Development of a Safe Walking Assistance System for Visually Impaired Persons Using MY VISION — Estimation of a Safe Passage from Sidewalk Information Based on Transfer Learning of VGG-16 Network

Takumi Yokote
Graduate School of Engineering, Kyushu Institute of Technology,
Kitakyushu, Fukuoka 804-8550, Japan
E-mail: yokote.takumi165@mail.kyutech.jp

Joo Kooi Tan
Faculty of Engineering, Kyushu Institute of Technology
Sensui-cho Tobata-ku
Kitakyushu, 804-8550, Japan
E-mail: etheltan@cntl.kyutech.ac.jp

Abstract

In recent years, the number of visually impaired persons has been increasing year by year, and outdoor accidents have also been increasing when they go out. It is difficult to detect hazards on sidewalks even with a currently popular technique, such as a semantic segmentation technique or YOLO, because sidewalk situations are complicated and change frequently. For this reason, we propose a method of recognizing sidewalk situations from a self-viewpoint video called MY VISION. Conventional methods detect objects surrounding the sidewalk by learning the objects' features beforehand and guiding visually impaired persons according to the position/direction of the detected object. The proposed method neither learns objects nor detects objects. We focus on sidewalk situations and use a multi-class classification technique based on transfer learning of VGG-16 to guide visually impaired persons' walk according to three kinds of sidewalk information to ensure more safety. The effectiveness of the proposed method was confirmed by experiments.

Keywords: Safe walking assistance, deep learning, visually impaired, MY VISION.

1. Introduction

According to a survey conducted by the Ministry of Health, Labour and Welfare in 2008, the number of visually impaired persons was 312,000 [1]. Regarding accidents involving visually impaired people on the street, a 2003 survey showed that approximately 42% of all respondents had experienced a walking accident involving injury while walking outdoors within past five years [2]. This indicates that walking outdoors is very dangerous for the visually impaired.

Currently, white canes, Braille blocks, and guide dogs are widely used to help the visually impaired recognize their surroundings. However, each of these methods has its own problems. This indicates that there is a need for a method that allows visually impaired people to explore their surroundings easily and extensively.

Therefore, we propose a method to recognize a surrounding environment from MY VISION by multi-class classification using deep learning. The MY VISION in this paper is based on a chest-mounted camera image. When the proposed method is completed, the visually impaired person will be able to know danger earlier. This method also guides the user to the center of the sidewalk, thus enabling the user to pass through a path with fewer hazards. The above is the significance of this research.

Previous studies include obstacle detection by object and area recognition using RGB-D sensors [3], and methods to recognize people and obstacles using YOLO and Faster-RCNN, that are excellent algorithms for object detection [4][5].

However, even if these methods produce objects and walkable areas as one task, instructions for a visually impaired person to avoid obstacles must be learned and processed as a separate task.

In this paper, by using instruction-based multiclass classification to understand the situation on the sidewalk, the learning and processing from the input to the instructions for the visually impaired person to walk can be performed as a single task. Another unique aspect of this research is that it uses images of the left and right edges of the sidewalk as well as the center of the sidewalk as a dataset, taking into account the actual situation in which a visually impaired person is walking.

2. Method

2.1 *Deep Learning model*

2.1.1 *Transfer learning*

In order to create a deep learning model for this method, we perform fine-tuning of the VGG16 model [6].

2.1.2 *VGG16 model*

VGG16 is the second-ranked model in ILSVRC in 2014. It uses ImageNet as its training dataset and classifies 1000 categories.

This method uses a model in which the output layer of this model is modified. The model is shown in Figure 1.

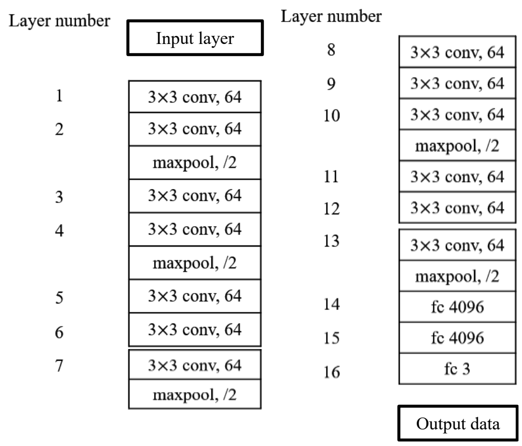

Fig.1 Structure of the model for gait instruction classification

2.2 Learning

2.2.1 *Loss function*

In this method, cross entropy is used as a loss function. Cross-entropy is a loss function for class classification, where n is the sample number, K is the number of classes, y_{nk} is the output data of the n th sample class k expressed as a probability, and t_{nk} is the teacher data of the n th sample class k expressed as 0 or 1, the error of each data is given by the following formula.

$$E_n = -\sum_{k=1}^{K} t_{nk} \log y_{nk} \tag{1}$$

When the total number of samples is N, the overall error is

$$E = \sum_{n=1}^{N} E_n \tag{2}$$

The more the output data and the teacher data match, the smaller the value.

2.2.2 Parameter optimization methods

In this study, Adam is used as a parameter optimization method. Adam's method is described below. First, let $m_{i,t}$ be the exponential moving average of the first-order moment of the gradient and $v_{i,t}$ be the exponential moving average of the second-order moment of the gradient, defined as follows.

$$m_{i,t} = \rho_1 m_{i,t-1} + (1-\rho_1)\nabla E(\boldsymbol{w}^{(t)})_i \tag{3}$$

$$v_{i,t} = \rho_2 v_{i,t-1} + (1-\rho_2)(\nabla E(\boldsymbol{w}^{(t)})_i)^2 \tag{4}$$

where the initial values are $m_{i,t} = v_{i,t} = 0$. In addition, ρ_1 and ρ_2 are decay rates, and $\nabla E(\boldsymbol{w}^{(t)})_i$ is the i th component of the gradient. However, since this equation takes an initial value of 0, the moment estimates are biased toward 0 at the beginning of the update. The moment estimates corrected for this bias are shown in the following equation.

$$\widehat{m}_{i,t} = \frac{m_{i,t}}{(1-(\rho_1)^t)} \tag{5}$$

$$\widehat{v}_{i,t} = \frac{v_{i,t}}{(1-(\rho_2)^t)} \tag{6}$$

The gradient descent method uses these estimates, and the equation becomes Eq. (7).

$$\Delta \boldsymbol{w_i}^{(t)} = -\eta \frac{\widehat{m}_{i,t}}{\sqrt{\widehat{v}_{i,t}+\epsilon}} \tag{7}$$

Here ϵ is a very small number that is introduced so that the denominator is not zero.

2.3 Multiclass classification

In this method, the gait instruction classification is multi-class classification into three classes. Multi-class classification is the process of sorting data into several categories. The following is the description of the training dataset used in this study.

The training dataset used in this study consists of 15 sidewalks in different locations, and at each location, three images were taken: an image of a person walking on the left edge of the sidewalk, an image of a person walking on the right edge of the sidewalk, and an image of a person walking in the center of the sidewalk. The images were then classified into three classes: "Go", "Avoid to the right" and "Avoid to the left.

The following is an explanation of the situations and obstacles assumed for each class. First, the obstacles are people and bicycles within 5 meters in front of the vehicle. The "Go" class is for situations where there are no obstacles on either side, such as in the center of a sidewalk, and there are no obstacles in front of you to stop. The "Avoid to the right" class is for situations where there is an obstacle on the left, such as when walking on the left side of the sidewalk, or when there is an obstacle in front of you and the sidewalk is visible to the right of the obstacle, making it safe. Finally, the "Avoid to the left" class is for the cases where there is an obstacle to the right, when walking on the right side of the sidewalk, or when the obstacle is in front of you and the sidewalk is visible and safe to the left of the obstacle.

3. Experiments and Results

3.1 Experiments

The number of datasets are shown in Table 1, and the wearers are shown in Figure 2. Image size is 224 x 224[pixel]. Precision was used as the evaluation function.

Table1 Number of data sets

Dataset	Learning[frame]	Test[frame]
Go	1529	505
Left	1622	421
Right	1656	405
Total	4807	1331

Fig.2 The wearer

3.2 Results

The accuracy for each class is shown in Table 2 shows how each correct answer label was estimated for each class.

Table2 Number of correct answers and Precision per correct answer label

	Go	Left	Right	Total
number of correct answers [frame]	494	276	306	1076
Precision	0.978	0.656	0.756	0.808

4 Discussion

First, Table 2 shows that the correct response rate is high for images with the correct label "Go", while the rate is low for images with the correct label "Avoid to the right" or "Avoid to the left. First, the high rate of "Go" is considered to be due to the fact that the basic "Go" label is applied when there are no obstacles on the left, right, or nearby, making it easier to make a decision. Second, the low rate of correct responses for the other two images may be due to the fact that, as shown in Figure 3, both the sidewalk and the roadway are shown in the image, and the roadway is misidentified as the sidewalk, and second, as shown in Figure 4, a passerby is in the center of the image, making it difficult to choose whether to avoid the right or left side of the road.

Fig.3 Example of mistaking a road for a sidewalk

Fig.4 Examples that are difficult to discern which way to avoid

5 Conclusion

In this study, we proposed a walking instruction classification method based on multi-class classification using deep learning. The proposed method was trained on a dataset of labeled left-most, center, and right-most images taken at 15 locations using MY VISION. Using the model created by the training, the images were classified into three multi-class categories: "Go", "Avoid left" and "Avoid right".

Experiments were conducted to verify the effectiveness of the proposed method.

Test images taken at different locations from the training images were classified into multiple classes, and the percentage of correct answers and the breakdown of the estimated labels for the test images were examined. The results showed that the correct answer rate for the label "Go" was 97.8[%], the correct answer rate for the label "Avoid to the left" was 65.6[%], the correct answer rate for the label "Avoid to the right" was 75.6[%], and the average of the correct answers for the three labels was 80.8[%].

Future issues include clarification of the classification criteria and the distinction between people and objects. In order to clarify the classification criteria, it is necessary to create a small region in the image and judge whether or not there is an object in the region. In addition, to distinguish between people and objects, it is necessary to take video images and judge people based on the motion between two adjacent frames.

References

1. Ministry of Health, Labour and Welfare, Social and Support Bureau, Disability Health and Welfare Department: "Summary of the Results of the 2016 Survey on Living Conditions, etc. (National Survey of Children and Persons with Disabilities at Home)," pp1-14, 2018. (In Japanese)
2. N. Abe and N. Hashimoto: "Walking accidents National survey to Maintain Visually Handicapped Persons Walking Environment", pp.81-92, 2005. (In Japanese)
3. A. Aladren, G. Lopez-Nicolas, L. Puig, J. J. Guerrero: "Navigation Assistance for the Visually Impaired Using RGB-D Sensor with Range Expansion", IEEE Systems Journal, pp922-932, 2016.
4. O. Masurekar, O. Jadhav, P. Kulkarni, S. Patil: "Real Time Object Detection Using YOLOv3", International Research Journal of Engineering and Technology (IRJET), pp3764-3768, 2020.
5. S. Ren, K. He, R. Girshick, J. Sun: "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp.1137-1149, 2016.
6. K. Simonyan, A. Zisserman: "Very Deep Convolutional Networks for Large-Scale Image Recognition", International Conference on Learning Representations, 2015.

Authors Introduction

Mr. Takumi Yokote

He received his B.E. degree in Department of Control Engineering in 2017 from the Faculty of Engineering, Kyushu Institute of Technology in Japan. He is acquiring the M.E. in the same University. His research interests are development of a safe walking assistance system for visually impaired persons using MY VISION, estimation of a safe passage transfer learning, and neural network.

Prof. Dr. Joo Kooi Tan

Prof. Tan is currently with the Department of Mechanical and Control Engineering, Kyushu Institute of Technology, as a Professor. 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 SICE Kyushu Branch Young Author's Award in 1999, the AROB Young Author's Award in 2004, the Young Author's Award from IPSJ of Kyushu Branch in 2004, and the BMFSA Best Paper Award in 2008, 2010, 2013 and 2015. She is a member of IEEE, The Information Processing Society, The Institute of Electronics, Information and Communication Engineers, and The Biomedical Fuzzy Systems Association of Japan.