A Deep Learning-Based Shopping Support Method for a Visually Impaired Person

Takaya Yamaguchi
Graduate School of Engineering, Kyushu Institute of Technology, 1-1 Sensuicho, Tobata-ku, Kitakyushu, 804-8550, Japan

Seiji Ishikawa[#], Yui Tanjo[*]
Faculty of Engineering, Kyushu Institute of Technology, 1-1 Sensuicho, Tobata-ku, Kitakyushu, 804-8550, Japan
[#] Emeritus Professor of Kyushu Institute of Technology

Email:yamaguchi.takaya571@mail.kyutech.jp, tanjo@cntl.kyutech.ac.jp
[*]Corresponding Author

Abstract

Shopping is one of the most important but challenging activities for a visually impaired people. This paper proposes a method to help them find convenience stores and their entrances based on deep learning using the images provided from a camera mounted on the chest of a user. This paper focuses on convenience stores because they have variety of goods and nationwide coverage. The proposed method employs a transfer learning model to recognize and detect the position of a convenience store, its signboards and logos, entrances of the store, and obstacles such as vehicles parked in front of the store. Then the method gives directional instructions to the user with a voice-function based on the detected store location to guide them going into the store. Experimental results under real environments show effectiveness of the proposed method.

Keywords: Direction instructions, voice guidance, deep learning, convenience stores

1. Introduction

According to the Ministry of Health, Labour and Welfare, the number of people with visual impairments in Japan in fiscal year 2022 was approximately 270,000[1]. Approximately 70% of people with visual impairments go out at least once a week[2]. Furthermore, in a questionnaire regarding difficulties in daily life, the highest percentage reported being unable to shop independently[1]. Furthermore, in a survey regarding difficulties encountered by visually impaired people when going out, 48% responded that finding stores was difficult, and 54% responded that finding store entrances was difficult[3]. Based on the above, the support related to shopping is considered indispensable for visually impaired people.

To address these issues, numerous studies have explored outdoor walking assistance for the people with visual impairments using a camera footage. Examples include the walking assistance application EyeNavi[4], which features route guidance, obstacle detection, and a walking recorder function, and the application Navilens[5], which allows users to scan tags with a camera and receive the scanned information via voice output.

These studies aim to assist travel to destinations by understanding the surrounding environment from camera footage. However, they face issues such that guidance ends upon reaching the destination and leaving directional decisions to the visually impaired persons, making it difficult to guide them to the exact location or entrance of the destination.

This paper proposes a method of detecting all signs, stores, and entrances within commercial facilities from camera footage and guide users to the entrance based on this information. The target commercial facility is defined as a convenience store in this paper. This study specifically targets the three major convenience store chains with the largest number of locations: Store A, Store B, and Store C[6]. This enables support for visually impaired individuals in navigating to the entrance of their destination, contributing to their ability to shop independently.

2. Method

2.1. Creating an Object Detection Model

An object detection model is created using transfer learning with YOLOv8[7]. Based on the result of object detection the proposed method provides guidance to the entrance by indicating the direction of movement. During the training, three different convenience stores Store A, Store B, and Store C are photographed. Each frame of the video is annotated to create the dataset. Object detection accuracy for each class is evaluated using Precision and Recall. They are given by Eqs. 1 and 2.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

Here, TP occurs when the predicted label is the target object and the ground truth label is also the target object. FP occurs when the predicted label is the target object but the ground truth label is not the target object. FN occurs when the predicted label is not the target object, but the ground truth label is the target object.

Note that car annotation is performed for each car appearing in the captured data. Additionally, the Kaggle Car Object Detection [8] dataset is also used. Fig.1 shows an example of training images for Store A, Fig.2 shows an example for Store B, Fig.3 shows an example for Store C, and Fig.4 shows training images for automobiles and entrances.

2.2. *Route Guidance Algorithm*

Based on the objects detected by the trained object detection model, movement instructions are provided. The movement instructions for each object are shown below.

2.2.1. *Signboard Guidance Algorithm*

When a signboard is detected, the signboard's position is indicated. The input image frame is divided horizontally into three sections. If an object is detected in the central 1/3 area, the guidance instruction 'center' is output; if in the right 1/3 area,'right'; and if in the left 1/3 area,'left'. This is because indicating the signboard's position allows the user to determine whether the store is nearby. Fig.5(a) shows an example of signboard detection and an example of the guidance content.

2.2.2. *Store Guidance Algorithm*

Similarly, when a store is detected, the input image frame is divided horizontally into three sections. If the object is detected in the central 1/3 area, the guidance instruction 'center' is output; if in the right 1/3 area, 'right'; and if in the left 1/3 area, 'left'. This allows the user to determine that a store exists within the camera's field of view and where the store is located. Fig.5(b) shows an example of store detection and the corresponding voice guidance content.

2.2.3. *Logo mark Guidance Algorithm*

When a logo mark is detected, the method issues a gaze guidance instruction to position the logo mark at the center of the camera's field of view. If the logo mark is centered, 'correct' instruction is issued, and the visually impaired individual is instructed to proceed straight ahead in that gaze direction. This method is adopted, because simply instructing the user to look in the direction of the logo mark would only let him gaze in that direction, failing to consider hazards outside the camera's field of view. By stopping the user, guiding his gaze, and then proceeding, the system ensures the user is always aware of potential obstacles in his path. An example of logo mark detection and the sample voice guidance content are shown in Fig.5(c). Furthermore, since the logo's location is critical information for guiding the user to the entrance, the types of gaze guidance instructions are finely tuned. Fig.6 illustrates the instruction types based on the logo's location.

2.2.4. *Entrance Guidance Algorithm*

When an entrance is detected, guidance is provided based on its position. Since the entrance is the final destination and even slight deviations in position guidance could have an impact on safety, the detection threshold is set more strictly than other targets. The input image frame is divided horizontally into three sections with the central area defined as 1/5 of the total. When an object is detected in this central 1/5 area, the system outputs the voice guidance 'center': When detected in the area to the right of this, it outputs 'right'; and when detected in the area to the left, it outputs 'left'. This enables safe and accurate guidance to the store entrance for a user with visual impairment. Fig.5(d) shows an example of entrance detection and the content of the voice guidance.

2.2.5. *Obstacle Guidance Algorithm*

Guidance to the entrance alone may result in collision with obstacles. Therefore, obstacle detection and avoidance instructions are provided. This study employs an object detection-based vehicle avoidance algorithm. When a vehicle is detected, if the detected vehicle is centered in the camera image and its bounding box area exceeds a threshold, a 'danger' instruction is issued. When both the logo mark and a vehicle are detected, the x-coordinates of the left and the right edges of the logo mark's bounding box and the vehicle's bounding box are referenced. Let the left edge x-coordinate of the logo mark be denoted by x_{ll}, the right edge x-coordinate be x_{lr}, the left edge x-coordinate of the vehicle be x_{cl}, and the right edge x-coordinate be x_{cr}. If the condition given in Eq.3 is satisfied, it is determined that there is no obstacle in the direction of movement, otherwise it is determined that an obstacle exists and the method issues an avoidance instruction.

$$(x_{ll} > x_{cr}) \cup (x_{lr} < x_{cl}) \quad (3)$$

Next, we explain the avoidance instructions. When avoiding an obstacle, we refer to the x-coordinates of the center of the logo mark and the center of the vehicle's bounding box. If the center coordinate of the logo mark is to the left of the vehicle, an instruction 'avoid left' is issued: If it is to the right, an instruction 'avoid right' is issued. This allows the user to approach the entrance while avoiding obstacles. Fig.5(e) shows an example of simultaneous detection of vehicles in front of a store and the logo of the store along with an example of the voice guidance content.

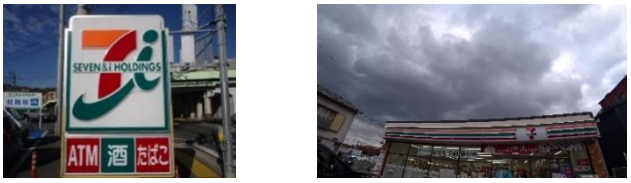


Fig.1 Example of learning images of Store A

Fig.2 Example of learning images of Store B

Fig.3 Example of learning images of Store C

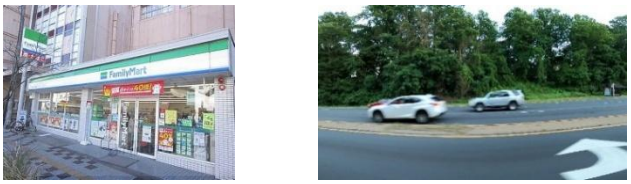


Fig.4 Example of training images of cars and entrances

2.2.6. *Arrival Guidance Algorithm*

In the proposed method, 'arrival' is considered that the user has reached within 3 meters of the convenience store entrance and the entrance is positioned at the center of the camera footage. To determine if the user is within 3 meters of the entrance, the vertical length of the bounding box is used with a threshold of 706 pixels. Fig.5(f) shows an example of the image of arrival and the audio guidance content upon arrival.

3. Experimental Results

Table 1 shows the number of images and detection accuracy for each class. The total number of images is 10,652. We conducted guidance experiments with respect to three stores — Store A, B and C— each 10 times, total 30 experiments, to evaluate the accuracy of the proposed method. Accuracy, representing the correct answer rate, was adopted as the evaluation metric. The definition of accuracy is given in Eq.3in which E_s is the number of successful experiments, and E_{all} is the total number of experiments.

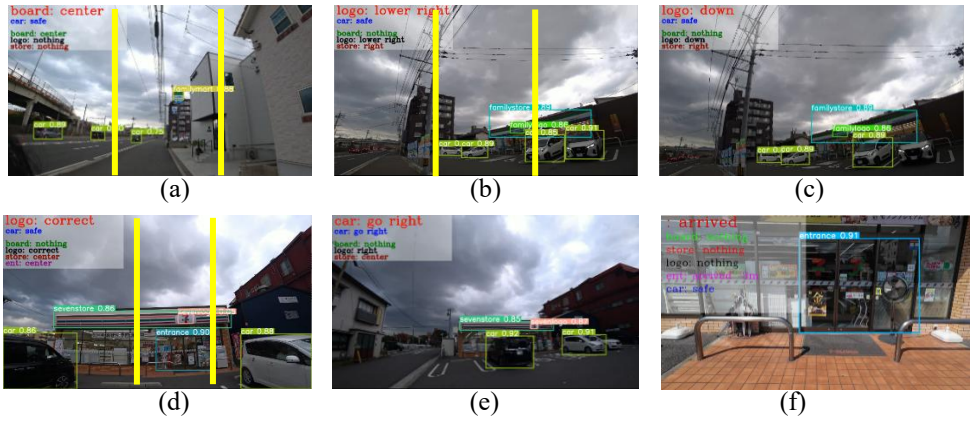


Fig.5 Example of guidance instructions
(a) Signboard, (b) store, (c) logo mark, (d) entrance, (e) obstacle, (f) arrival.

Table 1. The number of employed images and the result of recognition

Class	Training data	Test data	Precision	Recall
Store A Signboard	1513	371	0.93	0.90
Store A Logo mark	1069	858	0.97	0.89
Store A Store	871	968	0.77	0.82
Store B Signboard	582	266	0.98	0.99
Store B Logo mark	1066	264	0.84	0.93
Store B Store	1162	263	0.95	0.83
Store C Signboard	677	143	0.95	0.96
Store C Logo mark	635	280	0.97	0.99
Store C Store	697	282	0.93	0.97
Car	5545	1812	0.65	0.85
Entrance	790	397	0.87	0.50

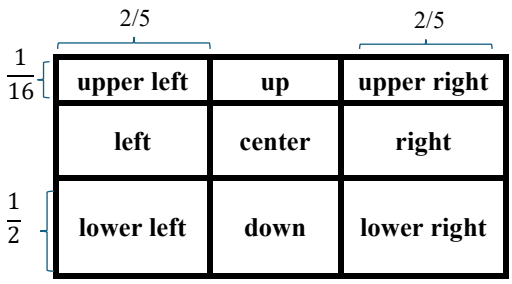


Fig.6 Eye-tracking guidance during logo detection

$$Accuracy = \frac{E_S}{E_{all}} \times 100[\%] \quad (3)$$

The accuracy was 70% for Store A, 70% for Store B and 60% for Store C. The average accuracy was 67% for the entire experiment.

4. Discussion

The accuracy of the proposed guidance method is examined. One of the main factors contributing to the decrease in the accuracy is that, when a car parks near the entrance parking lot, guiding solely on comparing the center coordinates of the car and the logo mark sometimes results in the guidance away from the entrance. This indicates the need to improve the method of guiding toward the entrance. However, under other environmental conditions, stable guidance to the entrance was consistently achieved. The proposed method demonstrated a certain level of effectiveness, and further improvement in accuracy is expected by addressing the issue of vehicles parked near the entrance.

5. Conclusions

This paper proposed a voice guidance method using GPS and camera footage to guide a visually impaired individual to the entrance of a store. Based on object detection employing camera footage, the method guides a user to a store entrance providing avoidance instructions when obstacles are detected. Outdoor guidance experiments yielded the average accuracy of 67%.

Future challenges include resolving the issue where guidance accuracy decreases when vehicles are parked in the lot adjacent to the entrance. We plan to implement space recognition using segmentation techniques. Additionally, we aim to expand the target stores and develop a highly versatile system applicable across various commercial facilities.

References

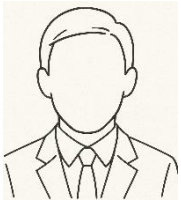
1. Ministry of Health, Labour and Welfare, Social and Welfare Bureau, Disability Insurance and Welfare Department: "Summary of the 2022 Survey on Difficulties in Daily Life (National Survey on the Actual Conditions of Children and Persons with Disabilities Living at Home)," pp. 2-3, 2024.
2. Social Welfare Corporation Japan Federation of the Blind: "Survey on the Current State of Mobility Support for Persons with Visual Impairments"pp.9, 2015.
3. Ministry of Health, Labour and Welfare, Specified Nonprofit Corporation Project Yuai: "Survey on the Current State and Future Direction of Audio Guidance Systems Supporting Walking Mobility for Visually Impaired Individuals" pp. 11-24, 2010.
4. Computer Science Research Institute, Inc.: "Development of a Mobility Support System for the Visually Impaired Using IoT and an Information-Sharing Cloud"
5. Junchi Feng, Mahya Beheshti, Mira Philipson, Yuvraj Ramsaywack, Maurizio Porfiri, John-Ross Rizzo: "Commute Booster: A Mobile Appreciation for First/Last Mile and Middle Mile Navigation Support for People With Blindness and Low Vision" IEEE, p524-525, 2023.
6. Nippon Soft Sales Co., Ltd.; 2023 Edition Convenience Store Outlet Rankings 2023.
7. Dillon Reis, Jordan Kupec, Jacqueline Hong, Ahmad Daoudi: "Real-time Flying Object Detection with YOLOv8" Georgia Institute of Technology, p.1-9, 2023.
8. Edward Zhang: "Car Object Detection" Kaggle, 2023.

Authors Introduction

Mr. Takaya Yamaguchi

Mr. Yamaguchi received his Bachelor's degree in Engineering in 2024 from the Faculty of Engineering, Kyushu Institute of technology in Japan. He is currently a master student in Kyushu Institute of Technology, Japan. His research interests include deep image processing for visually impaired individuals.

Emeritus Prof. Seiji Ishikawa



Dr. Ishikawa graduated from Tokyo University and was awarded BE, ME and PhD there. He is now Emeritus Professor of Kyushu Institute of Technology. He was Visiting Researcher of the University of Sheffield, UK, and Visiting Professor of Utrecht University, NL. His research interests include visual sensing & 3-D shape/motion recovery. He was awarded The Best Paper Awards in 2008, 2010, 2013 and 2015 from BMFSA. Professor Ishikawa is a life member of IEEE.

Prof. Dr. Yui Tanjo

Dr. Tanjo is currently a professor with the Department of Mechanical and Control Engineering, Kyushu Institute of Technology, Japan. Her current research interests include ego-motion analysis by MY VISION, three-dimensional shape /motion recovery, human detection, and its motion analysis from video. She was awarded the AROB Young Author's Award in 2004, and the BMFSA Best Paper Awards in 2008, 2010, 2013 and 2015. She is a member of IEEE, The Information Processing Society, The Institute of Electronics, Information and Communication Engineers of Japan.