Research Article

A wearable assistive system for the visually impaired using object detection, distance measurement and tactile presentation

Yiwen Chen[1], Junjie Shen[1], Hideyuki Sawada[2] (ID)

[1]Department of Pure and Applied Physics, Graduate School of Advanced Science and Engineering, Waseda University, Tokyo 169-8555, Japan.
[2]Department of Applied Physics, Faculty of Science and Engineering, Waseda University, Tokyo 169-8555, Japan.

Correspondence to: Prof. Hideyuki Sawada, Department of Applied Physics, Faculty of Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan. E-mail: sawada@ waseda.jp

Abstract

With the current development of society, ensuring traffic and walking safety for the visually impaired is becoming increasingly important. We propose a wearable system based on a system previously developed by us that uses object recognition, a distance measurement function, and the corresponding vibration pattern presentation to support the mobility of the visually impaired. The system recognizes obstacles in front of a user in real time, measures their distances, processes the information, and then presents safety actions through vibration patterns from a tactile glove woven with shape memory alloy (SMA) actuators. The deep learning model is compressed to achieve real-time recognition using a microcomputer while maintaining recognition accuracy. Measurements of the distances to multiple objects are realized using a stereo camera, and vibration patterns are presented through a tactile glove in response to these distances. Experiments are conducted to verify the system performance to provide safe navigation depending on the positions and the distances of multiple obstacles in front of the user.

Keywords: SMA, tactile display, wearable device, visually impaired, object detection, distance measurement, deep learning, model compression

1. INTRODUCTION

There are 36 million visually impaired people in the world, according to a report from BBC News Japan in 2017[1]. The number is anticipated to triple by 2050 due to our aging society. The percentage of people who are blind or have difficulty seeing is exceptionally high in South Asia and sub-Saharan Africa, making life even more inconvenient in developing countries. There are also 300,000 visually impaired people in Japan; thus, it is important to make the daily lives of the visually impaired easier through studies of assistive technology.

For example, Dionisi *et al*. developed a wearable object detection system to help the visually impaired using Radio Frequency Identification(RFID)[2]. The system detects tagged objects, and the strength of the signal is used to estimate its direction and distance from the user. Tags are detected at home easily; however, in the case of using the system in an outside environment, it is difficult to detect tags and for computer devices to execute the system.

Wagner *et al*. developed a tactile display by using metal pins and servomotors[3]. Each pin has a diameter of 1mm and is arranged in a 6 × 6 array. The display presents the tactile information of small objects since the metal pins are able to present vibratory information with frequencies up to 25 Hz. Servomotors are tightly packed together to achieve spacing between pins of 2 mm, and these pins can be vibrated up and down by rotating the servomotors. The large volume and the large number of motors needed to move the pins, however, make the stack of motors taller and less wearable.

Mukhiddinov *et al*. developed smart glasses for the visually impaired[4]. With these glasses, they employed a deep learning model for object detection, character recognition, and tactile presentation to provide information about the surrounding environment. The system uses a camera, a GPS sensor, and an ultrasonic sensor for information sensing. The deep learning model analyzes the images acquired by the camera to detect objects and characters. The ultrasonic sensor calculates the distance to the object and presents this information through the tactile display installed in the glasses.

Shen *et al*. developed a wearable system using object recognition and SMA actuators[5]. This research realized real-time object recognition by using a compressed YOLO network and presented the information using a small camera and tactile actuators[5]. The tactile presentation allows a user to know which direction to go to avoid an obstacle. In addition, since it was developed as a wearable system that uses Raspberry Pi, it is highly portable. Based on this research, we also consider the importance of providing information regarding the distance from a user to various obstacles through tactile patterns that provide alerts and avoidance information. In our study, we try to develop a new function for distance measurements.

Ashveena *et al*. developed a portable camera-based identification system for visually impaired people[6]. They introduced an algorithm to detect and recognize objects by using the SSD algorithm, a single camera, and an ultrasonic sensor. They introduced voice guidance through earphones as a means of communication with the visually impaired. The advantage of this system is its compactness, meaning that it can be executed with a microcomputer. However, precise distance measurements for multiple obstacles are difficult to carry out due to the limitations of the ultrasonic sensor. In addition, the use of earphones will disrupt the auditory system of users, making it difficult for them to pay attention to their surrounding environment.

In this study, we propose a system that recognizes obstacles in real time and presents tactile information, including the direction of the obstacle and the distance to the obstacle, through a wearable system. We introduce a tactile glove that presents this information through vibratory patterns. The tactile glove presents

different vibration patterns depending on the distance to the obstacle, which it determines by taking the distance measurement using a stereo camera. With this function of measuring the distance to the obstacle and the presentation of the tactile patterns, the system is able to provide better walking support for the visually impaired. To achieve high-performance object recognition and distance measurements while maintaining wearability, we use the deep learning model YOLO[7,8]. For realizing real-time recognition in an onboard system with limited performance, we reduce the number of parameters and layers to slim down the model and increase the inference speed without compromising accuracy. To measure the distances to obstacles, we employ a stereo camera system that uses the parallax of two cameras and implements the distance measurement of multiple objects using the feature point matching method. We also design and fabricate the control circuit to stably control the SMA actuators from the board. Small motors and SMA actuators are woven into the finger and palm parts in the tactile glove to achieve a silent alarm by employing micro-vibration.

2. METHODS

2.1. System configuration

Figure 1 shows the overall structure of the wearable system. The wearable system in this study consists of two major parts. The upper part of the figure shows the real-time object detection part that carries out object recognition and distance measurement using Raspberry Pi and a stereo camera. The lower part presents the tactile presentation part using Raspberry Pi Zero, a signal amplifier circuit, small vibration motors, and SMA actuators. Information acquired from the stereo camera, which consists of two cameras, is processed by the Raspberry Pi, and inference acceleration is carried out by a Neural Compute Stick 2 (NCS2)[9] to calculate the position and distance information of the detected objects. The calculated information is transferred to the Raspberry Pi Zero in the tactile presentation part through TCP communication. Then, according to the acquired position and distance information, a predetermined vibration pattern is transmitted to the tactile display through the signal amplifier circuit, and the tactile stimuli are presented to the user. Figure 2 shows the Raspberry Pi set-up for real-time object detection. A fan was installed to cool the chip, as its prolonged operation could result in overheating.

2.2. Real-time object recognition using compressed YOLO V3 and NCS2

The object identification set-up consists of a Raspberry Pi 3b+, a small camera, an NCS2, and a mobile power supply. One approach to avoid slowing down the inference speed is to adopt a stereo camera with a relatively low resolution (320 × 240). For a longer time operation, a high-capacity battery package powered by two 18,650 batteries (6,800 mAh) is employed to be attached to the Raspberry Pi. For the inference, it was necessary to compress YOLO V3 by reducing the number of parameters and layers due to the limited computing power of the Raspberry Pi.

2.2.1. The compression of YOLO V3

To compress the YOLO V3 model, unnecessary layers and channels in the network were removed by executing four specific steps: regular learning, sparse learning, layer and channel removal, and fine-tuning learning. These steps reduced the number of parameters in the compressed model to approximately 5% of the original structure. The training results and the number of parameters of the model before and after compression are described in the next section.

2.2.2. Acceleration of model inference by NCS2

By considering the mobility and portability of the system, a compact computer, Raspberry Pi 3b+, is employed for system control. Since it is not equipped with a GPU, it lacks the necessary computational power to perform object recognition with large-scale algorithms, such as YOLO V3, in real time. The small
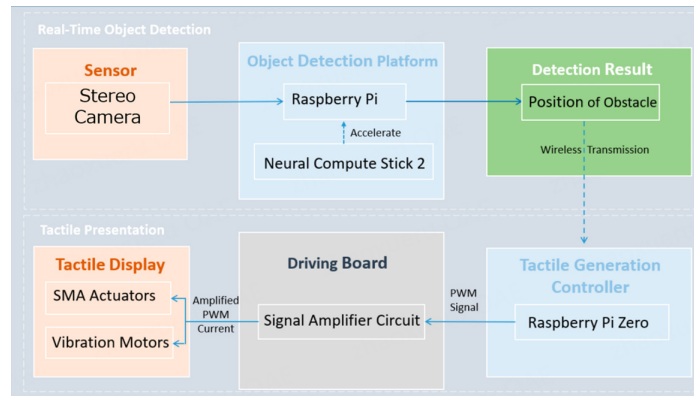
Figure 1. Overall structure of the wearable system.

Figure 2. Raspberry Pi for real-time object detection.

and compact size of the NCS2 is used to achieve edge computations for enhancing the computational power.

2.3. Stereo camera
A stereo camera consisting of two cameras is used to measure the distance to obstacles in real time. The dimension of the stereo camera is shown in Figure 3. To achieve more stable measurements, holes that are spaced 10 cm apart are drilled into a 3D-printed box, and cameras are fixed in the center of the holes. The baseline, which is the distance between cameras, is 10 cm.

2.4. Matching method
Identical object matching is performed to calculate the parallax of each object when multiple objects are found in a picture captured by a camera. In the natural environment, it is assumed that objects frequently enter or leave from the screen of the stereo camera. Thus, it is essential to match them correctly. Specifically, the following procedure is used for matching, and an example of the measurement results of the distances to objects is shown in Figure 4:
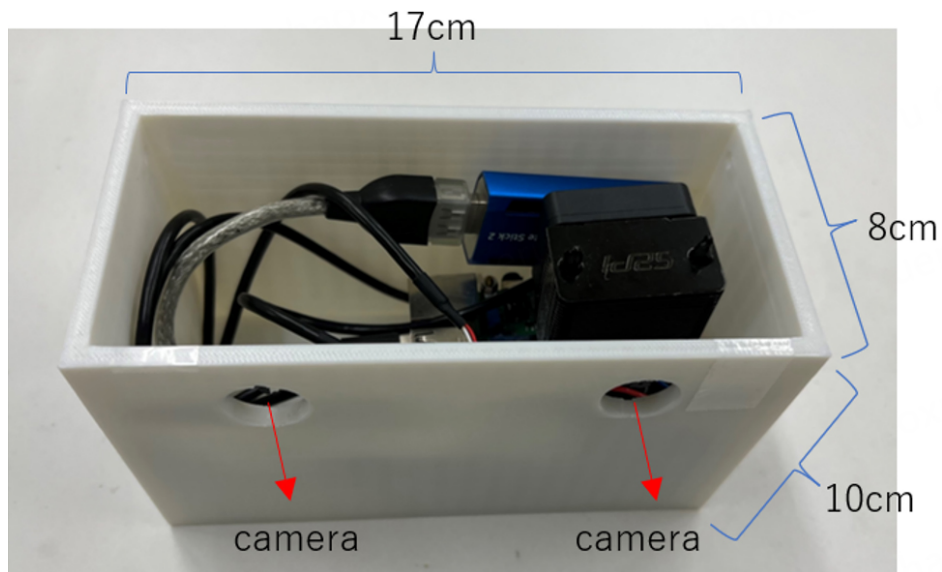
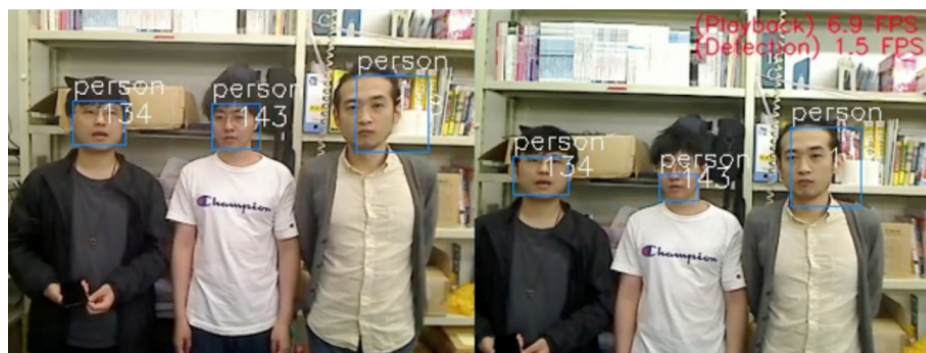**Figure 3.** Box size and stereo camera configuration.

Figure 4. Distance measurement by matching.

1. The number of objects in the two screens of the stereo camera is compared to determine misidentification.

2. Matching is performed from left to right according to the object positions within the screens.

3. The system uses the frame of the recognized object in the left screen as the template image for matching and calculates the similarity with the corresponding object in the right screen via the normalized difference of squares.

4. By using parallax, the distance to each object is calculated.

5. The results are compared with the results from the previous frame. If there is a significant difference in distance, the result is considered to be a recognition error and is ignored.

The normalized square difference, which is calculated using equation (1), is used for matching. The stereo measurement method using parallax is shown in Figure 5.

$$result(x, y) = \frac{\sum_{i,j}(T(i,j) - I(x+i, y+j))^2}{\sqrt{\sum_{i,j} T(i,j)^2}\sqrt{\sum_{i,j} I(x+i, y+j)^2}}$$
$$= \frac{\|\boldsymbol{u} - \boldsymbol{v}\|^2}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}$$

(1)

2.5. Face detection by skin color detection at a distance

Two cameras are separately connected to the Raspberry Pi, and the distance measurement is executed by using the parallax. Considering the calculation capacity, when the camera resolution was set to 320 × 240, the detection of objects far away was difficult. In this study, face detection using skin color is introduced.

Skin color detection is conducted by setting the HSK color space[10-12] and extracting only colors that fall into that range. The procedure of face detection is presented as follows, and an example of the face detection result is shown in Figure 6:

1. The *y*-coordinate with the most consecutive pixels is searched for based on its length, which should be equivalent to the face width.

2. By multiplying its width by the appropriate face ratio, the result is expected to be the length of the face.

3. The face ratio is determined empirically; the ratio of 1.2 works best in this study.

2.6. Vibration of SMA

The tactile actuator consists of a metal pin and an SMA wire with a diameter of 0.075 mm and a length of 5 mm. The structure is shown in Figure 7.

An SMA is a metal composed of Ti-Ni-Cu, which changes its shape when the temperature changes. When the temperature rises to the transformation temperature, the length of the SMA wire shrinks to 95% of its original length. The contraction of the SMA actuator is controlled by the PWM current, which consists of two states in one cycle: ON and OFF. The ON current generates heat inside the wire, by which an SMA wire shrinks. When the pulse current stops in an OFF state, the wire releases heat into the air and returns to its original length[13]. By adjusting the frequency and the duty ratio of the pulse current, different vibration patterns can be created that can be applied to human skin. In our previous studies, we discovered that the SMA actuator is able to generate vibrations with a frequency of up to 500 Hz. We also invented a structure to amplify the micro-contractions of the SMA wire, leading to greater vibrations, as shown in Figure 7B[14,15]. When an electric current is applied, the wire contracts, thereby lifting the pin upward. Then, when the current stops, the wire instantly returns to its original length and the pin also returns to its original position. The contraction and the return to the initial length cause the vibration of the pin to be sufficiently recognized by a user as a tactile sensation.

2.7. Signal amplifier circuit

The Raspberry Pi only provides a voltage of 3.3 V with a maximum current of 80 mA. The SMA wire requires approximately 100 mA to generate the maximum vibration, and thus, a current amplifier is required. The control circuit shown in Figure 8 was originally designed to control the vibration motors and

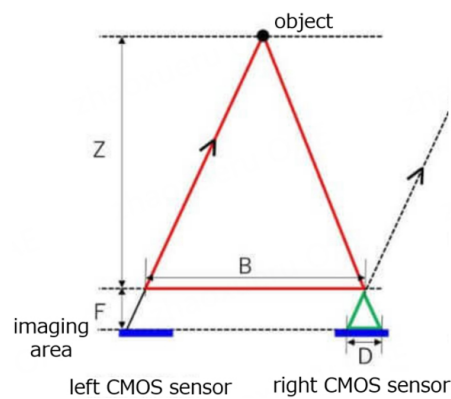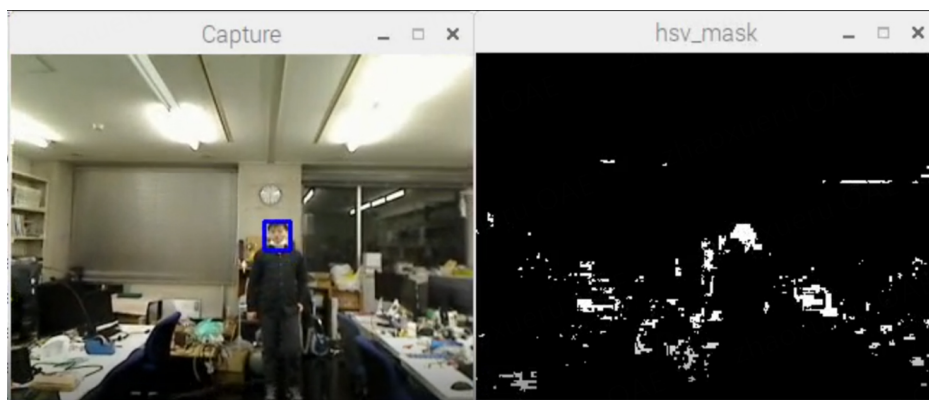**Figure 5.** Distance measurement method using parallax.
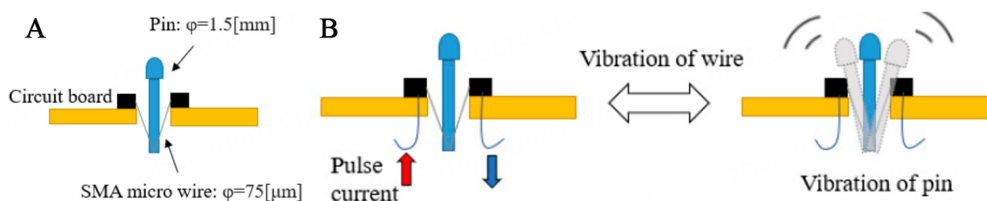

Figure 6. Skin color detection.

Figure 7. Configuration of the tactile actuator. (A) Structure of the tactile actuator; (B) Vibration of SMA wire and tactile pin.

tactile actuators. This circuit has terminals that connect to the vibration motors and SMA wires to drive them independently.

2.8. Tactile display system

The tactile display system consists of a Raspberry Pi Zero, a 9 V battery, a voltage conversion board, a signal amplifier circuit, vibration motors, and tactile actuators, as shown in Figure 9. The SMA actuators and vibration motors are mounted on a circuit board, as shown in the picture to the right. The Raspberry Pi Zero receives the information about obstacles and the distances from the recognition part via TCP
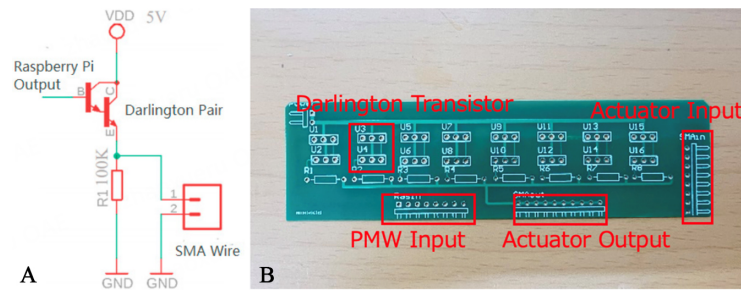
Figure 8. Structure of the signal amplifier circuit. (A) Darlington transistor; (B) Signal amplifier board.
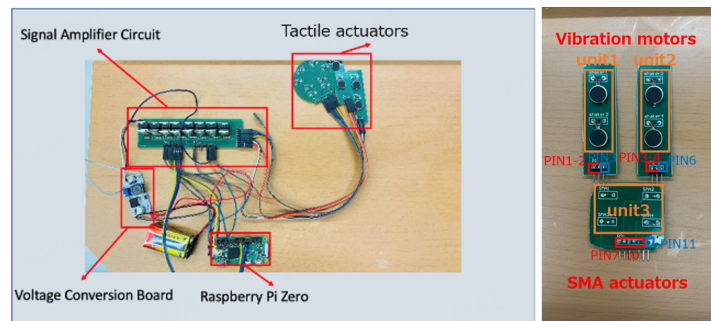

Figure 9. Tactile display system and tactile actuators.

communication and outputs the PWM current via GPIO to properly control the motors and the tactile actuators.

The SMA actuators and vibration motors are stitched into the glove, as shown in Figure 10A. In this tactile display, the SMA actuators present various vibratory stimuli with frequencies up to 300 Hz, which is a unique characteristic when compared with conventional vibration motors. Two vibration motors are stitched in the index and ring fingers of the glove, respectively, and four SMA actuators are stitched in the back side of the hand. By arranging eight actuators in a glove, various tactile patterns are presented by selectively driving SMA actuators with controlled pulse currents that have different frequencies.

All of the control circuits were put into a square box with a size of 16 cm × 10 cm × 5 cm so that during the experiment, the user could carry the whole system, as shown in Figure 10B.

2.9. Presentation of object location and distance through tactile patterns
The recognized object information is presented through tactile patterns. We consider intuitive understanding through tactile sensation by empirically trying to associate the location and distance of an object with different tactile patterns, as shown in Table 1.

The reason why the actuators are woven into gloves to present the vibrations is that a glove is easy to wear, and our fingers and hands are also sensitive to minute vibrations.

2.10. Use of two Raspberry Pis
In this system, two Raspberry Pis are employed, one for the control of the recognition part and the other for the control of the SMA actuators for the tactile presentation. Communication occurs using the socket communication protocol.

Table 1. Vibration pattern

Distance	Location of obstacle	Vibration position	Vibration speed
Less than 4 m	Right of center	Unit 1	2 Hz
More than 2 m	Left of center	Unit 2	2 Hz
Less than 2 m	Right of center	Unit 1	4 Hz
	Left of center	Unit 2	4 Hz
Nothing detected in sight		Unit 3	10 Hz
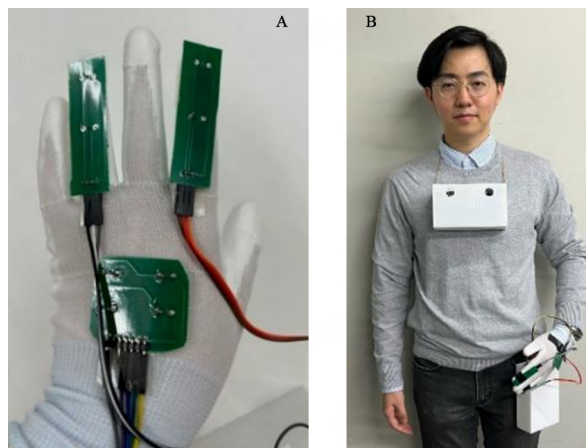
Figure 10. User with the wearable assistive system. (A) Tactile glove with tactile actuators; (B) User with the wearable system.

3. EXPERIMENTS WITH THE CONSTRUCTED WEARABLE ASSISTIVE SYSTEM

3.1. The training and compression of YOLO V3

3.1.1. Initial learning of YOLO V3

We used 8,439 pictures from the COCO dataset in our YOLO V3 training. The COCO dataset is a large-scale object detection, segmentation, and captioning dataset widely used in the image processing field. Of these, 7,551 images were used for training, and 888 were used for testing. As for annotations, the positional information of the frames drawn on the objects in the images is stored in a text file with the same name as the image, each in the form of a square with four coordinates. An example of a training photo is shown in Figure 11.

The training cycles were set to 100, 200, and 300 for the initial training, sparse learning, and fine-tuning, respectively. The learning rate was set to 0.001, and the batch size was set to 4 due to the GPU memory. The gradient descent (SGD) method was employed for learning. The loss value and the accuracy related to the learning results were recorded for the learning evaluation, which is shown in Figure 12.

The learning results show that after 100 training cycles, the average accuracy of the model becomes approximately 91%, and the loss value decreases to less than 0.7. After the initial training, sparse learning was performed to reduce the layers and channels in the model.

3.1.2. Sparse training reduces the number of learning parameters

We tried to reduce the number of parameters of the network as much as possible while maintaining a certain level of accuracy. Spaced learning was used to reduce the layers and channels. Here, we introduce a scaling factor γ to each channel in the model.

$$Loss = \sum_{(x,y)} l(f(x,W),y) + \lambda \sum_{\gamma \in \Gamma} g(\gamma) \tag{2}$$

where (x, y) denotes the train input and target, and W denotes the trainable weights. Function $l(\cdot)$ calculates the difference between the prediction and the correct answer, which is part of the loss. $g(\cdot)$ is a sparsity-induced penalty on the scaling factors, and λ balances the two terms. In the sparse training, we let $g(s) = |s|$.

By using Equation (2) as a loss function, γ asymptotically approaches smaller values. Since the output is the product of γ and the weights, the output of a channel with a γ near 0 will also be nearly 0, and since it contributes little to the calculation, the channel can be eliminated. In this way, the parameters of the entire model can be reduced. Figure 13 shows the accuracy and loss results after sparse learning. The distribution of γ values before and after learning is shown in Figure 14.

From the training results, it can be seen that the accuracy took an initial dive during sparse learning. However, it eventually returned to 80%. Losses increased sharply due to the regularization term but then dropped to 1.6.

From the results of the distribution of γ before and after the training, we can see that because of the sparsity-induced penalty on the scaling factors in Equation 1, the distribution of γ has been reduced to nearly 0. As a result, we will be able to remove redundant channels and layers.

3.1.3. Reduction of layers and channels
Through sparse learning, the channels and layers that could be deleted were identified. In this experiment, the threshold was set to 0.85, and the reduction was performed. A comparison of the number of parameters, accuracy, and inference time before and after deletion is summarized in Table 2.

The results show that although the accuracy was reduced only by 8%, the number of parameters was reduced to about 1.5% of those in the original network and the inference time was reduced to 40% of the original. With the use of NCS2, YOLO V3 was successfully run in Raspberry Pi.

3.1.4. Fine-tuning
After the channel and layer reductions have been completed, another 300 fine-tuning training sessions can be performed to increase the accuracy. The resulting accuracy and losses are shown in Figure 15.

The light blue lines in Figure 14 present the results of fine-tuning, which are compared to those from the initial and sparse learning. The results show that the fine-tuning increased the average accuracy to 85% and reduced the loss to 1.4.

3.2. Experiments
3.2.1 Verification of distance measurement
Since the distance measurement is an essential feature of this research, we experimented with its accuracy. For the experimental conditions, we set six different distances from the camera (0.5 m, 1 m, 1.5 m, 2 m, 3 m, and 4 m), used a person as an obstacle, and tested the extent of difference between the correct distance and the measured distance. The results are summarized in Table 3.

Table 2. Comparison before and after the reduction

	Before	Layer reduction	Channel reduction
Accuracy (mAP)	0.81	0.75	0.73
Parameters	61,523,734	1,424,654	963,752
Inference time (s)	0.014	0.0092	0.0057

Table 3. Accuracy of distance measurement

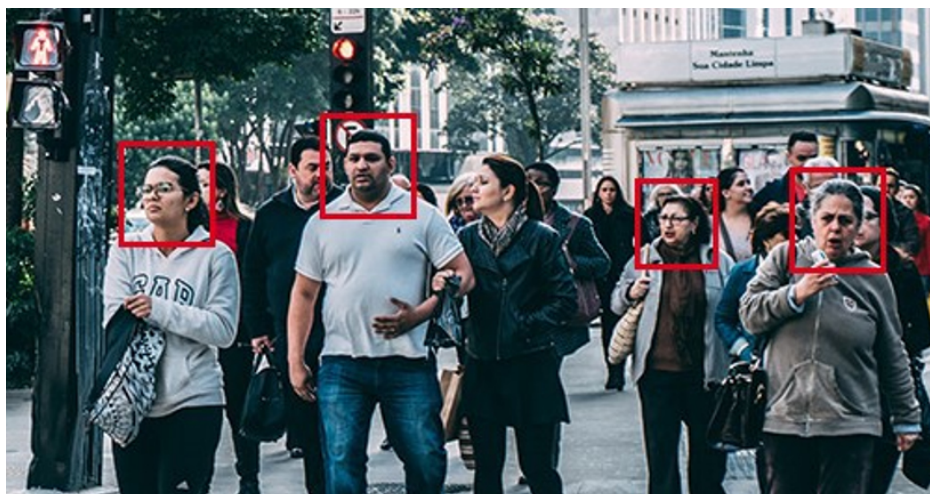|	0.5 m	1 m	1.5 m	2 m	3 m	4 m
1	0.54 m	0.95 m	1.43 m	2.00 m	2.86 m	4.00 m
2	0.53 m	1.06 m	1.43 m	1.82 m	2.86 m	4.00 m
3	0.47 m	0.95 m	1.54 m	1.82 m	3.07 m	4.44 m
4	0.57 m	1.06 m	1.43 m	2.00 m	3.07 m	3.64 m
5	0.57 m	0.91 m	1.54 m	2.00 m	2.86 m	4.00 m
Average error	0.048 m	0.062 m	0.058 m	0.072 m	0.11 m	0.16 m
Standard deviation	0.04	0.07	0.06	0.10	0.12	0.28

Figure 11. An example image from the COCO dataset.
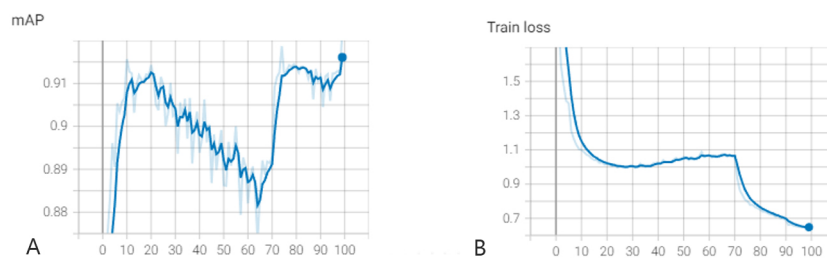

Figure 12. mAP and loss results of initial training. (A) mAP results of initial training; (B) Loss results of initial training.

The results show that within 2 m, the average error is small; however, as the distance increases, the measurement error increases. This is due to the characteristics of the stereo camera, which has a smaller parallax for objects at greater distances, resulting in a lower resolution. Another factor is that the baseline was set as 10 cm to account for the convenience of carrying the camera.
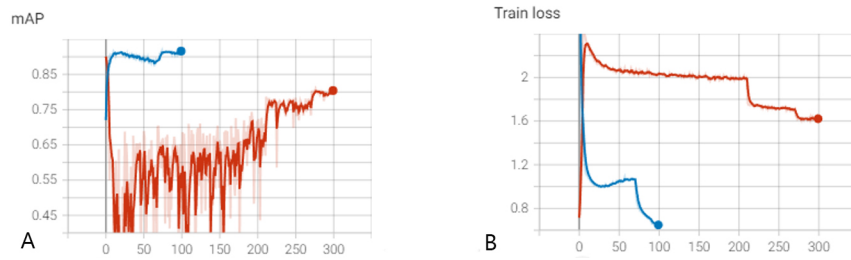
Figure 13. mAP and loss results of sparse training. (A) mAP results of sparse training; (B) Loss results of sparse training.
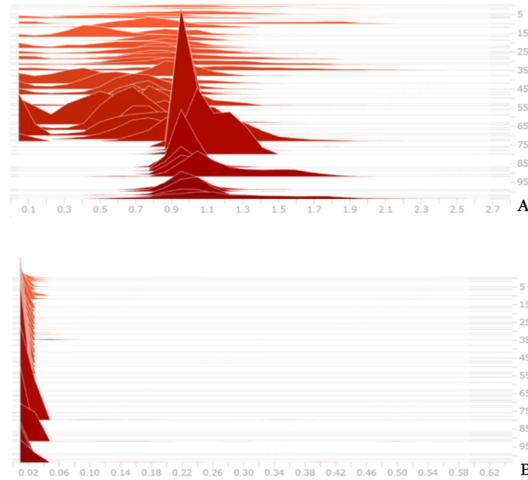

Figure 14. Distribution of γ before and after training. (A) γ values before training; (B) γ values after training.
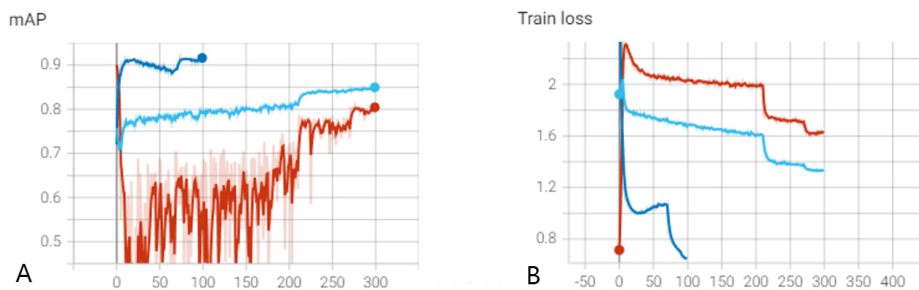

Figure 15. Results of fine-tuning. (A) mAP results of fine-tuning; (B) Loss results of fine-tuning.

3.2.2. Walking experiment using the wearable system

To confirm the system performance, we conducted a walking experiment in a natural environment. We tested whether a user could pass through obstacles only by referring to the vibration patterns presented by the tactile glove. This experiment was divided into two parts as follows:

1. One person is randomly located in front of a subject as an obstacle, and the subject tries to pass by the obstacle without collisions using the tactile glove to convey directional and distance information through vibrations.

2. Three people are randomly situated in front of the subject as obstacles, and the subject tries to pass through this group of people without any collisions.

Furthermore, to confirm the performance of the system in measuring the distance to objects by using the face detection algorithm described in Section 2.5, we conducted two experiments: one that included three people within two meters of the subject and one that included three people that were over two meters away. A schematic figure of each situation is shown in Figure 16.
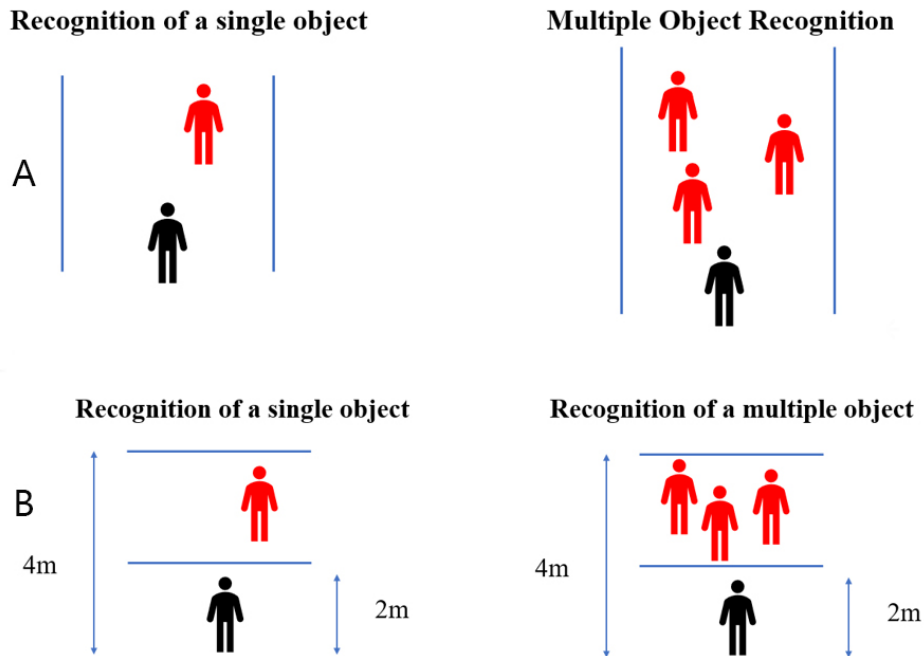
Figure 16. Set-up of walking experiments. (A) Objects situated at a distance less than 2 m away; (B) Objects situated between 2 m and 4 m.

The success rate is calculated for each directional and distance instruction given in these two experiments. For example, if the system gives a left indication when it should turn left, it is counted as a success, and the opposite or no indication is counted as a failure. As for the distance, the system provides a faster vibration if the object is within 2 m, and the system counts it as a success if it produces the correct vibration pattern for each distance. In the case of multiple objects, a success was counted if all of the obstacles were passed in one pass, and a failure was counted if all of the obstacles were not passed. The experiment was conducted 20 times in each case. The results are summarized in Tables 4 and 5. In the obstacle experiment, the estimation time for the entire system was within 0.7 s, which was considered sufficient as a support system for a visually impaired person in consideration of walking speed.

4. DISCUSSION

The results in Tables 4 and 5 show that the successful recognition rate was high for a single object and comparatively low for multiple objects. As explained in Section 2.4, the reason for this is that the matching method was performed for each object on the camera screen to measure the distance based on parallax, which increased the amount of calculations needed when there were multiple objects. The calculations could not keep up with the frame rate. In addition, in regard to the accuracy of the face detection, as the number of objects increased, the overall probability of correct detection also decreased. Moreover, the results when obstacles were within a distance of 2 m were better than those over 2 m away. This was because the detection by YOLO became less accurate in the range of 2 m or more due to resolution issues, and the system switched to using skin color detection.

As for the tactile part, users successfully avoided objects by referring to the presented vibration patterns given by the motors and SMA actuators in the glove. The distance information presented through the various vibration frequencies and their patterns also properly worked. In the experiment, for example, when the left motor vibrated quickly, that indicated that there was an obstacle within a 2-meter distance, and the

Table 4. Walking experiment with a single object

		Direction	Distance
More than 2 m	Number of successes	20	17
	Success rate	100%	85%
Less than 2 m	Number of successes	20	19
	Success rate	100%	95%

Table 5. Walking experiment with multiple objects

	More than 2 m	Less than 2 m
Number of successes	14	16
Success rate	70%	80%

user should move to the left; then, the user immediately moved forward and to the left. When the distance to the obstacle was still more than two meters, the motors vibrated relatively slowly, which could alert the user to prepare for a directional movement. However, if a mistake was made in the recognition part, the wrong vibration pattern was presented, causing the user to fail to pass by the obstacle.

As a future solution, if the performance of Raspberry Pi can be updated to be compatible with a camera with a higher resolution, YOLO alone will be able to detect faces at greater distances, which is expected to increase the overall accuracy and success rate. With more computing power, the matching algorithm for the detection of multiple objects would be faster and smoother.

5. CONCLUSIONS

Focusing on providing mobility assistance to visually impaired people, this study developed a wearable system equipped with a distance measurement function that included its tactile presentation via a corresponding vibration pattern.

Real-time object detection, which is a part of the system, used YOLO V3 models, stereo cameras, matching methods, and skin color detection to achieve obstacle detection. For the tactile display, by employing the vibration characteristics of SMA actuators and also by embedding them in the tactile glove, the signals provided due to real-time object detection were properly converted into vibration patterns to present proper tactile movements for avoiding objects.

Due to limitations in the computing power of Raspberry Pi, however, the detection accuracy and execution speed have yet to reach a satisfactory level. Thus, we are planning to improve the performance accuracy and processing speed through hardware updates in the future.

DECLARATIONS

Funding acquisition, project administration, resources, supervision, writing-review and editing: Sawada H
Software: Chen Y, Shen J
Writing-original draft: Chen Y

REFERENCES

1. Number of visually impaired people to "triple by 2050" worldwide. Available online: https://www.bbc.com/japanese/40810904 [Last accessed on 5 Sep 2023].
2. Dionisi A, Sardini E, Serpelloni M. Wearable object detection system for the blind. 2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings, Graz, Austria; 2012, pp. 1255-8. DOI
3. Wagner CR, Lederman SJ, Howe RD. A tactile shape display using RC servomotors. Proceedings 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. HAPTICS 2002, Orlando, FL, USA; 2002, pp. 354-5. DOI
4. Mukhiddinov M, Cho J. Smart glass system using deep learning for the blind and visually impaired. *Electronics* 2021;10:2756. DOI
5. Shen J, Chen Y, Sawada H. A Wearable assistive device for blind pedestrians using real-time object detection and tactile presentation. *Sensors* 2022;22:4537. DOI PubMed PMC
6. Ashveena A, Bala Deepika J, Mary SP, Nandini DU. Portable camera based identification system for visually impaired people. 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI). Tirunelveli, India; 2023, pp. 1444-50. DOI
7. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016, pp. 779-88. DOI
8. Liu Z, Li J, Shen Z, Huang G, Yan S, Zhang C. Learning efficient convolutional networks through network slimming. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV); 2017, pp. 2755-63. DOI
9. Intel neural compute stick 2 (in Japanese). Available online: https://www.intel.co.jp/content/www/jp/ja/products/sku/140109/intel-neural-compute-stick-2/specifications.html [Last accessed on 5 Sep 2023].
10. Kolkur S, Kalbande D, Shimpi P, Bapat C, Jatakia J. Human skin detection using RGB, HSV and YCbCr color models. Proceedings of the International Conference on Communication and Signal Processing 2016 (ICCASP 2016); 2016, pp. 324-32. DOI
11. Zou L, Li Y. A method of stereo vision matching based on OpenCV. 2010 International Conference on Audio, Language and Image

       Processing; 2010, pp. 185-90. DOI

12.   Extract skin color regions using HSV color space in Python, OpenCV (in Japanese). Available online: https://www.udemy.com/course/computervision_mediapipe/ [Last accessed on 8 Sep 2023].

13.   Jiang C, Uchida K, Sawada H. Research and development of vision based tactile display system using shape memory alloys (in Japanese). *IJICIC* 2014;10:837-50. Available online: https://waseda.elsevierpure.com/ja/publications/research-and-development-of-vision-based-tactile-display-system-u [Last accessed on 8 Sep 2023].

14.   Sawada H, Boonjaipetch P. Tactile pad for the presentation of tactile sensation from moving pictures. In Proceedings of the 2014 7th International Conference on Human System Interactions (HSI). Costa da Caparica, Portugal; 2014, pp. 135-40. DOI

15.   Zhao F, Fukuyama K, Sawada H. Compact Braille display using SMA wire array. RO-MAN 2009- The 18th IEEE International Symposium on Robot and Human Interactive Communication; 2009, pp. 28-33. DOI