Guidance for Visually Impaired Person through Braille Block Detection by Deep Learning

Dong-seok Lee[1], Seung-hu Kim[2], Soon-kak Kwon[2*]

Abstract

In this paper, we propose a passage guidance method for the visually impaired person through a braille block detection. The proposed method recognizes the passage information by detecting the braille blocks individually through a neural network based on a captured image. The braille blocks are detected through YOLOv7 which is a state-of-the-art object detection network. Then, the placements of the detected braille blocks are analyzed to find the groups of the straight blocks arranged in a single line and of the dot blocks gathered in a square shape. The passage information is recognized by comparing between the analyzed block placement and predefined placements. Objects in a sidewalk are detected together with the braille block to warn obstacles on the sidewalk. The passage information is guided in voice for the visually impaired person. In simulation results, the proposed method recognized the passage information with about 85% accuracy.

Key Words: Object Detection, Braille Block Detection, Visually Impaired Person Guidance.

I. INTRODUCTION

Braille blocks are installed on a sidewalk to help the visually impaired person walk safely. The type of the braille block is classified into a linear type, which indicates walking in the direction of the line, and a dot type, which indicates stopping and looking around. The visually impaired person taps the braille block with a stick to acquire passage information such as the passage direction. However, it makes the walk speed of the visually impaired person slow. In addition, the obstacles on the sidewalk may be harmful to the visually impaired person waking.

Even though mounting RFID (Radio-Frequency Identification) into the blocks [1-2] can be utilized for detecting the braille blocks, it is unrealistic because the existing sidewalk infrastructure should be replaced. Kuzume et al. [3] propose a detection method of the braille block through the measurement of a staff pressure change. However, it cannot solve the problem of slow walking speed. Image-based methods can detect the braille blocks quickly and accurately at a low cost through deep learning networks. Okamoto et al. [4] proposes a braille block detection method through CNN (Convolutional Neural Network) layers. The captured image is divided into 3×4 grids. It is recognized whether the blocks exist for each grid through the CNN layers. The method helps finding the braille blocks to the visually impaired person by guiding the camera capturing. Kang et al. [5] find the bounding boxes of the braille blocks through CNN layers. The shapes of the blocks are detected by binarizing the detected bounding boxes. The passage information represented by the braille blocks is recognized through the vertex detection of the block shapes.

The detections of specific objects through neural networks are widely used in various fields. The neural networks can detect various objects accurately and quickly. The passage information about the sidewalk can be recognized through the neural network by detecting braille block groups which have specific meaning. However, it is difficult to find the unlearned shape of the braille block group. In addition, it is hard to recognize a braille block group with damaged blocks by the methods. We proposed a recognition method based on the placement of the braille blocks. The braille blocks are individually detected. The meaning of the braille blocks is recognized by analyzing the placement of the braille blocks. The placement-based method can recognize the information of the braille blocks more accurately even though there are some damaged blocks.

In this paper, we propose the recognition method of the braille blocks by object-based detection. Each braille block

is detected by YOLOv7 [6], which is a state-of-the-art object detection network in real time. Then, the passage information such as a walk direction and crosswalk are recognized by analyzing the placement of the detected blocks. In addition, the obstacles are also detected for the walking safety of the visually impaired person. The passage information is guided by a voice for the visually impaired person.

II. RELATED WORKS

2.1. Deep Learning Object Detection Networks

Objects in an image can be detected through a deep learning network with CNN layers. The CNN layer performs a convolution operation with the input image and its 2D kernel. An output to be passed to the next layer is determined through its activation function. Through the CNN layer, various features in the image can be extracted while maintaining the structure of the image. After a noticeable improvement of the object detection performance through AlexNet [7] in 2012, researches on object detection through the deep learning networks have exploded.

AlexNet [7] and VGGNet [8] detect the region and class of a single object through the CNN layers. The detection accuracy increases as more CNN layers. However, deep layers in the network cause difficult network training due to a gradient vanishing that the value of error backpropagation converges to 0. ResNet [9] solves the problem to introduce a skip connection that is a shortcut between the input and output in a layer. The skip connection prevents the overfit of the network due to excessive network training. FPN [10] can detect the object with various sizes through multiple resolution sampling layers. SSFPN [11] utilizes scale-invariant features which are extracted by 3D CNN. Multi-scale objects can also be detected by a super resolution method [12].

It is very difficult to design networks with outputs for detection results about multiple objects because the number of objects to recognize is variable. Instead, the networks divide the input image before the object detection to detect the single object for each divided region. The networks for the multiple objects are classified into two types based on the image segmentation method. One of them obtains the regions by region proposal, which divides the image based on colors, textures, or feature vectors in order that each region has a single object with high probability. It is called a 2-stage detector because the object detection is performed in a region proposal stage and a detection stage. The other network divides the image into regions of equal size rectangles. The network needs only one stage, so then this network is called a 1-stage detector. The detection accuracy of the 1-stage detector is faster than the 2-stage detector due to the region proposal stage in the 2-stage detector. On the other hand, the 2-stage detector is better for the accuracy of the object detection.

R-CNN [13] is an early detector for multiple objects. R-CNN detects the multiple objects through 2-stage, that is the region proposal and the object detection. However, the CNN is applied to only extract the features of the image, while the region proposal and the object detection are not performed through the network. Therefore, the detection speed is very slow. Fast R-CNN [14] and Faster R-CNN [15] improve the detection speed by introducing networks for the object detection and for the region proposal, respectively. Unlike the R-CNN families, YOLO families and SSD are a 1-stage detector. YOLO [16] divides the image into regions of equal size and performs object detection on each region. YOLO does not require the region proposal, so the detection speed is faster than the 2-stage detectors such as Faster R-CNN. YOLO has been revised to improve the detection accuracy and processing time [6, 17-18]. SSD [19] utilizes the features from image scaled to various sizes by a pyramidal feature hierarchy to detect the multi-scaled objects.

III. BRAILLE BLOCK GUIDANCE FOR VISUALLY IMPAIRED PERSON

3.1. Braille Block Detection

The braille blocks are classified into the following three types: a straight block in front direction, a straight block in side direction, and a dot block. Fig. 1 shows the types of the braille block.

YOLOv7 [6] is a network for detecting the braille blocks. YOLOv7 can detect multi-objects in real time. We train the detection network through the images of sidewalks including the braille blocks in AI HUB [20]. The braille blocks in the training images are labeled and are trained into YOLO. Obstacle objects such as scooters and bicycles in the images are also trained in order to warn them on the sidewalk during walking. The number of the images for the network training is 200. The training images have about 1,635 braille blocks and 48 obstacle objects. 80% of the training images are utilized for the network training and the others validate the training. In the network training, the number of epochs and the learning rate are 500 and 1×10^{-3}, respectively, and
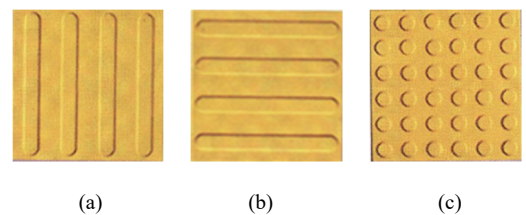

(a)　　　　　　(b)　　　　　　(c)

Fig. 1. Types of braille block. (a) Straight block in front direction, (b) straight block in side direction, and (c) dot block.

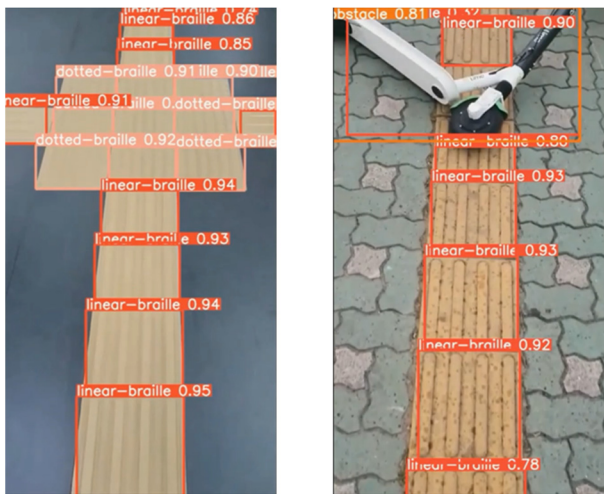Fig. 2. Sample of training image with braille blocks.

Fig. 3. Braille block detection by object detection network.

an input image size is 640×640. Fig. 2 shows the samples of the images for network training.

Fig. 3 shows the detection results. The braille blocks and the obstacles are detected well for each object.

3.2. Passage Information Recognition

In order to provide the passage information of the sidewalk for the visually impaired person, the braille blocks, whose types are the dot and the straight, are placed in special shapes. The passage information according to the placement of the braille blocks are as follows: three or more consecutive straight blocks is a straight section; a set of the consecutive dot blocks is a stop section; a placement of the

dot blocks in two lines horizontally after the consecutive straight blocks vertically means that the crosswalk is in front; a placement of sets of horizontal and vertical straight blocks at right angles is a turn section. Fig. 4 shows the passage information according to the placement of the braille blocks.

A straight section is recognized by finding 5 or more straight blocks in a single line. Consecutive blocks arranged in a single line have the same x coordinate or the same y coordinate of their top-left points if the capture and the movement directions are equal. If the image is tilted, the coordinates are not equal. However, the bounding boxes of the blocks overlap as shown in Fig. 5(b). Therefore, the straight section can be recognized by detecting 5 or more overlapping bounding boxes of the straight blocks in a vertical direction. The stop section is recognized by detecting
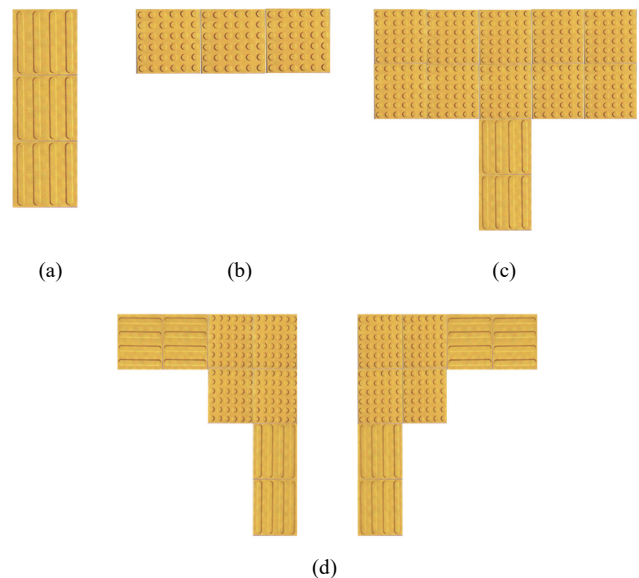

Fig. 4. Passage information according to placement of braille blocks. (a) Straight section, (b) stop section, (c) crosswalk section, and (d) turn sections.
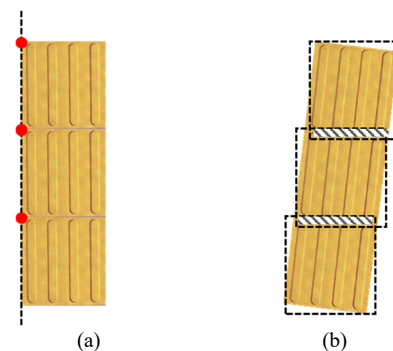

Fig. 5. Detection of consecutive blocks. (a) Not-tilted image and (b) tilted image.
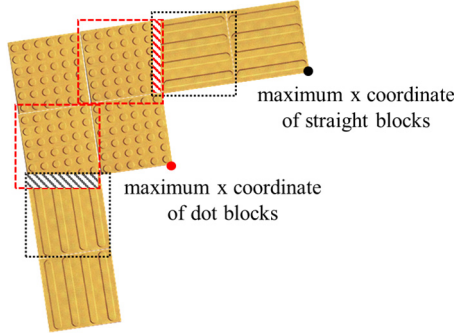
Fig. 6. Right turn section recognition.

3 or more overlapping bounding boxes of the dot blocks in a horizontal direction. The crosswalk section is recognized by detecting dot blocks in double lines, that is by detecting 5 or more bounding boxes of the braille blocks overlapping in both of the vertical and the horizontal directions. The turn section is recognized by detecting the overlapping bounding boxes between the dot and the straight blocks in a horizontal direction and a vertical direction as shown in Fig. 6. The turn direction is right in case that the maximum x coordinate among the bottom-right points of the straight blocks is larger than the maximum coordinate among the dot blocks.

3.3. Passage Information Guidance for Visually Impaired Person

In order to guide the passage information for the visually impaired person by detecting the braille block, images are captured through a camera mounted on a smartphone that the people always have. The captured image is transmitted to a server device having high-performance computing power for the real-time detection. The recognized passage information is guided by voice for the visually impaired person after receiving the recognition result from the server. If the obstacles are detected, the warning information about the obstacles is guided preferentially over the passage information. Fig. 7 shows the voice guidance of the passage information about the sidewalk.

The passage information may be very early guided due to far braille blocks. For the guidance at a right time, the regions of a 1/5 height at top and bottom of the captured
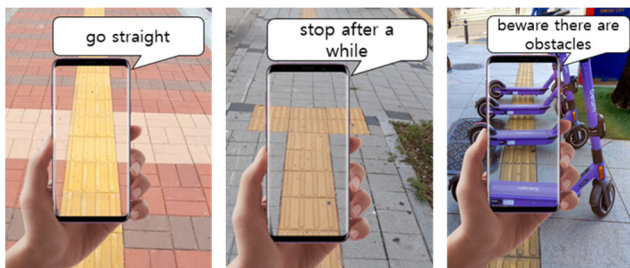

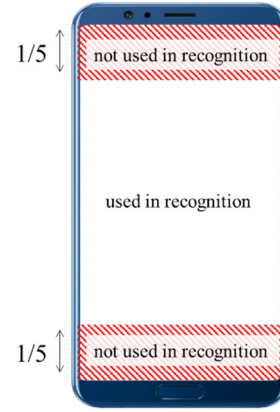Fig. 7. Voice guidance of passage information for visually impaired person.

Fig. 8. Recognition region for guidance at right time.

image are not used to recognition, respectively as shown in Fig. 8.

IV. SIMULATION RESULTS

The accuracies of the braille block detection and the passage information recognition are measured. We capture the sidewalk including the braille blocks which indicate the sections of only straight, of only turn, of combined turn and straight, of a crossroad, and of a crosswalk. Similar to actual capture situations for the sidewalk, the sidewalk videos for the simulation are captured with 25 degrees, 30 degrees, and 35 degrees angles relative to the ground as shown in Fig. 9. First 60 frames of the captured videos are utilized for the simulations. Table 1 shows the number of the simulation videos for the cases.

Table 2 shows the accuracy of the braille block detection.

Table 1. Number of simulation videos.

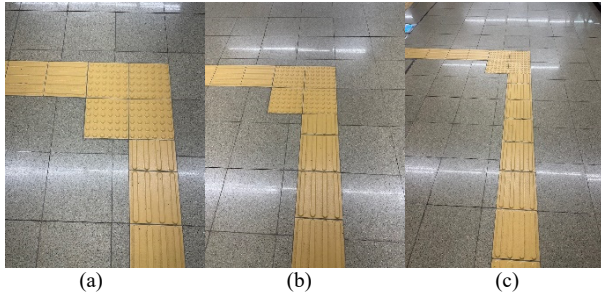| Section type	Capture angles	Number of videos
Only straight	25°	15
	30°	15
	35°	15
Only turn	25°	10
	30°	10
	35°	10
Combined straight and turn	25°	10
	30°	10
	35°	10
Crossroad	25°	10
	30°	10
	35°	10
Crosswalk	25°	5
	30°	5
	35°	5

Fig. 9. Videos for simulation according to capture angle. (a) 25 degrees, (b) 30 degrees, and (c) 35 degrees.

Table 2. Accuracy of braille block detection.

Capture angles	No. blocks	Correct detection	False detection	Not detected
25°	431	414	15	2
30°	532	505	22	5
35°	641	588	38	15

The braille blocks are almost accurately detected in case of the videos with the low capture angles. As the capture angle is larger, the far blocks are sometimes not detected or their types are incorrectly detected.

The accuracy of the passage information recognition is shown in Table 3. The passage direction is almost recognized accurately for 25 degrees of the capture angle. When the capture angle is 30 or 35 degrees, the straight and crosswalk sections are well detected, but the recognition of the turn sections are sometimes wrong.

Table 3. Accuracy of passage direction recognition.

Section type	Capture angles	No. videos	Correct recognition	False recognition
Only straight	25°	15	15	0
	30°	15	15	0
	35°	15	15	0
Only turn	25°	10	10	0
	30°	10	10	0
	35°	10	8	2
Combined straight and turn	25°	10	10	0
	30°	10	7	3
	35°	10	5	5
Crossroad	25°	10	9	1
	30°	10	6	4
	35°	10	4	6
Crosswalk	25°	5	5	0
	30°	5	4	1
	35°	5	4	1

Table 4. Accuracy of obstacle object detection.

Number of obstacles	Correct detection	Not detected
43	39	4

Table 4 shows the result of the obstacle object detection. The obstacle objects are accurately detected. However, the unlearned object cannot be detected. In order to solve this problem, more various types of obstacle objects need to be trained in the network.

V. CONCLUSION

In this paper, we proposed the recognition method of the passage information about the sidewalk by detecting the braille block through deep learning. The braille blocks and the obstacles were individually detected by YOLOv7. The passage information about the sidewalk was recognized by analyzing the placement of the braille blocks. Either the passage information or the warning information about the obstacles was guided by voice for the visually impaired person. In the simulation results, the accuracy of the passage information recognition was about 85%. The proposed method may be helped for not only the guidance for the visually impaired person but also autonomous driving of robots on the sidewalk.

ACKNOWLEDGMENT

REFERENCES

[1] M. Murad, A. Rehman, A. A. Shah, S. Ullah, M. Fahad, and K. M. Yahya, "RFAIDE — An RFID based navigation and object recognition assistant for visually impaired people," in *Proceeding of the International Conference on Emerging Technologies*, 2011, pp. 1-4.

[2] S. Chumkamon, P. Tuvaphanthaphiphat, and P. Keeratiwintakorn, "A blind navigation system using rfid for indoor environments," in *Proceeding of the International Conference on Electrical Engineering/ Electronics, Computer, Telecommunications and Information Technology*, 2008, pp. 765-768.

[3] K. Kuzume, Y. Watanabe, H. Masuda, and T. Masuzaki, "Inference system for automatic identification of braille

blocks using a pressure sensor array," in *Proceeding of the IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events*, 2022, pp. 46-49.

[4] T. Okamoto, T. Shimono, Y. Tsuboi, M. Izumi, and Y. Takano, "Braille block recognition using convolutional neural network and guide for visually impaired people," in *Proceeding of the International Symposium on Industrial Electronics*, 2020, pp. 487-492.

[5] J. K. Kang, V. Bajeneza, S. Y. Ahn, M. W. Sung, and Y. S. Lee, "A method to enhance the accuracy of braille block recognition for walking assistance of the visually impaired: Use of YOLOv5 and analysis of vertex coordinates," *Journal of KIISE*, vol. 49, no. 4, pp. 291-297, 2022.

[6] C. Y. Wang, A. Bochkovskiy, and H, Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *arXiv:2207.02696*, 2022.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceeding of the International Conference on Learning Representations*, 2015, pp. 1-14.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.

[10] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117-2125.

[11] H. J. Park, Y. J. Choi, Y. W. Lee, and B. G. Kim, "ssFPN: Scale sequence (S^2) feature based-feature pyramid network for object detection," *arXiv:2208.11533*, 2022.

[12] Y. H. Lee, D. S. Jun, B. G. Kim, and H. J. Lee, "Enhanced single image super resolution method using lightweight multi-scale channel dense network," *Sensors*, vol. 21, no. 10, pp. 1-17, 2021.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceeding of the Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580-587.

[14] R. Girshick, "Fast R-CNN," in *Proceedings of the International Conference on Computer Vision*, 2015, pp. 1440-1448.

[15] R. Shaoqing, H. Kaiming, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, 2017, pp. 1137-1149.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779-788.

[17] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv:2004.10934*, 2020.

[18] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, and L. Li, et al., "YOLOv6: A single-stage object detection framework for industrial applications," *arXiv:2209.02976*, 2022.

[19] A. Kumar, Z. J. Zhang, and H. Lyu, "Object detection in real time based on improved single shot multi-box detector algorithm," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 204, pp. 1-18, 2020.

[20] AI HUB, Sidewalk Video, https://www.aihub.or.

AUTHORS

Dong-seok Lee received the B.S., M.S., and Ph.D. degrees in Computer Software Engineering from Dong-eui University in 2015, 2017, and 2021, respectively, and is currently a research professor in AI Grand ICT Research Center at Dong-eui University. His research interest is in the areas of image processing and video processing.

Seung-hu Kim is currently an undergraduate student in the Department of Computer Software Engineering at Dong-eui University. His research interest is in the areas of image recognition.

Soon-kak Kwon received the B.S. degree in Electronic Engineering from Kyungpook National University, in 1990, the M.S. and Ph.D. degrees in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), in 1992 and 1998, respectively. From 1998 to 2000, he was a team manager at Technology Appraisal Center of Korea Technology Guarantee Fund. Since 2001, he has been a faculty member of Dong-eui University, where he is now a professor in the Department of Computer Software Engineering. From 2003 to 2004, he was a visiting professor of the Department of Electrical Engineering in the University of Texas at Arlington. From 2010 to 2011, he was an international visiting research associate in the School of Engineering and Advanced Technology in Massey University. Prof. Kwon received the awards, Leading Engineers of the World 2008 and Foremost Engineers of the World 2008, from IBC, and best papers from Korea Multimedia Society, respectively. His biographical profile has been included in the 2008~2014, 2017~2019 Editions of Marquis Who's Who in the World and the 2009/2010 Edition of IBC Outstanding 2000 Intellectuals of the 21st Century. He is an associate editor for IEICE Nolta journal, a topic editor for MDPI Electronics journal, and a reviewer board member for MDPI Signals journal. Also he is working as a reviewer for several journals such as Sensors, Applied Sciences, Information, Symmetry, Entropy, IEEE TCSVT, and IEEE Access. His research interests are in the areas of image processing, video processing, video transmission, depth data processing, and AI object recognition.