Support for Visually Impaired Persons' Understanding of Proximity Space and Action Recognition Based on Pointing

Yasuyuki Murai, Yumiko Ota, Hisayuki Tatsumi, and Masahiro Miyakawa

***Abstract*—It is essentially important for the visually impaired to accurately understand their surroundings. However, the only way for visually impaired people to understand their surroundings is by hearing and touch. In human support, a pointing operation is effective in which the supporter holds the hand of a visually impaired person and points his or her hand toward the surrounding object to indicate its position. The purpose of this research is to develop a system that assists visually impaired people in understanding their surroundings based on this pointing motion. In this report, we propose a system that uses wearable camera and AI to recognize what is around and corrects the veering tendency when a visually impaired person walks.**

***Index Terms*—Visually impaired, veering tendency, deep learning, wearable camera.**

I. Introduction

The act of teaching while holding the hand of a visually impaired person giving assistance to pointing is an effective means of spatial recognition for the visually impaired person. For example, when indicating the position of an object on a desk to a visually impaired person, it is effective not only to indicate the clock position, direction, but also to guide the visually impaired person with your hand. This is considered to be the act of correcting the eccentricity of the direction instead of the sight.

The purpose of this study is to automate the support of near space recognition for visually impaired people [1] and [2]. The target system in this report is a machine learning system that helps the creation of cognitive maps based on finger pointing and calling by the visually impaired.

II. Research Background

There is a behavior recognition / support method for understanding human motion using a wearable camera. The goal of this method is to analyze the meaning of the action from the viewpoint of the attached camera and to provide support to the action. However, if the user is a visually impaired person, it is difficult to learn and estimate by analyzing the meaning of the action from the viewpoint video of the camera. The reason is that, due to visual impairment, the camera's viewpoint image is not always an image that leads to action recognition. This is because the operation of the user includes a redundant part, which makes machine learning difficult.

In order to support cognitive maps for visually impaired people, many studies have been conducted to support walking outdoors [3]. The main approach of the studies was to analyze walking environment images and identify characters such as signboards to identify the position of movement, and to identify pedestrian crossings and signals. In these studies so far, it was difficult to avoid danger in situations where sufficient information could not be obtained during walking.

III. Related Work

Many image recognition services for the visually impaired have appeared, such as OrCam, which uses a dedicated wearable device, and Microsoft Seeing AI, a smartphone app. In addition, "EyeRing" that attaches a camera to the fingertip has been developed by Nanayakkara, S. *et al.* [4]. The camera used in EyeRing was originally developed and includes a control device that fits user's fingertip. This system recognizes an object by pointing the camera attached to the finger to the object. Pointing at the target is a natural motion and can be done by the user without difficulty. In our research, the pointing operation is used as the user interface. These services recognize images captured by a smartphone camera or a dedicated device, and read out characters, identify currency, and show surroundings to the visually impaired. In particular, Seeing AI uses Microsoft's Azure Cognitive Services and can develop applications, so we are considering using it in this research.

IV. Study on Waking Support for The Visually Impaired

Most of the conventional studies that support visually impaired people to walk outdoors have analyzed the image of the walking environment, identified the characters on the signboard, the moving position, and identified pedestrian crossings and signals. In these studies, it was difficult to avoid danger in situations where sufficient information could not be obtained while walking.

One of the reasons why it is difficult to avoid danger is the veering tendency when walking, that is, the characteristic of deviating from the trajectory. Regarding

Manuscript received September 19, 2021; revised March 22, 2022.

Yasuyuki Murai and Yumiko Ota are with Nihon Pharmaceutical University, Tokyo, Japan (email: {murai, arisue-888}@nichiyaku.ac.jp).

Hisayuki Tatsumi and Masahiro Miyakawa are with Tsukuba University of Technology/Department, Ibaraki, Japan (email: tatsumi@cs.k.tsukuba-tech.ac.jp; mamiyaka@gmail.com).

doi: 10.18178/ijmlc.2022.12.6.1119

veering tendency, Sekita et al. Of Tsukuba University of Technology investigated the walking characteristics in a situation where visual information was not obtained by wearing an eye mask on 5 visually impaired persons, 1 blind person and 4 amblyopia [5]. Fig. 1 and Fig. 2 show the results. Fig. 1 shows the subject facing the target at the starting point and walking straight for 11m. As a result of walking each subject three times for a total of 15 cases, there were 11 cases of deviation to the right and 4 cases of deviation to the left. The results were processed by non-linear least-squares approximation and normally distributed. The root mean square veering can be approximated by the cube of the distance, and the probability of leaving 5m or more from the center line when walking 20m was 30.8%.

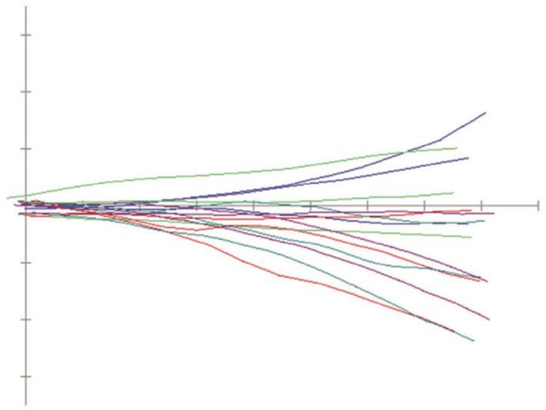


Fig. 1. Veering tendency in straight walking.

The Fig. 2 shows the result of turning around the rectangle line counterclockwise. It can be seen that the deviation veering increases with each turn. The veering tendency of visually impaired people when walking is also a cause of the accident at the station platform.

The system to be developed in this study is to understand the surroundings of the visually impaired by using the user's pointing motion, the wearable camera and its image recognition. At the same time, the system presents to the user, by its voice, the object pointed by the finger out of the surrounding conditions understood by the image recognition. The user selects the target object required for the next action from the presented objects. The target selected by the user is detected from the latest image of the wearable camera and the difference from the previous position is calculated. The obtained difference means the veering tendency. The purpose of this study is to correct the motion of the visually impaired based on the detected veering tendency.

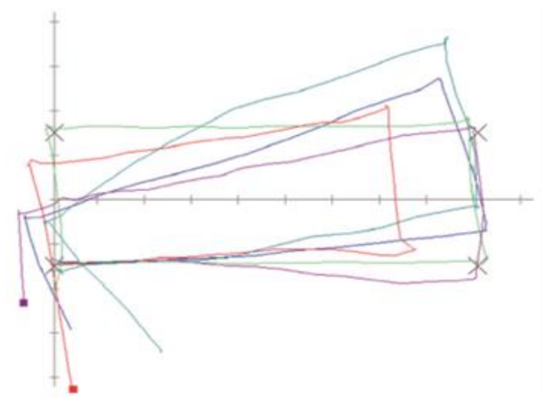


Fig. 2. Veering tendency in orbit.

V. Purpose of This Study

The purpose of this study is to recognize the behavior of visually impaired people and assist them to understand the proximity space, especially by mobile terminals. Specifically, it aims to provide visual impairment compensation support for various actions, including walking support and living support, as well as learning support. A support method to correct various situations including the veering tendency is constructed. For example, a case is considered where a visually impaired person eats using the proposed method. First, the visually impaired performs a pointing operation to check what is located on the table. At the same time as the pointing operation, the first-person viewpoint image of the wearable camera is input to the support system. The system has already machine-learned about meal etiquette. The system understands the situation on the table and advises them on the appearance on the table and the order in which to use the tableware.

The aim is to develop a support technology that performs such a situation explanation, which enables a visually impaired person to make a correction without a helper. The machine-learned system can also be used as a simulator for training the visually impaired person in the action recognition required in the close space.

VI. Prototype of System

To verify the possibility of developing a system for assisting the recognition of visually impaired persons. We prototyped a system to support the recognition of nearby space. This system recognizes an object that is pointed by a user with a finger.

First, a small camera connected to a PC is attached to the fingertip, Fig. 3 Practically, a method of holding a smartphone and taking a picture is also conceivable. Next, the image captured by the fingertip camera is recognized by a VGG16 convolutional neural network that has been trained in advance, Fig. 4. VGG16 is a convolutional neural network that has been trained with over one million images from the ImageNet: http://www.image-net.org/ database, often used in machine learning. The network is 16 layers deep and can classify images into 1000 categories, clocks, desks, PCs, mice, many animals, etc.

Table I shows the development environment of the prototype system.

A program created with Python takes about 100 steps. From the image of the USB camera at the fingertip (1280 × 720 pixels), cut out 400 × 400 pixels at the center position. The extracted image is recognized by the VGG16, and the top five recognition results are displayed above the camera image.

The results of experiments using the developed program are shown below. The PC used was a MacBook Air, and the processing power of the CPU was not so high, and the GPU could not be used, so the frame rate was about 1.8 frames / sec.

Fig. 5 shows the clock on the wall recognized. The image at the upper left of the screen is the image to be recognized. This image is obtained by cutting out 400 × 400 pixels around the center position of the camera image, which is shown by the center cross in Fig. 5 that is equivalent to the position of the fingertip. The image has been reduced by half. The display on the right side of Figure 6 shows the top five recognition results.

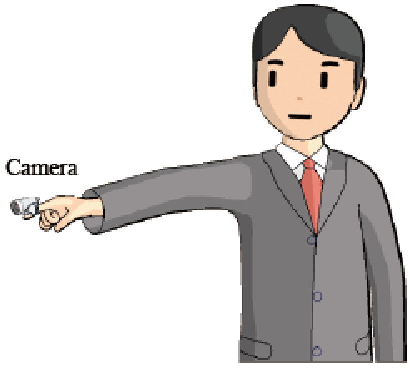


Fig. 3. Camera mounted on fingertip.

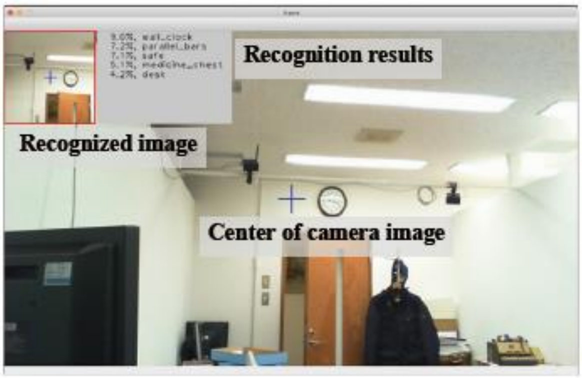


Fig. 4. System output image.

TABLE I: THE PROTOTYPE SYSTEM

Programming tool	PC
Python 3.7	MacBook Air (Retina, 13-inch, 2018)
OpenCV 4.1	1.6 GHz Dual Core Intel Core i5
Keras 2.2.4 VGG16	8 GB 2133 MHz LPDDR3
	Intel UHD Graphics 617 1536 MB
	USB camera: Microsoft LifeCam HD-3000

Fig. 5. Recognizing the clock on the wall.

The numbers in Fig. 6 are recognition scores, and the letters represent recognized objects. The result was that the recognition score was low at 9.0% because the distance to the clock was about 5m, but the "wall clock" could be recognized.

Fig. 7 shows the result of pointing on the desk. The PC indicated by the position of the fingertip is recognized as "laptop" by 32.9%. The second place is "notebook" and the third is "monitor".

The recognition result changes one after another if the camera is moved a little. It is important to determine what is necessary for the visually impaired person from the recognition results, and it is a future subject.

In addition, the prototype program was able to recognize the clock located 5 m away with a recognition rate of 9%. This program, however, cannot show the distance between the camera and the recognized objects though it can recognize objects themselves. It is also necessary that recognition range of the program should be modified and added.

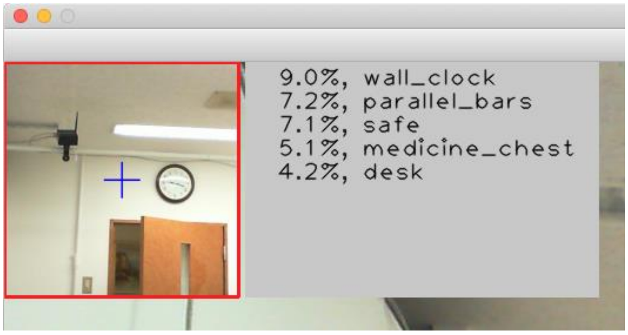


Fig. 6. Top 5 recognition results.

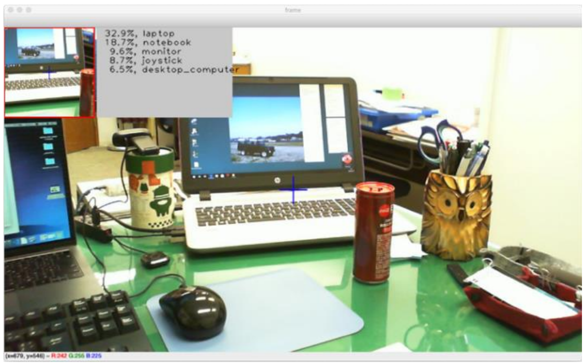


Fig. 7. Recognizing the laptop on the desk.

VII. DETECTION AND CORRECTION OF VEERING TENDENCY

Veering is detected and corrected by the following procedure. First, the target object is determined from the objects captured and recognized by the camera. Next, the difference between the current viewpoint position of the camera and the determined position of the target object is obtained. This difference is the veering from the target. An instruction is given to reduce the obtained veering, and the veering tendency is corrected. The red cross on the mouse in Fig. 8 is the target. The other red cross on the laptop in Fig. 9 is the current viewpoint of the camera, and the red arrow is veering.

As a method for obtaining veering, a function that calculates the optical flow of OpenCV can be used. A test program in Python is, then, created that uses OpenCV's optical flow function to detect movement of objects. The outline of the created program is shown below.

1) Image capture from USB camera
2) Cut out 400 × 400 pixels in the center of the captured color image
3) Make cutout image gray skeleton
4) Calculate the optical flow by processing the grayscale

image with the OpenCV function "calcOpticalFlowFarneback"

5) Illustrate the direction traveled using optical flow

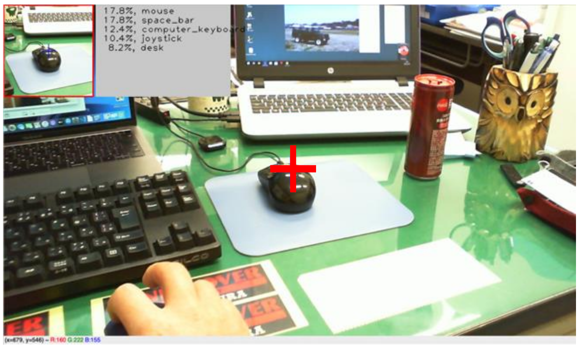


Fig. 8. Target.

Fig. 9. Current viewpoint of the camera.

Optical flow is the pattern of apparent motion of image objects between two consecutive frames caused by the movement of an object or a camera. It is 2D vector field where each vector is a displacement vector showing the movement of points from the first frame to the second. The "calcOpticalFlowFarneback" function used this time calculates the optical flow for all pixels in the image.

The results of the test program are shown in Fig. 10 and Fig. 11. Fig. 10 right is an image showing the optical flow in a stationary state. The optical flow is calculated for all pixels in the image. However, it is difficult to see the display for all pixels, so in this program, the image is divided into 36 by 6x6, and the optical flow is shown only for the pixel at the center, the red dot in the image. Since this figure is stationary, no line indicating movement is displayed. The figure on the left shows the averaged optical flow at each point. Fig. 11 is an image after the camera has moved to the right from Fig. 10. In the figure on the right, the red line extends to the right, indicating that the camera has moved to the right. The figure on the left shows the average value of 36 points, showing that the line extends to the right and moves to the right as a whole.

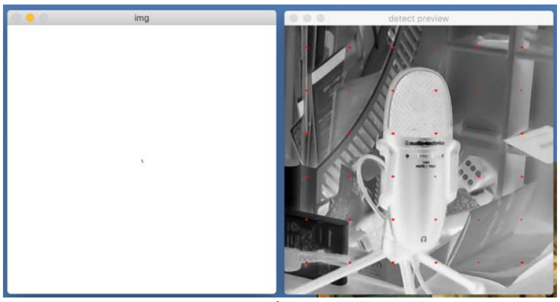


Fig. 10. Before moving.

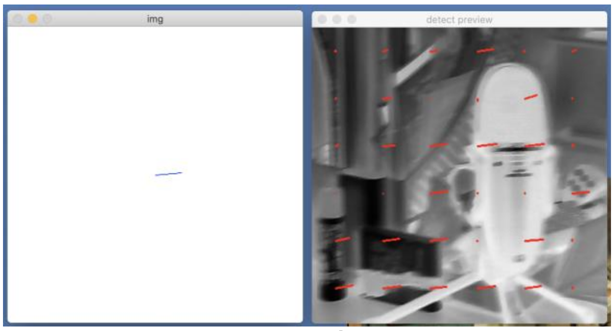


Fig. 11. During the move.

It was confirmed that the moving direction of the camera was detected using the OpenCV function. In the future, we are to develop a method of correcting the user's behavior by obtaining the user's veering tendency from the detected moving direction. We plan to implement the developed method in the prototype system and perform a verification test.

VIII. Conclusion

With the prototype system, the camera at the fingertip was able to recognize what was pointed by the finger.

In the future, based on the results of the experiment, it is to be possible to respond to the environment where the dynamic object exists, and also incorporate the measurement of the distance between the camera and the object and the setting of the recognition range of the object. Additionally, we will construct a machine learning system to correct the excursion based on the analysis of the walking environment image and the pointing and calling of the visually impaired person.

Furthermore, it is necessary to develop a camera that can be worn by the visually impaired and to experiment whether it is viable to use a smartphone.

The constructed system acts on behalf of the visually impaired and corrects the excursion when walking, so that the visually impaired can walk independently.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Hisayuki Tatsumi and Yumiko Ota cooperated in the experiment; Masahiro Miyakawa was an advisor to knowledge about the visually impaired; Yasuyuki Murai wrote the paper; all authors have approved the final version.

Acknowledgment

This work is supported by JSPS KAKENHI Grant Number JP17K01092.

References

[1] Y. Murai, H. Tatsumi, and M. Miyakawa, "Recording of fingertip position on tactile picture by the visually impaired and analysis of tactile information," in *Proc. International Conference on Computers Helping People with Special Needs*, July 2018, pp. 201-208.

[2] Y. Murai, M. Kawahara, H. Tatsumi, T. Araki, and M. Miyakawa, "Congestion recognition for arm navigation, --aids for the visually

impaired," in *Proc. 2010 IEEE Int. Conf. on Systems, Man and Cybernetics*, October 2010, pp. 1530-1535.

[3] T. Yanagihara, "Improvement of pedestrian environment and visual and spatial cognition of people with visual impairment," *Infrastructure Planning Review*, vol. 27, no. 1, pp. 19-31, 2010.

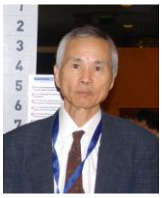
[4] S. Nanayakkara, R. Shilkrot, K. P. Yeo, and P. Maes, "EyeRing: a finger-worn input device for seamless interactions with our surroundings," in *Proc. the 4th Augme Human International Conference*, March 2013, pp. 13-20.

[5] I. Sekita *et al.*, "Difficulties encountered by visually-impaired students while traveling," *TCT Education of Disabilities*, vol. 18, no. 1, pp. 74-79, 2010.

Yasuyuki Murai received the B.A. in 1997 from the SANNO University and the M.S. and Dr. Eng. in 1999, 2004 respectively from the Kanagawa Institute of Technology. In 2003 he joined the Kanagawa Institute of Technology as an assistant professor. In 2006 he moved to the Nihon Pharmaceutical University as a lecturer and works a professor from 2020. His current interest includes soft computing, affective engineering, welfare engineering, and visual impairment compensation technology using haptic and force sensory devices.

Yumiko Ota received the B.P. in 2001 from the Daiichi University of Pharmacy and the M.A. in 2020 from the International University of Health and Welfare. In 2020 she enrolled in the doctoral program at the Nihon Pharmaceutical University Graduate School. The research theme is medical information analysis.

Hisayuki Tatsumi received the B.A., M.S. and Dr. Eng. in 1979, 1981 and 1985 respectively from the Meiji University. In 1986 he joined the Kanagawa Institute of Technology where he worked on enumeration of multi-valued logic functions, failure diagnosis by neural network, and distance field model concept for space representation. In 2002 he moved to the Tsukuba University of Technology as an associate professor and works a professor from 2008 to the present where he teaches visually impaired students in the Department of Computer Science. His current interest includes soft computing, affective engineering, welfare engineering, and visual impairment compensation technology using haptic and force sensory devices.



Masahiro Miyakawa received the B.A. and M.S. in 1966 and 1968 respectively from the University of Tokyo and the Dr. Eng. in 1988 from Osaka University. In 1968 he joined the Electrotechnical Laboratory where he worked on compiler design, optimum searching algorithms, classification and basis enumeration of many-valued logical functions, and optimization of decision trees. In 1992 he moved to the Tsukuba University of Technology as a professor and belonged to the Department of Computer Science. Currently he is an emeritus professor at the university. His current interest includes structure of clones of many-valued logic functions, decision table and Boolean complexity.