib, Optimal Path Planning using RRT* based Approaches: A Survey and Future Directions. International Journal of Advanced Computer Science and Applications(IJACSA), 2016. 7: p. 11.

6 Jun, M. and R. D'Andrea, Path Planning for Unmanned Aerial Vehicles in Uncertain and Adversarial Environments, in Cooperative Control:

Models, Applications and Algorithms, S. Butenko, R. Murphey, and P.M. Pardalos, Editors., Springer, Boston, MA: United States. p. 95-111.
7 Analyzer, C.A.-S.R., 'UAV' search results analysis: Research Areas Categories (1999-2019), in Clarivate Analytics - Web of Science, U.s.r.a.R.A.C. (1999-2019), Editor. 2020, Clarivate Analytics.

Development of Interactive Robot
-Emotion Estimation System Using Speech by 1dCNN-

Yugo Kawachi
Department of Mechanical Information Science and Technology, Kyushu Institute of Technology
680-4, Kawazu, Iizuka-City, Fukuoka, 820-8502, Japan

Eiji Hayashi
Department of Mechanical Information Science and Technology, Kyushu Institute of Technology
680-4, Kawazu, Iizuka-City, Fukuoka, 820-8502, Japan

E-mail: kawachi.yugo846@mail.kyuutech.jp, haya@mse.kyutech.ac.jp
http://www.kyutech.ac.jp/

Abstract

In order for robots to interact smoothly with people, they need to recognize human emotions and express its own emotions through both verbal and non-verbal communication. In non-verbal communication, we naturally do things such as estimating emotions from the tone of the other person's voice while talking. In this study, we developed a system to estimate emotions from features of speech rather than the speaker's words by comparing two different data sets, and compared what features each data set has and how they differ for each subject.

Keywords: personal robot, emotion estimation, nonverbal communication, 1dCNN

1. Introduction

With the expansion of the robotics industry market, the development of service robots is becoming more and more popular. These robots are intended to be used in the home, medical care, welfare, and other places where people communicate with each other, and it is necessary for them to behave and talk in a friendly manner. In this research, we are developing an interactive robot that pursues human-like movements by focusing on non-verbal interactions (cooperative behavior) such as facial expressions and body language.
In conversations, we can infer the other person's emotions from the inflection of their speech. We thought it was necessary to have a function to estimate emotions from speech. Therefore, in this study, we developed an emotion estimation system for speech using a machine learning method called 1 dimensional convolutional neural network(1dCNN).

2. Emotion Estimation System for Speech Using 1dCNN

In this system, we used a model called emotion-classification-Ravdess[1]. In addition, user status was defined 'Positive' and 'Negative' in line with previous studies.

2.1. *Structure of 1dCNN*

The structure of the 1dCNN for emotion estimation is shown in Fig. 1. In Fig. 1, A is Convolution layer and B is Pooling layer. The parameters of the Convolution and Pooling layers are shown in Table. 1, and the ReLU and softmax functions are used as activation functions.

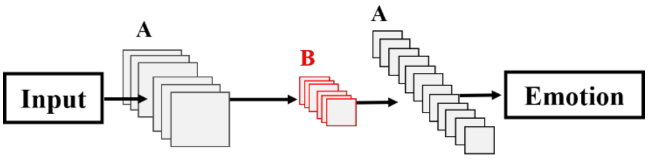


Fig. 1 Structure of network

	1dCNN	MAX_pooling	1dCNN
Size of filter	5	8	5
stride	1	1	1
Number of filters	128	128	128

Table. 1 parameters of each layers

2.2. *Feature Values*

The Mel Frequency Spectrum and Mel Frequency Cepstrum Coefficient (MFCC) were extracted as features from speech, and the features were used as input data for training. The MFCC is a discrete cosine transformed coefficient of the Mel Frequency Spectrum. The discrete cosine transform is said to improve the performance of the feature.

2.3. *Leaning data*

For training, we used the Ravdess[2] dataset and a newly created dataset.

Ravdess was recorded two types of sentences, "Kids are talking by the door" and "Dogs are sitting by the door," read by 24 professional actors (12 men and 12 women) in eight different emotions. The total number of data was 1440. The eight emotions were Classified “neutral, calm, happiness, sad, anger, fearful, disgust, surprised”. In this research, neutral and calm were removed, joy and surprise classified as positive, and sadness, anger, fear, and disgust classified as negative. The new dataset was created using the author's voice, and as in Ravdess. It was recorded by emotion and classified into positive and negative. The new dataset contains English greetings of two words or less, not sentences. The total number of data is 120, 60 for positive and 60 for negative.

2.4. *Accuracy Evaluation*

Table. 2 shows the classification results by each model of the results of using 1dCNN with MFCC and Mel frequency spectrum as input, and it used author's voice.

Dataset	Ravdess		Newly Created Dataset	
Feature Values	MFCC	Spectrum	MFCC	Spectrum
Positive	0%	0%	92%	89%
Negative	67%	67%	93%	90%

Table. 2 Classification Accuracies for each feature

First of all, focusing on the classification accuracy of each dataset, the classifier using Ravdess was classified as negative in both MFCC and Mel frequency spectrum. On the other hand, the evaluation using the newly created data set showed that the classifier was able to classify more unbiasedly than the classifier using Ravdess. The reason is conceivable that Ravdess dataset is a sentence. When words are given as input, it is necessary to focus on the inflection of the words for classification. However, in the case of sentences, the recognition rate drops due to the influence of syllables in between sentences.

Next, classification accuracy for each feature value in the newly created dataset, we can see that MFCC has a higher accuracy of about 3%. However, since there was only one difference in the number of data that failed to be classified, it was concluded that there was no difference in the classification accuracy between the two features.

Based on the above results, the classifier trained on the newly created data set using MFCC and Mel frequency spectrum features respectively was adopted as the emotion estimation system.

3. Verification

Test data containing the voices of four people were input to the emotion estimation system for classification. Fig. 2 shows the results of classification using MFCC and Mel Frequency Spectrum. A and C are the results of the system's classification of the positive emotions, and B and D are the results of the system's classification of the negative emotions. Table. 3 shows the classification accuracies of each class obtained from the classification results of this experiment.

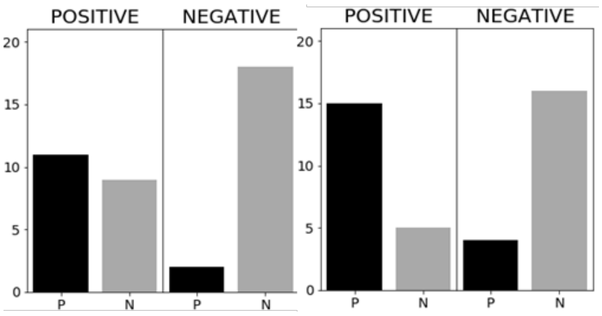


Fig. 2 Result of classification for each feature

Table. 3 Classification accuracy of emotion for each feature

	MFCC	spectram
Positive	**66.7%**	**76.9%**
Negative	**76.6%**	**78.0%**

From the results of this study, it was found that the classification accuracy of the model using Mel Frequency Spectrum was on average 5.9% higher than the model using MFCC for feature extraction. In addition, the model using MFCC showed a bias towards the negative, however it was not the case when the Mel Frequency Spectrum was used. The shape of each feature was output as a graph for comparison.
Fig. 3 shows the MFCC features of subject A's speech, and Fig. 4 shows the same speech as in Fig. 3 with the Mel frequency spectrum. From these figures, It can be seen that in MFCC, there is not much difference in shape between positive and negative and it is difficult to judge, whereas in Mel frequency spectrum, there was a clear difference in shape.

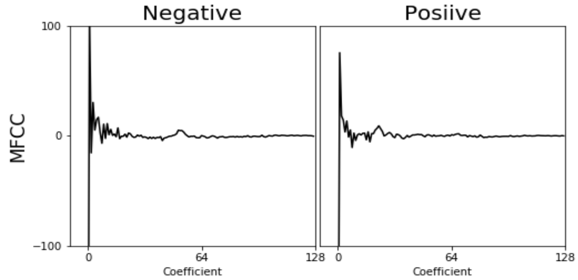


Fig. 3 Shape of MFCC

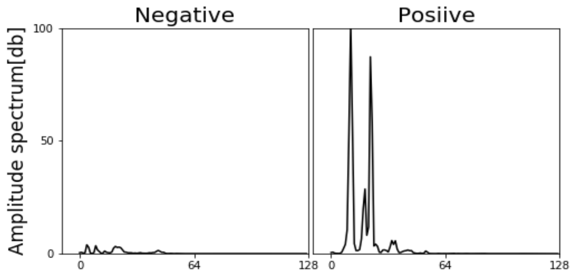


Fig. 4 Shape of Mel Frequency Spectrum

Table. 4 shows the emotion classification accuracy for each subject. First, focusing on subject C, the accuracy is lower than the others in the classification using the Mel frequency spectrum. This is thought to be because the distance between the subject and the microphone is close and the volume of the sound is larger. This suggests that in the classification using the Mel frequency spectrum, the factor of the loudness of the sound is related to the classification.

Feature Value	emotion	A	B	C	D
MFCC	Positive	75.0%	75.0%	88.9%	50.0%
	Negative	83.3%	83.3%	90.9%	66.7%
Mel Frequency Spectrum	Positive	57.1%	88.9%	71.4%	88.9%
	Negative	76.9%	90.9%	33.3%	90.9%

Table. 4 Accuracy of classification for each subject

4. Conclusion

In this study, we created and evaluated a classification system for emotions using 1dCNN. As a result, we were able to develop a system to classify users' emotions with an accuracy of 76.9%. However, we are able to identify emotions even when we use speakers, which means that a system in which the loudness and height of the voice affect the classification is not appropriate. This means that a system in which the loudness and height of the voice affect the classification is not appropriate. As a future prospect, it is necessary to search for an emotion estimation system that does not depend on the volume and height of the voice, and also to construct an emotion estimation system that combines facial expression and body expression recognition.

5. References

1. **Marco Giuseppe de Pinto,**https://github.com/marcogdepinto/Emotion-Classification-Ravdess(2019)

2. Livingstone, Steven R., Katlyn Peck, and Frank A. Russo. "Ravdess: The ryerson audio-visual database of emotional speech and song." *Annual meeting of the canadian society for brain, behaviour and cognitive science*. 2012.

Deep Learning Methods for Semantic Segmentation of Dense 3D SLAM Maps

Pei **Yingjian**
MIST, Kyushu Institute of Technology, 680-4 Kawazu
Iizuka-shi, Fukuoka 820-8502, Japan

Sakmongkon Chumkamon
MIST, Kyushu Institute of Technology, 680-4 Kawazu
Iizuka-shi, Fukuoka 820-8502, Japan

Eiji Hayashi
MIST, Hayashi Lab, 680-4 Kawazu
Iizuka-shi, Fukuoka 820-8502, Japan
E-mail: yingjian.pei801@mail.kyutech.jp, m-san@mmcs.mse.kyutech.ac.jp, haya@mse.kyutech.ac.jp
www.kyutech.ac.jp

Abstract

Most real-time SLAM systems can only achieve semi-dense mapping, and the robot lacks specific knowledge of the mapping results, so it can only achieve simple positioning and obstacle avoidance, which may be used as an obstacle in the face of the target object to be grasped, thus affecting the realization of motion planning. The use of semantic segmentation in dense SLAM maps allows the robot to better understand the map information, distinguish the meaning of different blocks in the map by semantic labels, and achieve fast feature matching and Loop Closure Detection based on the relationship between semantic labels in the scene. There are many semantic segmentation datasets based on street scenes and indoor scenes available for use, and these datasets have some common tags. Based on these training data, we can derive a semantic segmentation model based on RGB images by using the Pytorch platform for training.

Keywords: 3D SLAM, Semantic Segmentation, Point Cloud, ROS

1.Introduction

Active obstacle avoidance motion planning is an important element in the autonomous motion planning of the robot arm. Active obstacle avoidance can not only avoid collision damage during the operation of the robot, but also improve the robot's ability to sense the environment and avoid causing safety accidents. In active obstacle avoidance motion planning, one of the conditions is the level of the robot's perception of the workspace. Effective obstacle avoidance motion planning can only be achieved if the robot has a prior awareness of the obstacles in the workspace and is able to update environmental information at any time during the work. The advantage of 3D SLAM is that the robot can get the complete spatial information of the current environment, and get the abstract modeling of the real environment in the virtual environment through the octomap, and through the Rviz plugin of ROS platform, the obstacle avoidance motion planning can be realized with the octomap as the reference.

For semantic segmentation of dense maps, we can either use a direct segmentation method on 3D point cloud data or a semantic segmentation method based on 2D RGB images, and we use the second method due to the convenience of training data. We use the second method for the convenience of training data. There are many semantic segmentation datasets based on street scenes and indoor scenes available for use, and these datasets have some common tags. Based on these training data, we can derive a semantic segmentation model based on RGB images by using the PyTorch platform for training.

There has been some progress in scene marking research based on this approach, and our project has now applied some of this technology to detect target types in 3D point cloud data. Combining this progress with obstacle avoidance motion planning, the robot will be able to accurately distinguish between grasping and obstacle avoidance based on an understanding of the meaning of the scene, and it will be easier to integrate the two systems into the same framework.

2.Segmentation Model Training

We refer to the work of Xuan Zhang et al. in selecting the semantic segmentation model and improve it to some extent based on our system. Our study was mainly trained using the ade20k dataset, a training set that includes a large number of images and labels of common household items and can be applied to most scenarios. For the laboratory scenario we will use, we used annotation software to additionally annotate a portion of the images and add them to the dataset for training. Figure.1 shows Example image of our own data.

In terms of neural network selection, we chose PSPNet as the basis, which is very suitable for semantic segmentation work such as scene parsing, and demonstrated a certain degree of accuracy when tested in real scenes using pre-trained models. We used the PyTorch platform to initially train the dataset with PSPNet50. Due to the limited GPU computing resources and the complex environment during the initial training, the semantic segmentation effect was not perfect, but a clear segmentation boundary could be demonstrated under sufficient lighting conditions.

2.1. *Segmentation Models Comparison*

In choosing the model, we mainly examined the training accuracy of different neural network models on the ade20k dataset, and the indoor part of the ade20k dataset was chosen for testing, mainly because our project is based on indoor scenes, and the part of the dataset chosen cannot accurately evaluate the accuracy of some labels, but it can greatly accelerate the training.

Figure.1 represents the real-time display of 2D semantic segmentation under different models, from top to bottom, the original RGB image, PSPnet_50, PSPnet_50 with Bayesian filter on, and the unmodified ResNet_50 image. From the comparison graphs, it can be seen that Bayesian filter does not contribute much to the semantic segmentation accuracy of indoor scenes, while PSPnet, as an improved network based on ResNet, reflects better results on ade20k dataset. It can be seen from the figure that PSPnet accurately identifies the screen that appear only partially in the scene (green part), while the original ResNet is more ambiguous.

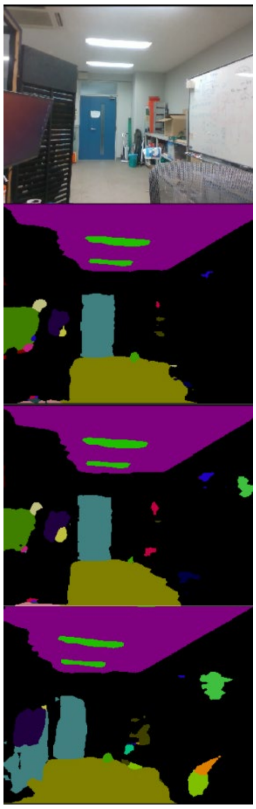


Figure. 1 Comparison between different models

Table.1 Comparison of segmentation speed (FPS)

Model	FPS (Average)
PSPnet_50	20
PSPnet_50 (Bayesian filter)	15
BasicResNet_50	23

2.2. *Segmentation Speed Comparison*

Since SLAM pursues real-time, accuracy and robustness, the frame rate should be used as the reference value for speed determination. The comparisons in Table.1 are all made under Ubuntu 18.04, ROS Melodic environment using CPU calculations, and the relevant point cloud generation module is not activated to save system resources during the frame rate comparison.

2.3. *Conclusion of Model Selection*

It can be seen from the table that turning on the Bayesian filter will have a significant impact on the frame rate, and the original ResNet is better than PSPnet in terms of fluency. Taking all factors into consideration, the PSPnet semantic segmentation model without bayesian filter has achieved a good balance between accuracy and speed, and we will use this scheme in the actual test.

3. Overall System Construction

In this section, we will discuss the setting up of the experimental environment and camera selection. The first section will give a general overview of the experimental environment and the experimental format, and the second section will present our considerations when selecting a depth camera.

3.1. *Environment Construction*

In order to allow flexible observation of the workspace by depth cameras, we placed the depth camera for SLAM at the manipulator of the robot arm, and to compensate for blind spots in the observation, we also placed a plurality of depth cameras at other locations in the workspace. In the following sections, we will mainly demonstrate and illustrate the main camera fixed on the manipulator.

After our depth camera acquires the semantic segmentation image, it publishes and projects the 2D image to the point cloud, and generates Octomap using the point cloud image.

Figure.2 shows how the system looks like in a rviz UI in ROS, you can see Semantic Image on the corner and a octomap view in the main window.

Figure.3 shows the overall view of system frame.

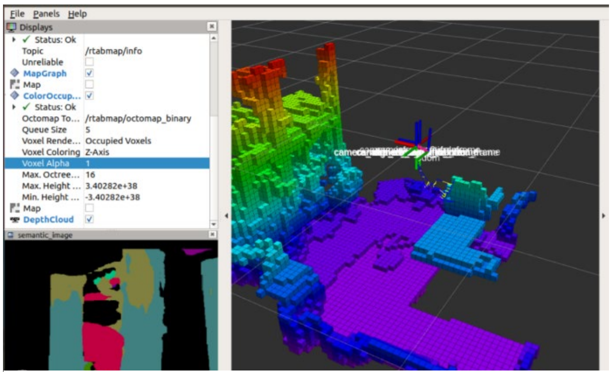


Figure. 2 Rviz scene demo

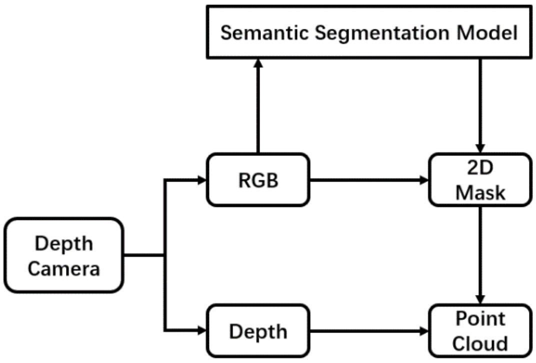


Figure. 3 Conceptual Frame

3.2. Camera Selection

In terms of camera selection, Intel Realsense D400 series was chosen for this study. The D435i with IMU module and the D435 without IMU module were used in practice and they both performed consistently in the tests. The reason for using D435 series is that when building 3D SLAM maps, it is necessary to provide accurate depth data and clearer images for feature recognition and reference, and the overall size of the camera has to be taken into account as well, so the camera should not be too heavy or too big to be mounted on the operation end of the robot arm. The D435i, which comes with an IMU module, is useful for 3D SLAM mapping and offers more possibilities for optimization in subsequent development.

4. Mapping

The platform we use is ROS Melodic based on Ubuntu 18.04, and the graphical interface is ROS Rviz, which is also used to facilitate the subsequent obstacle avoidance motion planning.

Due to the relatively small workspace used in this study, some of the parameters of the official launch file are not applicable to the realities of this study, so we have adjusted some of the parameters.

Through the actual test, the overall map building resolution is relatively high and the map building is nearly perfect. The semantic octomap built can be successfully saved through the relevant functions of octomap_server, and can be published directly in octomap_server with .bt format into Marker Array Topic under the condition that the actual working scene of the robot remains unchanged.

After completing the scene mapping, the obstacle avoidance motion planning is performed using the RRTconnect algorithm in MoveIt! and output directly to the Motoman robot arm. In the Gazebo simulation environment, the obstacle avoidance planning effect on octomap is good.

In practice, due to the problem of hand-eye calibration and coordinate conversion, we can use static coordinate conversion to fix the coordinates of the camera and robot model to realize the map building and obstacle avoidance function in MoveIt! It is also possible to save and publish octomap at the same time to realize real-time map building and obstacle avoidance planning in different Rviz terminals.

Since the actual runtime is based on the Python 2 platform, and the latest version of PyTorch no longer supports Python 2, we use the older version 0.4.0, which is only supported by the CPU, and therefore runs with a certain degree of lag and has some shortcomings in terms of time efficiency.

5. Conclusion

After simulation and practical testing, the deep learning method can be integrated with our system to a certain extent, which is of great help in the subsequent development of intelligent control for Motoman robot. Semantic 3D SLAM's mapping results are in line with expectations, the obstacle avoidance function works normally, and the constructed map can be reused, which eliminates the need to model the scene in a virtual environment. This module will continue to be refined and developed as part of the project.

6. Discussion & Future Works

In practical tests, we also found the following problems:
1) Inefficient map building. When using the deep learning method described in this paper, because the feature point matching method not use the semantic data to match the point, and the rate at which the camera extracts keyframes from the video stream depends on the processing speed of the computer and the camera, in the case where the robot arm moves too fast, the camera will collect two pictures with almost no similar features, and thus cannot form a closed loop to build a map. In the actual test we have to use slower movement speed to get the complete map, which affects the overall map building efficiency.
2) Due to the existence of multiple coordinate nodes in Motoman's virtual model, multiple coordinate conversions are required when binding the camera coordinates, which will result in the camera coordinates being incorrectly bound to the world coordinate system or the overhang position. We use a compromise between opening two terminals to create the final effect, and using fixed coordinates is also a last resort.
3) We placed more than one depth camera in the workspace, but we only used one of them in practice, and we hope that in the future we will be able to integrate the depth data from all the cameras into a complete map.

Our next plan is to replace the modules with models trained by other deep learning methods to improve the efficiency of map building. We plan to use Siamese Neural Network-based semantic feature recognition matching algorithms and implement the object-level building, and we refer to the methods used in Ref. 4 and Ref. 5 for this work.

References

1. Labbe, Mathieu & Michaud, François. (2013). Appearance-Based Loop Closure Detection for Online Large-Scale and Long-Term Operation. Robotics, IEEE Transactions on. 29. 734-745. 10.1109/TRO.2013.2242375. (IEEE Xplore)
2. Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid Scene Parsing Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6230-6239.
3. Xuan, Zhang and David, Filliat, Real-time voxel based 3D semantic mapping with a hand held RGB-D camera, 2018, GitHub, GitHub repository, \url{https://github.com/floatlazer/semantic_slam
4. Hu, L., Xu, W., Huang, K., & Kneip, L. (2019). Deep-SLAM++: Object-level RGBD SLAM based on class-specific deep shape priors. ArXiv, abs/1907.09691.
5. Puligilla, Sai Shubodh, Satyajit Tourani, Tushar Vaidya, Udit Singh Parihar, Ravi Kiran Sarvadevabhatla and K. Krishna. “Topological Mapping for Manhattan-like Repetitive Environments.” 2020 IEEE International Conference on Robotics and Automation (ICRA) (2020): 6268-6274.

Robot Motion Generation by Hand Demonstration

Sakmongkon Chumkamon, Umaporn Yokkampon, Eiji Hayashi, Ryusuke Fujisawa
School of Computer Science and Systems Engineering, Kyushu Institute of Technology, Japan
E-mail: m-san@mmcs.mse.kyuetech.ac.jp, may@mmcs.mse.kyuetech.ac.jp, haya@mse.kyutech.ac.jp, fujisawa@ces.kyutech.ac.jp

Abstract

Since traditional robot teaching requires time and instruction to the robot motion, we present a systematic framework based on deep learning and experiment for generating robot motion trajectories from human hand demonstration. In this system, the worker could teach robot easier rather than assigning the instruction to the robot controller manually. Therefore, the robot can imitate the action in a new situation instead of directly teaching the robot arm. Our contributions include three points 1) the real-time extracting method of hand movement without marker using hand detection in 3D from human 2) the motion generalization of the hand trajectories from human 3) Robot path planning for grasping and place the object to the target. We also present the experiment conducted by the user movement for real data and evaluate the system using the manipulator robot. The investigation shows the pick-and-place task of the robot for food by hand demonstration.

Keywords: List four to six keywords which characterize the article.

1. Introduction

In order to push country economic, smart automation in the industrial sector is critical to improving nowadays. Since the merit of artificial intelligence research could facilitate various tasks in industrial. Therefore, we can develop a factory that could produce products faster with higher precision. In a smart factory, the robot is essential to assembly because it is more reliable and accurate than human.

The manipulator robot actually is controlled by joint position; however, we usually control the robot in Cartesian space; therefore, we have to calculate the inverse kinematic (IK). One of IK that is popular recently is Trac-IK and KDL, which can figure the solution more than traditional methods [1]. The robot also needs to plan the trajectory according to the environments such as avoiding the collision or moving to the target. The recent motion planning method is Open Motion Planning Library which is popular, reliable and fast for solving the trajectory motion in various movement such as robot arm motion, vehicle, drone[2].

In robot learning by demonstration, recent research develops the robot to predict and learning the motion from a human. However, the research develops from the device which is attached to the human body to extracting the information of motion[3]. Additionally, some research develops the teleoperation to the robot to learn how to move following the human motion[4].

In this paper, we propose and contribute the robot system that can generate the robotics motion from human hand demonstration and generalize the hand motion by Dynamic Movement Primitive (DMP)[5]. Then we apply the motion to the robot. Besides, this system uses Trac-IK for kinematic and OMPL for path planning after DMP process. Fig. 1 shows the overview of our proposed, which consists of hand demonstration with grasping object and perception using RGBD cameras with hand position estimation then robot motion using the manipulator and visualization using PCs.

For the rest of the paper, section 2 presents the framework of robotic motion generation approach by human demonstration. Section 3 explains the experiments as well as evaluation of the robotics motion. Finally, we

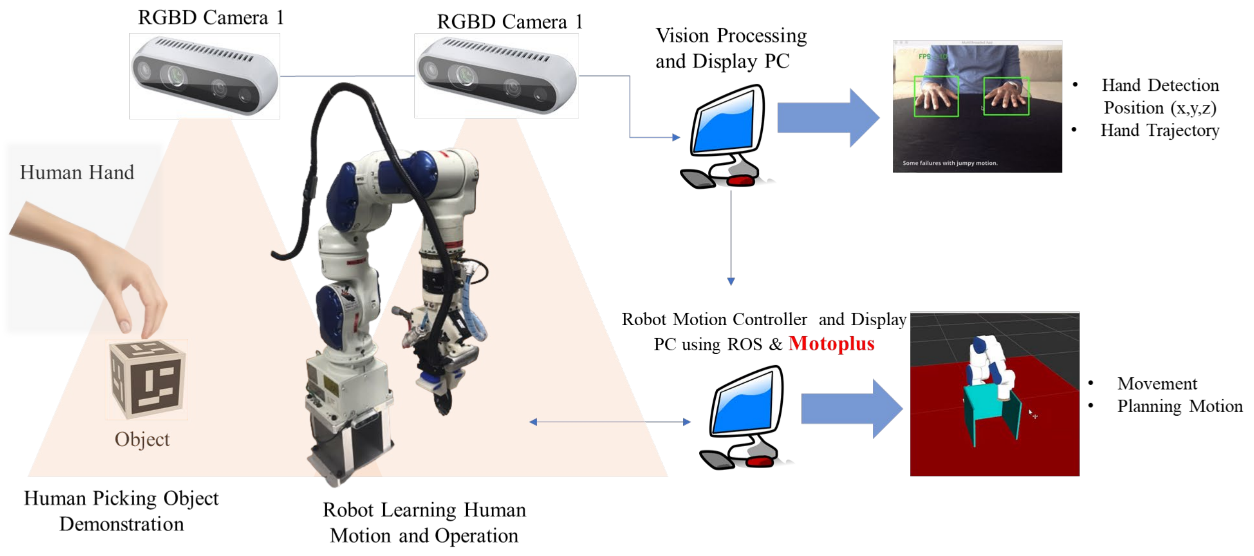


Fig. 1. The overview system of the Robot Motion Generation by Hand Demonstration.

present the conclusion, discussion and critical finding in this paper.

2. Approach

In this section, we explain the overview system including robot hardware configuration, hand recognition, hand position and trajectory determining. Finally, the motion repeating following the human hand demonstration.

2.1. *Hand Position Extraction*

Firstly, we implement the method to collect the data from the demonstration by record human hand motion. We divide the process into two steps. The First step is hand detection. We introduce the deep learning-based model which is Single Shot Detection (SSD) for real-time hand detection from a colour image[6]. SSD could outcome the detection box of hand area then we apply the human skin filter to getting only hand region. Later, we determine the hand position by calculating the centroid of the hand region. Therefore, we could get the hand position in 2-dimension. Since we use the RGBD camera, we apply both colour and depth image to map each other. After that, we map the 2D hand position into depth image to get the hand position in 3-dimension cartesian coordinate. Finally, we concatenate the point while the human moves to be the trajectory of hand motion and record to send to the robot.

2.2. *Robot Trajectory Generation and Motion*

Since the robot knew the hand motion trajectory, we can then plan the robot's motion according to the workspace constraint and the trajectory. Firstly, we set the home pose to initialize the robot joints to prepare a starting position to move following the trajectory. When the robot gets the new trajectory from demonstration, it will plan motion using the Open Motion Planning Library (OMPL) and Trac-IK inverse kinematic. In this implementation, we utilize the MoveIt tool with Robotics Operation System to organize motion planning and collision protection among the workspace environment. After complete recording the trajectory, then the robot moves following the human hand trajectory.

2.3. *Experiment Setup*

We use the industrial robot with 7 degrees of freedom and soft gripper tool of end-effector for the hardware configuration system since we aim our robot to grasp various food. For robot control system is implemented in the desktop PC with GPU Nvidia RTX 2080Ti. The PC also operates the SSD for hand detection, archiving the average hand detection 55 frame per second. In experiment setup, the robot is mounted on the robot base station with the camera mounted at the base of the robot. We then calibrate the camera position translation to the robot position and orientation.

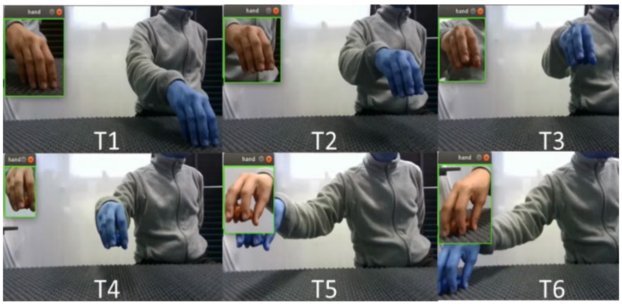


Fig. 2. The Trajectory Data collection procedure. The participant is working on the demonstration by moving his hand for trajectory collection.

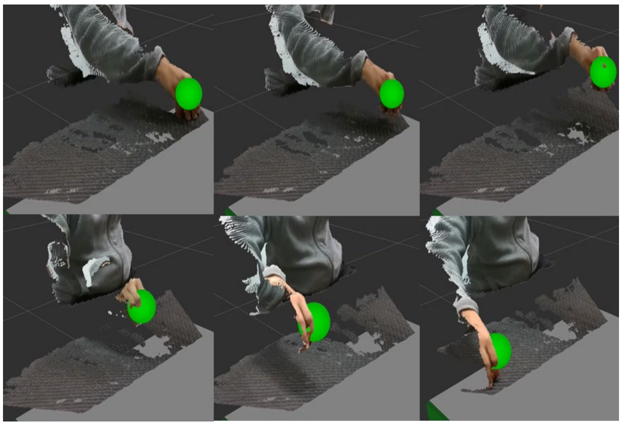


Fig. 3. Human Demonstration in 3-dimension point cloud with hand position detection represented by the green marker.

3. Experiment and Results

In our experiment, the user performs the demonstration, which is pick-and-place tasks for food. The robot could detect the hand position in cartesian coordinate then the robot reproduces the motion following the user's hand trajectory by robot end-effector. Firstly, the robot system performs hand detection, then extract the hand position in 3D cartesian coordinate and record the user's hand trajectory. After that, the robot performs the motion planning and move the end-effector following the trajectory.

3.1. *Human Demonstration by Moving Hand*

The participant makes the motion by his hand in front of the robot. The trajectory of the hand movement follows to the Fig. 2, which the participant moves from right side of the robot then move up. Additionally, Fig. 3 presents the marker of the hand represent in 3D cartesian coordinate. After that, the user moves the hand down a little from the centre of the trajectory. Finally, the user moves the hand down to the target to the left side of the robot. In hand position estimation, we determine the position using colour image then we map the colour image to depth image to get the cartesian coordinate of hand position. Finally, the trajectory is concatenated by the point to be the waypoint of hand movement. The example of the hand trajectory is shown in Fig 4, which is represented by the marker.

3.2. *Robot Repeating the Human Hand Motion*

Since the participant made the demonstration of hand motion using grasping the object, which is the chicken fry. Since our research would like to develop the robot to assist the worker in the food assembly industry, the participant picks and places the chicken fry from starting point to the target point to teach the robot move to the same motion and target as in Fig. 4 which shows the desired trajectory. After we finish the desired trajectory, the robot moves following the trajectory. In Fig 5, the robot starts from T1, which the robot moves to the starting position. At T2, the robot moves up to the left side of the robot. At T3, the robot moves the end-effector slightly down as the human demonstration. At T4, the robot moves up to the left side of robot to the target chicken fry. At T5, the robot is approaching to the target to the left side of it. Finally, at T6, the robot could move to the target as the human demonstration before at the same place that hand placed the chicken fry.

4. Conclusion

In this paper, to fill the gap of the related research, we present and implement the robot framework of robot motion generation by learning from human hand demonstration and generalize the motion applying to industrial manipulator. It aims toward natural teaching process relating to human learning from experience such as learning from human hand motion.

In this paper, we contribute the method of hand trajectory extraction using deep learning based which is single shot detection to detection hand position. The hand detection using deep neural networks could perform in real-time hand position estimation using RGBD camera to collect

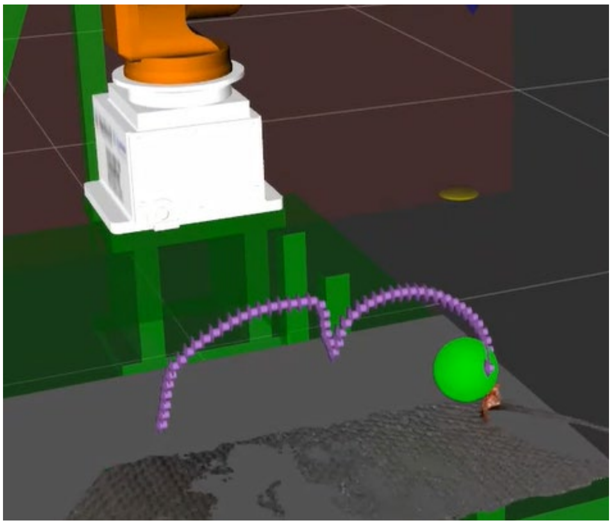


Fig. 4. The overview of the Robot Motion Generation by Hand Demonstration.

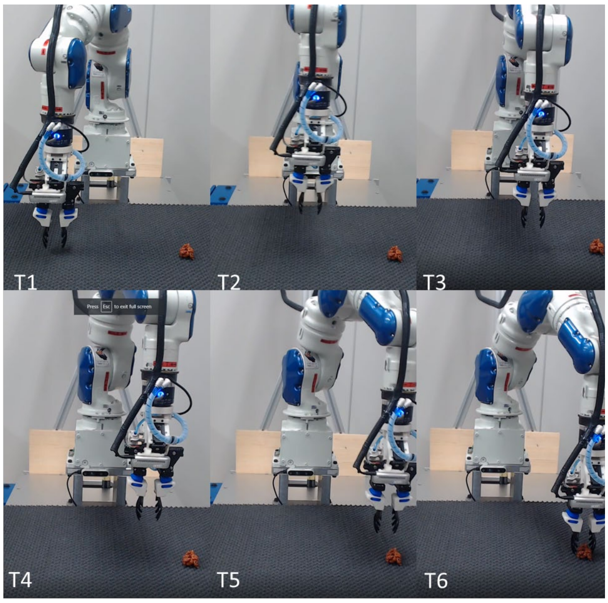


Fig. 5. The capture images of robot motion capture according to the human demonstration.

the trajectory path of hand from the frame by frame from human demonstrating. The second, we proposed the motion generalize motion from hand and generate the motion to the robot. Our approach can generate and generalized motion from hand demonstration without a marker, making the system more reliable and convenient from the real-world tasks. We also successfully experiment and evaluate the performance of the approach.

In future works, we would like to implement with constrain workspace and improve the pose estimation of hand for more reliable learning from humans.

Acknowledgements

This research was partially supported by the Religion revitalization project in Kitakyushu by the Cabinet Office of Japan.

References

1. Beeson, Patrick, and Barrett Ames. TRAC-IK: An open-source library for improved solving of generic inverse kinematics. In 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids), pp. 928-935. IEEE, 2015.
2. Sucan, I.A., Moll, M. and Kavraki, L.E., 2012. The open motion planning library. IEEE Robotics & Automation Magazine, 19(4), pp.72-82.
3. Wang, Weitian, Rui Li, Yi Chen, and Yunyi Jia. "Human intention prediction in human-robot collaborative tasks." In Companion of the 2018 ACM/IEEE international conference on human-robot interaction, pp. 279-280. 2018.
4. Jha, Abhishek, and Shital S. Chiddarwar. "Robot programming by demonstration using teleoperation through imitation." Industrial Robot: An International Journal (2017).
5. Schaal, Stefan, Jan Peters, Jun Nakanishi, and Auke Ijspeert. "Learning movement primitives." In Robotics research. the eleventh international symposium, pp. 561-572. Springer, Berlin, Heidelberg, 2005.
6. Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. "SSD: Single shot multibox detector." In European conference on computer vision, pp. 21-37. Springer, Cham, 2016

Medical Telerobotic: IRAPs SHaRE-aGIVeR

Noppadol Pudchuen[1], Jiraphan Inthiam[2], Wisanu Jitviriya[3]
Amornphun Phunopas[4], Chirdpong Deelertpaiboon[5]
Center of Innovative Robotics and Advanced Precision Systems: iRAPs,
Faculty of Engineering, King Mongkut's University of Technology North Bangkok,
1518 Pracharat 1 Road,Bangsue, Bangkok 10800, Thailand

Aran Blattler[6]
Department of Mechanical Information Science and Technology
Kyushu Institute of Technology,
680-4 Kawazu, Iizuka, Fukuoka 820-0067, Japan
E-mail: noppadol.p@eng.kmutnb.ac.th[1], jiraphan.i@eng.kmutnb.ac.th[2], wisanu.j@eng.kmutnb.ac.th[3],
amornphun.p@eng.kmutnb.ac.th[4], chirdpong.d@eng.kmutnb.ac.th[5], aran.blattler703@mail.kyutech.jp[6]
https://iraps.eng.kmutnb.ac.th

Abstract

The Coronavirus disease 2019 (COVID-19) pandemic has affected the global population. In particular, the medical personnel in direct contact with patients have been exposed to high risk. To reduce the spread of COVID-19 and protect health-care workers and patients, we would like to present the fully automated medical telerobot as the IRAPs SHaRE-aGIVeR robot. Our robot is capable of generating both 2D and 3D maps automatically, delivering medical supplies, food, or medical devices such as blood pressure monitors, pulse oximeters, and so on. In addition, the user interface system is also vital part. Users are able to connect and control the robot using a computer, a mobile device, or a tablet via the wireless network which is installed inside the robot. Currently, our robots are being operated on with medical personnel at hospitals, regarding the feedback from the formal caregivers that can prove our robot's efficiency in reducing the risk of COVID-19 spread.

Keywords: Medical Telerobotic, Perception, Planning, IRAPs SHaRE-aGIVeR.

1. Introduction

On 30th January 2020, the COVID-19 pandemic was declared as "Global Emergency" by The World Health Organization (WHO) because the virus had spread rapidly [1]. The COVID-19 pandemic has brought along several threats, including economic and social disruption. The group of people at high risk of catching COVID-19 are health-care workers, who spend much time up close with the patient. Therefore, the technologies involving medical robotics systems are mentioned to control and reduce infection spread to a large population [2].

Considering the current COVID-19 pandemic situation, robots are well suited for caring for the well-being of patients, sharing the workload of health-care workers in hospitals such as utilizing them for cleaning and food preparation jobs in infected hazardous areas. Robots are mainly classified with various applications in

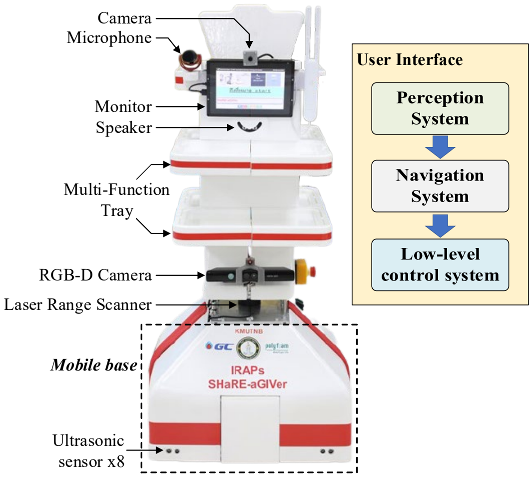


Fig.1 IRAPs SHaRE-aGIVeR robot

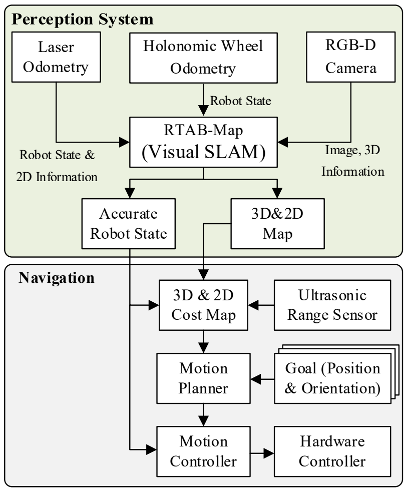


Fig.2 Perception Module and Navigation system

healthcare and related fields. **Receptionist robots** are used at a hospital's reception to give information about various sections of the hospital and guide visitors and patients [3]. **Nurse robots** are assigned to assist doctors in the hospital in the same manner as human nurses. These robots are commonly used in nursing and healthcare individuals undergo high stress and exhaustion due to patient load. For example, Robear: a robotics nurse to lift patients in Japan [4], Moxi: nursing robot placing medicines in bins [5], Paro robot and Pepper robot are used to assisting elderly patients in providing therapeutic assistance. **Serving robots** are used to deliver food and beverages, dispensing of drugs, or removing unclean laundry etc., inside the hospital [6],[7] and so on. For new alternative of the health-care robotics system, we would like to propose the fully automated medical telerobot as the IRAPs SHaRE-aGIVeR robot. Our robot is a holonomic movement platform, including the capability of generating both 2D and 3D maps automatically, path planning and navigation system, delivering medical supplies, food, or medical devices (e.g., blood pressure monitors, pulse oximeters) to patients.

Our paper is organized as follows: Chapter 2 explains the IRAPs SHaRE-aGIVeR robot overview system, which has three systems. In Chapter 3, our results are shown how our robot can operate with medical personnel at hospitals. In the final section, we conclude the paper and describe our future works.

2. Overview system of IRAPs SHaRE-aGIVeR

The robotic platform consists of the aluminum profile construction weighing over 35 kg and capable of a maximum speed of 0.8 m/s. The Mecanum wheeled drive is used with the robot's mobile base, which provides good properties and is popular in modern robotic designs. It can move sideways or rotate around its axis. The dimensions of the robot are 80 x 70 x 150 cm (length, width, height), it has a total weight of 60 kg and can carry an additional 80 kg. The appearance of IRAPs SHaRE-aGIVeR robot is shown in Fig.1.

2.1. *Perception System*

The first main core of our system is the perception system. The perception system provides a common state such as position, orientation velocity, and acceleration. Besides, it also provides the overall environment in the form of a 3D map. The system overview of our perception system is shown in Fig. 2. Each sub-system described as follows:

2.1.1. *Holonomic Wheel Odometry*

The first essential key for the perception-system is Holonomic wheel odometry. Therefore, we use the by-products of our driving system by employing the encoder wheel information was attached to the driving system with an algorithm in [8] to estimate the robot state such as position, orientation, and velocity. However, because of our drive system, which we use to be our odometry has four wheels. The odometry matrix that shows in [8] non-invertible due to the equation system is an overdetermined system because we have four odometry wheels, but we have only three state variables to solve due to our robot laying on the plane. In order to solve this problem, we permutate the configuration of the odometry wheel such as (1,2,3), (2,3,4), (3,4,1), (4,1,2). Therefore, the odometry matrix turns to be invertible and has a configuration as expressed in Eq. (1).

$$\begin{bmatrix} v_x \\ v_y \\ \omega \end{bmatrix}_{a,b,c} = \begin{bmatrix} \cos(\theta_a) & \sin(\theta_a) & L \\ \cos(\theta_b) & \sin(\theta_b) & L \\ \cos(\theta_c) & \sin(\theta_c) & L \end{bmatrix}^{-1} \begin{bmatrix} \dot{\phi}_a \\ \dot{\phi}_b \\ \dot{\phi}_c \end{bmatrix} \quad (1)$$

When the left side vector is a state of the robot. θ is the angle of velocity vector for each wheel reference to robot local coordinate. $\dot{\phi}$ is the angular velocity of odometry wheels, and (a, b, c) is odometry wheel configuration. However, the robot state vector received from the wheel odometry still has uncertainty due to it does not have an error correction system. So, we will employ the following algorithm to correct the accumulative error.

2.1.2. *Lidar Odometry and Visual Odometry*

As mentioned before, wheel odometry has uncertainty such as friction force and wheel slippage. In order to deal with the uncertainty of an environment, we employ the lidar odometry techniques because the core algorithm of lidar odometry relies on external features information such as plane and corner. The core concept for our lidar odometry includes the following steps:

- Point Cloud Filtering: The input point cloud is downsampled by voxel grid filter and statistical outlier removal filter [9] in order to reduce bandwidth, computational power and remove the outlier which comes from an environment such a highly reflective surface and the reflection of sunlight
- ICP Registration: In order to estimate the state vector of the robot, we employ iterative-closest-point (ICP) [10] implemented in libpointmatcher [11] by registering the new point cloud to the last keyframe, which called scan to scan method (S2S).

For now, we received the state vector of the robot from lidar odometry. Nonetheless, it still has uncertainty like wheel odometry. The lidar odometry cannot deal with a long corridor or large flat plane environment. Due to the system does not have enough features to track between two frames.

2.1.3. *Visual SLAM*

Our last perception system is the visual SLAM system. The two systems mentioned before still have problems such as accumulative error and cannot deal with a low geometry feature frame. Both problems can be solved by SLAM (Simultaneous Localization and Mapping) or, in other words, it is odometry with a close loop. At present, the state-of-the-art SLAM has both sensors: lidar sensor and camera. For our robot, we choose RTAB-Map [12] that is visual or camera SLAM. Because it relies on visual features, so it does not affect by low geometry feature frames like lidar odometry. Besides that, it is also combined with loop closure detection and non-linear optimization, which is used to deal with the accumulative error. Moreover, RTAB-Map supported combining the other types of odometry: wheel odometry and lidar odometry, to reduce or eliminate accumulative error and uncertainty around an environment. RTAB-Map has two working modes. The first one is the mapping mode. We used it to create a consistent map as the memory of the robot. Another one is localization mode, which fuses all information from wheel odometry, lidar odometry, and visual odometry to estimate the robot's accurate state vector references to the created map.

2.2. *Navigation System*

The next system is the navigation system, which plans the trajectory path and controls the robot movement. The system overview of the navigation system is shown in Fig.2. Each sub-system described as follows:

2.2.1. *3D and 2D Cost mapping*

The first sub-module of the navigation system is 3D and 2D cost mapping used to transform the 2D occupancy grid map provided from RTAB-Map into costmap used for planning route. The structure of our cost mapping module references from [13]. It was also implemented inside the ROS [14] costmap_2d module. However, we also developed a new range sensor layer to deal with noisy ultrasonics using a probabilistic model. The overall system of costmap is divided into two parts: global and

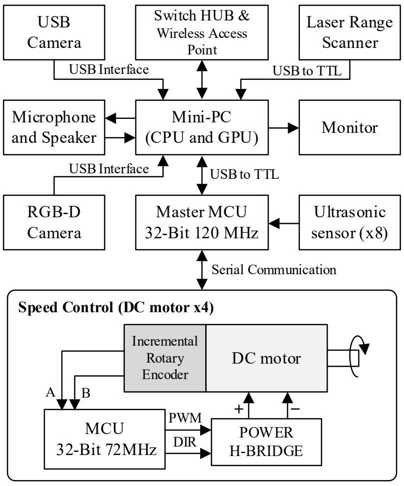


Fig.3 Low-level Control System

local costmaps, which both parts will be used in the next sub-module separately.

2.2.2. *Motion Planner and Motion Controller*

The next sub-module inside the navigation system is the motion planner or trajectory planner module. The motion planner consisted of two parts. The first one is the global planner. It is used for planning a trajectory to reach the goal in the long term. It is called as the global path. Another one is a local planner. It used to deal with immediate situations by trying to reach the global path. We called it the local path. If there is no obstacle the local and global paths will overlap. The first time we build the robot we apply the Dijkstra's/A*global planner together with the dynamic window approach (DWA) [15], [16] local planner which both implemented inside ROS navigation stack. Then, we found that the DWA local planner was not suited for our robot. Because the robot uses a holonomic drive system which increases the complexity for the DWA local planner by “n” times. When “n” is a number of samples used to exploring the velocity space for each axis. For example, the differential drive robot has sample space for all velocity space: O(n2) because it has two of freedom. Nevertheless, for a holonomic robot, the complexity increases to be O(n3) because it has three degrees of freedom. Besides, it also has unpredictable behavior because its algorithm is a sampling base. The user may be confused about its behavior and feel uncomfortable. Therefore, we decided to develop a new local planner which we refer to elastic band planner [17]. Because it has lower complexity,

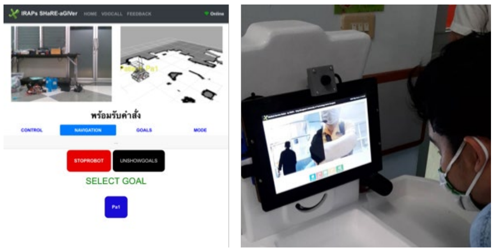


Fig.4 (Left) The medical staff can save position and select goal for the robot. (Right) Telemedicine using peerCalls localhost.

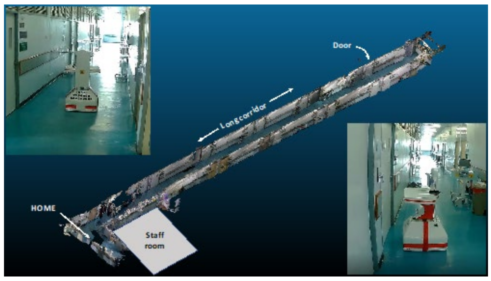


Fig.5 The robot's working area in memory of the mixed map and image reconstruction in 3D

suitable for a holonomic drive system, and also has a predictable behavior. The user can get acquainted in a short time. In summary, we used Dijkstra's/A* for the global planner and elastic band for the local planner. The final step for the Motion Planner and also Navigation system is Motion Controller. We apply a simple PD controller to control both the position and orientation of the robot for each time step.

2.3. *Low-level Control System*

The robot contains a powerful Mini-PC (CPU and GPU units) that can handle harder computational tasks. For the module interface of Mini-PC that can communicate with all the sensors and other modules. A detailed scheme of the low-level control system is shown in Fig. 3. The drive subsystem, four DC motors are equipped with incremental rotary encoders (resolution 12 bits) for rotational speed measurement. PID controller is implemented in the control loop of the speed controller of each DC motor.

3. Verification

The robot was tested in the Covid-19 state quarantine from June to September 2020 at the Queen Sirikit Naval Hospital, Thailand. The medical staff could autonomously send the robot to deliver food and things to the front of infected people's rooms. When the robot had reached the goal, the medical team could make a video call from the staff's room to communicate with the patients without personal contact, as shown in Fig. 4. The robot moved along the long corridor to stop at the selected room on the left or right side in Fig. 5. The robot has logged data for four months to a google sheet, and we can summarize the operational performance. The robot gets the active command 5,448 times in 94 days to go to the rooms. The robot has an average operating time of 12 hours and 58 orders per day. It can reach the goals of 94% and 6% missing because there is something on the goal's position. The robot waits for clearing another object at the destination until a timeout and cancel the order by itself. The robot can continue working from 6 to 8 hours when starting with a full charge, depending on the number of orders.

4. Conclusions

The robot is a success to use in a real situation of state quarantine. The medical staff can look after the patients from a distance using the robot for delivery and communication. The robot can go to the destination automatically with obstacle avoidance and is easy to control through the web application by any device. However, the team is developing more functions to deploy to the robot by the doctor's advice.

Acknowledgements

This work is supported by King Mongkut's University of Technology North Bangkok, Queen Sirikit Naval Hospital, Navamindradhiraj University, PTTGC and Polyfoam Manufacturing Co., LTD.

References

1. WHO. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19); WHO: Geneva, Switzerland, 2020.
2. Yang, G. Z. et al., "Combating COVID-19—The role of robotics in managing public health and infectious diseases," ed: Science Robotics, 2020.
3. France-Presse, A. Robot Receptionists Introduced at Hospitals in Belgium. The Guardian. Available online: https://www.theguardian.com/technology/2016/jun/14/ro bot-receptionists-hospitals-belgiu m-pepper humanoid.
4. Hurst, D. Japan Lays Groundwork for Boom in Robot Carers. The Guardian. Available online: https://www.theguardian.com/world/2018/feb/06/japan-robots-will-care-for-80-of-elderly-by-2020
5. Care is a Team Effort. Diligentrobots. Available online: https://diligentrobots.com.
6. Panasonic Autonomous Delivery Robots—HOSPI—Aid Hospital Operations at Changi General Hospital.
7. TUG-Change Healthcare. Aethon. Available online: https://aethon.com/mobile-robots-for-healthcare
8. J. Inthiam and C. Deelertpaiboon, "Self-localization and navigation of holonomic mobile robot using omni-directional wheel odometry," TENCON 2014 - 2014 IEEE Region 10 Conference, Bangkok, 2014, pp. 1-5, doi: 10.1109/TENCON.2014.7022281.
9. R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," 2011 IEEE International Conference on Robotics and Automation, Shanghai, 2011, pp. 1-4, doi: 10.1109/ICRA.2011.5980567.
10. P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pp. 239-256, Feb. 1992, doi: 10.1109/34.121791.
11. Pomerleau, F., Colas, F., Siegwart, R. et al. Comparing ICP variants on real-world data sets. Autonomous Robots 34, 133–148 (2013). https://doi.org/10.1007/s10514-013-9327-2
12. M. Labbé and F. Michaud, "RTAB-Map as an Open-Source Lidar and Visual SLAM Library for Large-Scale and Long-Term Online Operation," in Journal of Field Robotics, vol. 36, no. 2, pp. 416–446, 2019.
13. D. V. Lu, D. Hershberger and W. D. Smart, "Layered costmaps for context-sensitive navigation," 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, 2014, pp. 709-715, doi: 10.1109/IROS.2014.6942636.
14. Quigley, Morgan, et al. "ROS: an open-source Robot Operating System." ICRA workshop on opensource software. Vol. 3. No. 3.2. 2009.
15. O. Brock and O. Khatib, "High-speed navigation using the global dynamic window approach," Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C), Detroit, MI, USA, 1999, pp. 341-346 vol.1, doi: 10.1109/ROBOT.1999.770002.
16. D. Fox, W. Burgard and S. Thrun, "The dynamic window approach to collision avoidance," in IEEE Robotics & Automation Magazine, vol. 4, no. 1, pp. 23-33, March 1997, doi: 10.1109/100.580977.
17. S. Quinlan and O. Khatib, "Elastic bands: connecting path planning and control," [1993] Proceedings IEEE International Conference on Robotics and Automation, Atlanta, GA, USA, 1993, pp. 802-807 vol.2, doi: 10.1109/ROBOT.1993.291936.

Crack Detection in a Concrete Structure Using an Underwater Vehicle

Yuya Nishida*
Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka 808-0196, Japan

Naoto Sohara*2
Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka 808-0196, Japan

Shinsuke Yasukawa*3
Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka 808-0196, Japan

Kazuo Ishii*4
Kyushu Institute of Technology
2-4 Hibikino, Wakamatsu, Kitakyushu, Fukuoka 808-0196, Japan
*E-mail: y-nishida*1, s-yasukawa*3, ishii*4@brain.kyutech.ac.jp, sohara.naoto395@mail.kyutech.jp*2*
https://www.kyutech.ac.jp/

Abstract

In this paper, to realize efficient underwater infrastructure inspection, automatic crack detection by image processing is proposed. In first process of our method generates enhanced image based on the absorbance from the turbidity meter and removes background component, and then detects crack from the enhanced image by using decision tree learning algorithm. This paper explains the algorithm of our method and shows evaluation experiment results.

Keywords: Crack detection, image processing, Underwater vehicle

1. Introduction

Almost social infrastructures that were built at the time of high economic growth have become deteriorated and have a risk of accidents and traffic restrictions [1]. Japanese Ministry of Land, Infrastructure, Transport and Tourism made the road law that the infrastructures such as a bridge and tunnel should be inspected every five years, in 2014. However, infrastructure inspection is difficult with human hands alone because there are about 730,000 bridges and more than 10,000 tunnels in Japan. In particular, the inspection of underwater infrastructures is not progressing because the inspection is inefficient and dangerous to divers. Several institutes have been researched underwater infrastructure inspection method by using a ROV (Remotely Operated Vehicle) and image processing [2]. However, cracks on the underwater infrastructures are difficult to automatically detect by the image processing, because underwater images taken by the ROV depends on the water turbidity that varies from the site.

Our laboratory has been developing inspection method by using the USV (Unmanned Surface Vehicle) and ROV to realize efficiency inspection of underwater infrastructure [3]. This paper proposes crack detection method by using photo image from a camera and

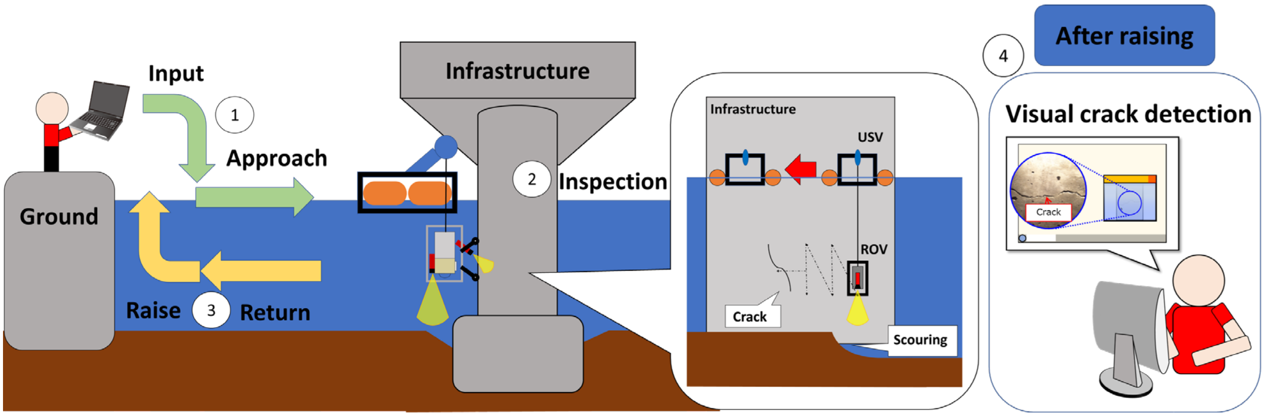


Fig. 1 Underwater infrastructure inspection by the USV and the ROV

absorbance from a turbidity meter installed on the ROV, and experiment results for the evaluation of our method are shown.

2. Underwater infrastructure inspection

2.1. *Inspection procedure*

The underwater infrastructures like valley highway pillar are located in a lake and a river that are hard for people to go. The ROV which its movement is restricted by the umbilical cable is hard to approach the infrastructures for them inspection. Our laboratory developed inspection method that a USV supports a ROV for underwater infrastructure inspection. Figure 1 shows operation overview and inspection procedure by using the USV and the ROV. The USV has four thrusters for horizontal movement, a GPS, IMU for positioning, wireless LAN for communication to the operator, and batteries. Fist, the operator deploys the USV with the ROV in the water and the USV moves to target infrastructure by heading and position controls. Second, the ROV dives by a winch mounted on the USV and captures the infrastructure images by using a front camera and a LED light. At this time, the ROV moves to horizontal direction by the USV thrusters and the ROV heading is controlled its thrusters. The ROV and the USV are connected an umbilical cable, and the operator on the quay monitors infrastructure image captured by the front camera of the ROV during inspection. After the inspection, the USV that stores the ROV returns to the quay near the operator and is recovered by the operator. Finally, cracks on the infrastructure is detected from captured images by offline image processing that is explained in Chapter 3.

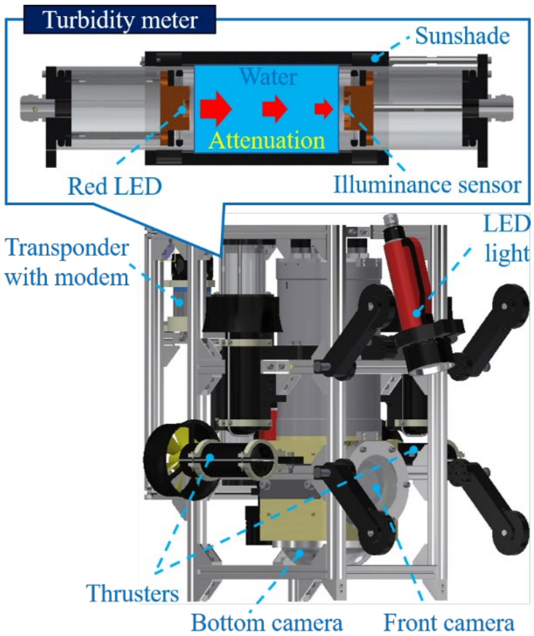


Fig. 2 The ROV equipped a turbidity meter

2.2. *The ROV system for inspection*

The ROV shown in Fig.2 is used for inspection of underwater infrastructure. The ROV has a front camera for crack inspection of the infrastructure and a bottom camera for scouring inspection of seafloor near the infrastructure as observation device. Two thrusters are used for that the ROV direction is controlled and the infrastructure is approached. The transponder with modem function includes the depth sensor and IMU sensor, and the USV measures the ROV position by using it. A turbidity meter on the rear consists of a red LED, an illuminance sensor and sunshade, and surrounding water gets into the sunshade. The illuminance meter measures the red LED light intensity I' that is attenuated by the water. Relationship between Light intensity I_0 which is irradiated by the LED and I' is expressed by following the Lambert-Beer law:

$$A = -\log\left(\frac{I'}{I_0}\right) \qquad (1)$$

where A denotes the absorbance that depends on the distance traveled by the light. For distance-independent values are desirable to represent the underwater light environment, equation (1) is rearranged as following.

$$\beta = -\frac{1}{d}\log\left(\frac{I'}{I_0}\right) \qquad (2)$$

d denotes the distance the LED and the illuminance sensor, our turbidity meter outputs β that denotes the attenuation coefficient.

3. Crack detection method

3.1. *Preprocessing*

Cracks on the underwater infrastructure are detected by image processing from the images captured by the ROV. The image processing is separated the preprocessing and the detection process, as shown in Fig.3. The contrast between the cracks and the infrastructure surface is important for crack detection, because the cracks is detected based on the brightness, not color. First step generates high contrast images by using the absorbance coefficient from the turbidity. Second step searches crack position from the images by using the decision tree learning.

Because light in the lake and the river diffuses by floating particles such as the sand and the plankton, degraded images are taken by the ROV during the underwater infrastructure inspection. The degraded image $\boldsymbol{g}$ is represented by using raw image $\boldsymbol{f}$.

$$\boldsymbol{g} = \boldsymbol{H} * \boldsymbol{f} \qquad (3)$$

where * means the convolution operation. $\boldsymbol{H}$ in equation (3) denotes point spread function and is represented multiplication of line spread functions in u and v axis of image coordinate, and those functions are approximated by the gaussian function. Undegraded image $\boldsymbol{f}$ is obtained

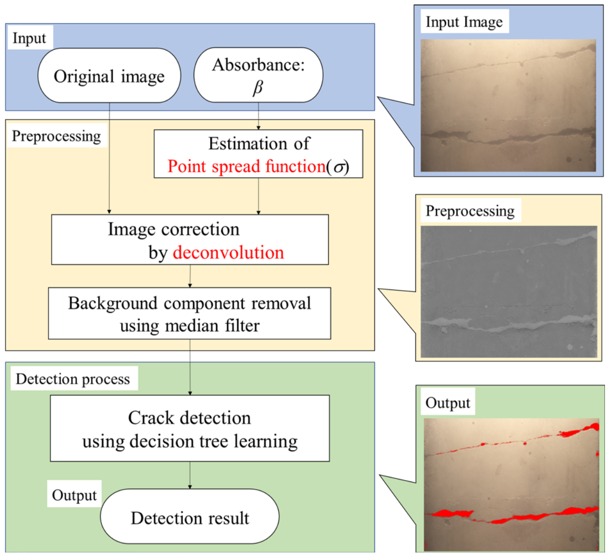


Fig. 3 Crack detection procedure

by deconvolution operation of $\boldsymbol{H}$ on equation (3). In preprocessing, this research generates high contrast image that background component is removed from $\boldsymbol{f}$ by using median filter.

3.2. *Detection process*

This process uses the decision tree learning of classification tree that can obtain complex identification boundaries by combinations of simple identification roles. The classification tree is learned by CART (classification and regression tree) that threshold for the classification is decided by following Gini coefficient [4].

$$L(t) = \sum_{j\neq 1} P(C_i|t)P(C_j|t) \qquad (4)$$

$P(C_i|t)$ denotes the probability that node t in the tree selects data of class i, and $P(C_j|t)$ denotes the probability that node t in the tree selects data of class j. In the Cart, the decision tree learning is learned to maximize following equation expressed by the Gini coefficient.

$$\Delta L(t) = L(t) - \{p_L L(t_L) + p_R L(t_R)\} \qquad (5)$$

p_L and p_R represent the probability of being classified into left and right node. $L(t_L)$ and $L(t_R)$ denote Gini coefficients left and right nodes. In this research, the decision tree learning is learned by using equation (5) until that the node in the tree reaches 100 or more.

The input vector to the decision tree learning is 32 demission that consists of the brightness of target pixel (1), each brightness after applying 3 square median filters

(3), each brightness after applying 2 non-square median filters from 8 directions (16), each brightness after multiscale line enhancement by 11 types Hessian matrix (11) and noise candidate data (1). The brightness of target pixel is used as input vector, because the decision tree learning in this research classifies target pixel into crack and back background pixel. If only the brightness is used as input vector, false recognition can occur due to the surface color of the infrastructure. Because the brightness after applying the gaussian filters is included to input vector, classification results in the learning are less sensitive to the background color [4]. Each brightness after applying by the Hessian matrixes is used as input vector, because enhanced cracks are extracted by multi scale line enhancement using the Hessian matrixes [5]. The brightness like noise is included to the input vector, to reduce the effect of the impulse noise.

4. Experiment

4.1. *Parameter estimation*

As mentioned in chapter 3, the image taken by the ROV in the turbid water degrades based on the point spread function that is represented multiplication of the line spread functions approximated by gaussian function. However, nobody knows how the light is spread by turbid water. To analyze the relationship of the attenuation coefficient and the standard deviation for the point spread function, this research performed photographing experiment in the turbid water using the ROV with the turbidity meter. The clack scale which has lines from 0.03mm width to 1.50mm width was used for the photographing target. The turbid waters from 0 degree to 20 degrees were made by using the Kaolin which is also used for JIS turbidity standard. Figures 4 shows each picture and the brightness in the turbid waters of 0 degree and 25 degrees. At this time, target clack was 0.2mm width line that is the minimum requirement for the repair. The turbidity meter outputted 3.6 in 0 degree water and 4.64 in 25 degrees water, as attenuation coefficient. The picture in 25 degrees water was darker overall than the picture in 0 degree and the clack line were blurry. Figure 5 shows the standard deviations of approximated gaussian function in each brightness versus absorbance coefficient from the turbidity meter. This research used approximated quadratic function which fits to the

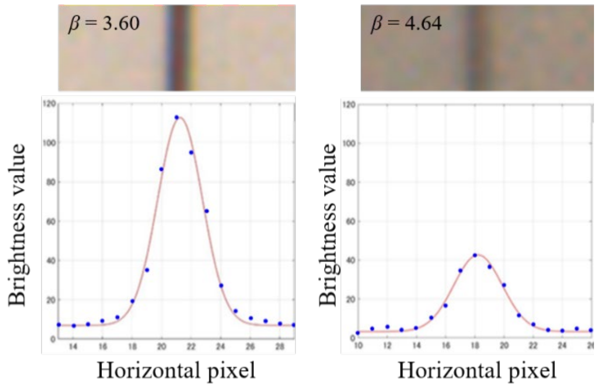


Fig. 4 Comparison of brightness distribution

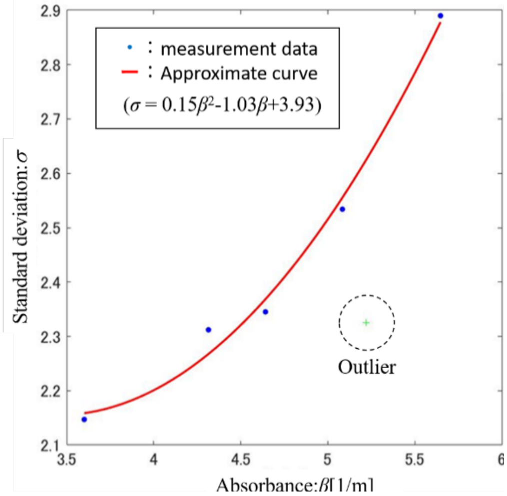


Fig.5 Absorbance coefficient vs. standard deviation

standard deviations without the one outlier data, to detect the cracks.

4.2. *Crack detection*

The cracks on the concrete block was detected by our method for evaluation. Figure 6 shows the image taken by the ROV in 25 degree water, grayscale image corrected by using the point spread function, the image with background components removed and detection results. As a result of comparing the raw image and the corrected image, the contrast of the crack edge on the corrected image was higher than the contrast of raw image cracks. Brightness unevenness and concrete pattern were removed by process using median filters. Although there were some cracks that could not be detected, our method detected the cracks with few false positives. Table 1 shows accuracy rage and sensitivity with and without the correction by the point spread

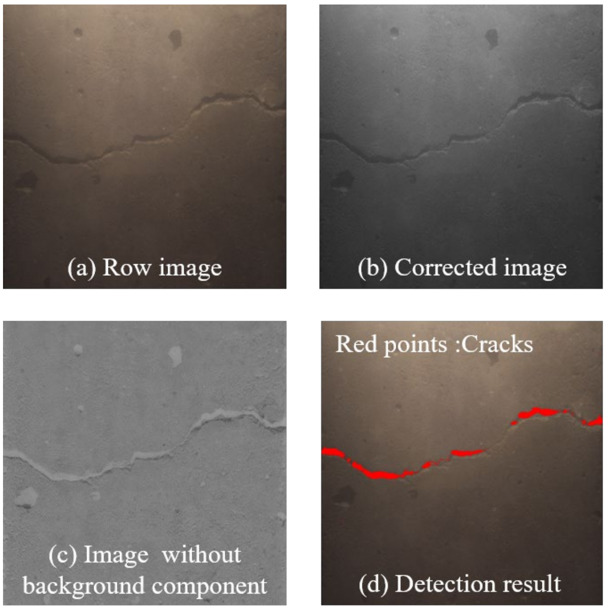


Fig.5 Absorbance coefficient vs. standard deviation

function. Although performance of the crack detection with and without the correction was almost same until 10 degrees water, the accuracy rate and the sensitive were improved with turbidity from 15 degrees to 20 degrees, and the sensitivity of our method was greatly increased. This means that the detection rate was improved by the correction using the point spread function.

5. Conclusions

This paper explains efficient underwater structure inspection by using the USV and the ROV, and the crack detection method in the turbid water is proposed. The underwater infrastructure is observed by the ROV equipped the turbidity meter which can measure the absorbance coefficient of the turbid water. In preprocess, the image taken by the ROV is corrected by using the point spread function and high contrast image is generated by using median filters. Our method detected cracks on the underwater infrastructure in the turbid water from the image enhanced by the preprocess, with high accuracy rate.

References

1. Road Bureau, Aging of social infrastructure, Current statu s of aging and issues related to aging countermeasures, Ja panese Ministry of Land, Infrastructure, Transport and To urism, https://www.mlit.go.jp/hakusyo/mlit/h25/hakusho/h26/html/n1131000.html, *in Japanese*
2. Sugimoto Hideki, et al., Underwater Structure Diagnosis System by Remotely Operated Vehicle, Journal of the Robotics Society of Japan, Vol.34, No.8, pp.505-506, 2016, *in Japanese*
3. Takumi Ueda, et al., Inspection system for underwater structure of bridge pier, Proc. of International Conference on Artificial Life and Robotics, DOI: 10.5954/ICAROB.2019.OS21-2, Beppu, 2019
4. Naoya Kuramoto, et al., Detection of pavement crack from image by decision tree learning, Journal of Japan Society of Civil Engineers, Vol.71, No.2, pp.823-830, 2015, *in Japanese*
5. Yusuke Fujita, et al., Automatic and exact crack extraction from concrete surfaces using image processing techniques, Doboku Gakkai Ronbunshuu F, Vol.66, No.3, pp.459-470, 2010, *in Japanese*

Table 1 Accuracy rate and sensitive in each turbidity

Turbidity	Accuracy rate		Sensitivity	
	With correction	No correction	With correction	No correction
0	0.992	0.990	0.583	0.358
5	0.996	0.995	0.729	0.713
10	0.995	0.995	0.655	0.672
15	0994	0.990	0.625	0.586
20	0.994	0.974	0.642	0.425
25	0.994	0.972	0.5851	0.294

Development of Shock Sensitive Tiny Dummy Robot for Junior High School Rescue Robot Challenge

Kazuo Kawada, Kesuke Iuchi, Keita Murai and Hiroyuki Y. Suzuki
Graduate School of Humanities Social Sciences, Hiroshima University, 1-1-1 Kagamiyama Higashi-Hiroshima, Hiroshima 739-8524, Japan
E-mail: kawada@hiroshima-u.ac.jp, hiro-suzuki@hiroshima-u.ac.jp

Abstract

Junior High School Rescue Robot Challenge is an annual activity held by Hiroshima University with sponsorship of a construction machinery company. The theme in 2020 was to convey injured people "tenderly" from top of half-demolished building to ground. We developed shock sensitive tiny dummy robot, controlled by M5STACK microcomputer since it rigged with accelerometer. The remote (wired) controlled robots developed by junior high school students conveyed the dummy and evaluated their performances including "tenderness" of the robots using the dummy.

Keywords: Rescue Dummy, Robot Evaluation, Technology Education, Junior High School

1. Introduction

Robot contest is an attractive activity for both students and instructors in junior high school technology education. Planning, processing, assembling and manipulating robots by themselves make students foster ingenuity and creativity in a great deal. Such experiences can increase interest in today's science and technology developments as well.

A number of robot contests[1-3] have been held heretofore, in Japan, in which they claimed that those contests have been successful to foster the capacities of invention, logical thinking and creativity in young participants. Those activities, on the other hand, focused on "winning or losing of a game" for evaluation of the robots. Introduction of competition is a good way to arouse incentive in participants' minds, but it also has side effect of narrowing the view on the subject. Any robots contain complexity in them, regardless of the level of the product. They are not machines of single function. We always have to treat the robots as system. Evaluation of robots also, therefore, must be systematic. In the other word, it must be done from various viewpoints. Remind that the guideline for junior high school technology education (stated by Ministry of Education, Culture, Sports, Science and Technology) also claims that we have to deal with humanities in a context of development of technologies. Manufacturing education must be humanity development education as well. Robots of only high performances must not be required, but robots which can consider and cooperate with others will be expected.

In such context, we abandoned the word "Contest" from our robot activity and chose "Challenge" as an alternative. Faculty of Education in Hiroshima University has been holding "Rescue Robot Challenge" from 2014 for junior high school students[4]. It formerly had been named "Rescue Robot Contest" and held since 2003 with the aim of deepening their understanding of social issues and the relationship between people and robots through the creation of robots with an awareness of tenderness to human[5].

The basic theme of our "Challenge" is human rescue from disastrous conditions. Japan is known as an earthquake-prone country. We have faced, in recent years for example, the Kumamoto Earthquake in 2016, the Hokkaido Eastern Iburi Earthquake in 2018, etc. It was natural to desire a rescue robot. In this paper, we are going to describe theme, results and evaluating points of the "Challenge" in 2020.

2. Challenge Theme in 2020

From the era of the "Contest", our activity has used the theme of "quickly rescuing a dummy doll (of 1/8 of real human) from a simulated earthquake site and transport it to a safety zone". Time of rescue has been evaluated as basic data, of course, but ideas for rescuing methods, humanitarian standpoint have been also evaluated. In particular, we have focused on the "tenderness" of the robots to the doll during evacuation.

In 2020's activity, we developed a small shock-sensitive dummy robot. The dummy was put on the remote (wired) controlled robots developed by the middle school students and checks how much shocks and/or vibrations are given during evacuation.

Figure 1 shows the rescue field used in 2020. The robots with dummies are set in the starting area on the roof of the half-demolished building, then start to rescue dummies passing through form Zone A to C. They can choose different rout in Zone B of, i) stairs, ii) a gap with vertical walls and iii) a steep slope. On the ground (Zone C) there are obstacles. The robots finally arrive at the safety zone.

A team consists of 2-5 members of Junior high school students. Every teams firstly give a 4-minute presentation using screen showing about the characteristics of the robot they developed, then followed by a 3-minute rescue activity using real field.

3. Development of Rescue Dummy

For simplicity, we used a small tapper box (83 mm in length, 83 mm in width, and 45 mm in height) as dummy's body (Fig. 2), although there are not protrudes of legs, arms or a head. We chose a microcomputer module of M5STACK (Fig. 3) as controller since it equipped with accelerometers. Gross weight of the dummy was 81 g.

We set thresholds of the accelerometer of ±1.5g[m/s^2] in two directions of X and Y. When the dummy got two acceleration of threshold or more, LCD (Liquid Crystal

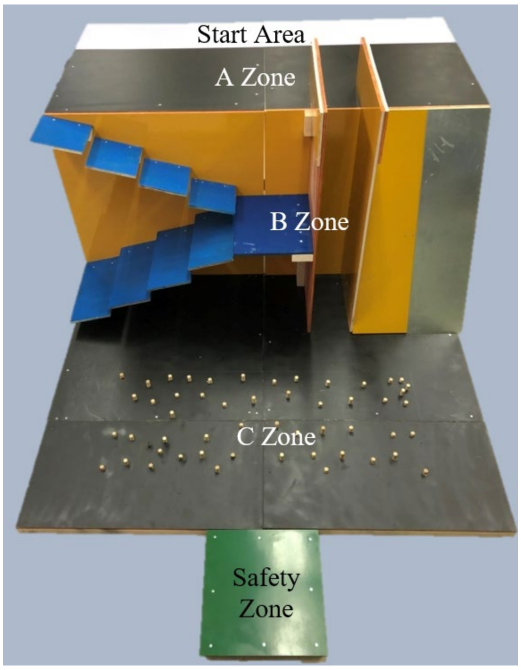


Fig.1 Rescue Field.

Display) of the M5STACK turns red and emits a warning sound. Introduction of the dummy equipped with accelerometer made possible to evaluate more precisely the degree of tenderness of the robot to the dummy.

Fig.2 Rescue Dummy with M5STACK.

4. Result of "Challenge" and Evaluation

The 6th Rescue Robot Challenge in 2020 were held on Feb. 16th, 2020 (still in 2019 as fiscal year). A total of 11 teams were participated. The points they got are summarized in Table 1. Performances of the robots are evaluated from two viewpoints of, i) goal points and ii) tenderness points. Maximum points of each viewpoint are 20, so that the maximum total points are 40. For the goal points, if the robot conveys the dummy to the safety

Fig.3 M5STACK with Accelerometer.

zone in 3 minutes without dropping the dummy, it gets 20 points. The goal points are decreased according to degree of problems happened during the rescue activity.

For the tenderness points, 20 points are given if no warning sound is made, 15 points for 1-3 times of warning, 10 points for 4-6 warnings, 5 points for 7-9 warnings, and 0 points if the warning surpasses 10 times.

Let's see an example. Rescue robot C is shown in Fig.3. Robot C was succeeded to transport the dummy to the safety zone without dropping the dummy, so that it got goal point of 20. On the other hand, the robot C gave 14 to 19 shocks to the dummy, therefore the tenderness points were decreased much. This robot was using magnets to slowly descend the steel slope in zone B. Their trials were successful during practices. But unluckily, the magnet fell off from the steel slope in final performance, resulting in a big shock to the dummy.

In contrast, the rescue robot G (Fig. 4) could transport the dummy without giving shocks. This robot used a pantograph mechanism to descend between the gap of two walls in Zone B.

Another interesting idea was found in the Robot H (Fig. 5). It used crawlers for traveling the stairs in Zone B, which result in a speedy transportation of the dummy to the goal. Although it gave several shocks to dummy, the number and degree of the shocks were much less than other robots which chose the stairs in Zone B. It had ingeniously attached long arms in front and behind the body, by which the robot could prevent to stagger by manipulating the arms.

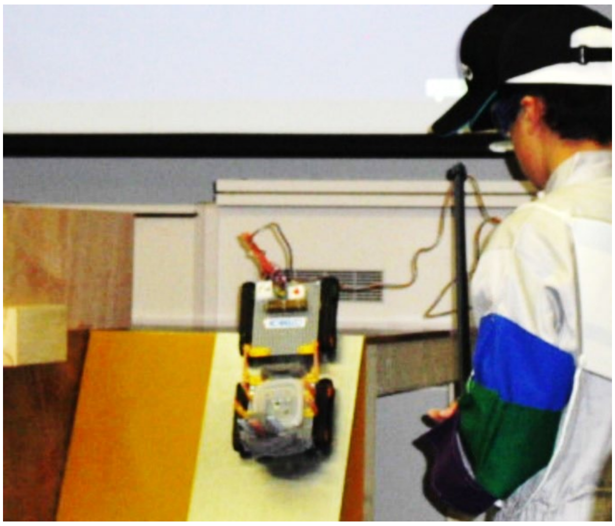


Fig.3 Rescue Robot C.

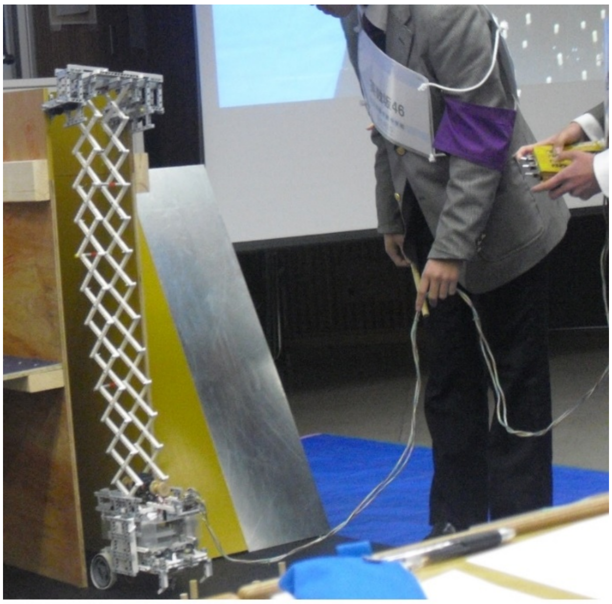


Fig.4 Rescue Robot G.

Table 1. Goal points and tenderness points for each rescue robot.

Rescue robot	A	B	C	D	E	F	G	H	I	J	K
Goal points	0	0	20	0	0	0	20	20	20	0	0
Tenderness points	20	15	5	20	20	0	20	15	15	20	20
Total points	20	15	25	20	20	0	40	35	35	20	20

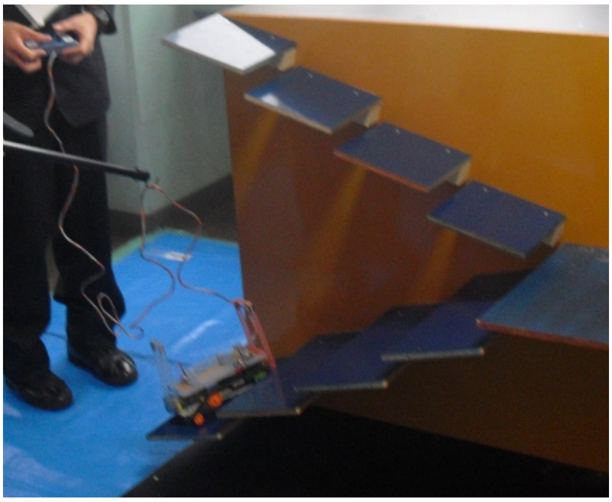


Fig.5 Rescue Robot H.

5. Discussions

The 11 rescue robots can be classified according to descending rout in Zone B. Robots of type 1 chose the gap between two walls. Six robots of A, B, E, G, I, and J chose this way. Type 2, four robots D, F, H, and K chose the stairs. Only the robot C was classified to Type 3, descended a steel slope using a magnet, as we explained in previous section.

The average goal points of Type 1, Type 2, and Type 3 were 6.67 points, 5 points, and 20 points, respectively, indicating that the Type 3 had higher manipulability during transportation in Zone B. On the other hand, the mean of tenderness points for Type 1, Type 2 and Type 3 were 18.33, 13.75 and 5 points. Comparison of the points made clear that the robots of the Type 1 had higher stability during transportation in Zone B. Nonetheless, there are no considerable differences in averages of the total points between Type 1, Type 2, and Type 3, which were 25, 18.75, and 25 points, respectively. These results may indicate that the shortcoming of evaluation from single factor. As we mentioned before, the robots essentially contain "complexity" in their design, mechanism, manipulating manner, and so on, therefore their performances also must be evaluated from various aspects.

6. Conclusions

We developed a dummy robot for "Rescue Robot Challenge in 2020". The dummy was sensible to shocks during rescue transportation. Introduction of shock sensible dummy made possible to evaluate quantitatively the "tenderness" of the robots developed by the middle school teams. Adding evaluation factors can improve the correctness of evaluation of robot performances.

References

1. C. Kasuga, An Analysis of Contesting Robots – Micro Mouse and Sumo Robots, *JRM*, 10(1), 1998, pp. 22-26.
2. M. Asada, RoboCup-97, *JRM*, 10(1), 1998, pp. 30-33.
3. T. Yamamoto, T. Ichikawa and S. Fujisawa, A Consideration on Rescue Robot Contest from Viewpoint of Technology Education, *JRM*, 15(5), 2003, pp. 516-520.
4. K. Okamoto, Y. Shoda, T. Ueno, K. Kawada and M. Nagamatsu, An Approach to 1st Rescue Robot Challenge for Junior High-Shool Students -Present Conditions and Problems-, *Technology Education*, 6, 2016, pp. 8-15. (in Japanese)
5. K. Kawada, M. Nagamatsu and Toru Yamamoto, A Practice of Rescue Roboto Contest in Junior High School, *J.of JSEE*, 58-2, 2010, pp. 33-39. (in Japanese)