Detecting a Pedestrian's Walk Direction Using MY VISION for Supporting Safe Walk of a Visually Impaired Person

Shinya Iizumi*
Graduate of Engineering, Kyushu Institute of Technology, 1-1 Sensui-cho, Tobata, Kitakyushu Fukuoka, 804-8550, Japan

Joo Kooi Tan*, Yuta Ono, Seiji Ishikawa, Masuhiro Nitta
Faculty of Engineering, Kyushu Institute of Technology,, 1-1 Sensui-cho, Tobata, Kitakyushu Fukuoka, 804-8550, Japan
E-mail:{iizumi-shinya, etheltan}@ss10.cntl.kyutech.ac.jp
https://www.kyutech.ac.jp/~etheltan/

Abstract

In this paper, we propose a method of recognizing multiple objects using MSC-HOG (Multiple-Scale-Cell Histograms of Oriented Gradients) features and intensity models of both pedestrians and bicyclists. We also propose a method of detecting approaching passersby using different discriminators without using time-series information such as Optical Flow. The effectiveness of the proposed method is verified by experiments.

Keywords: MSC-HOG, Passerby Detection, MY VISION, Visually impaired person, Walk support

1. Introduction

"White canes" and "guide dogs" are two main tools for helping a visually impaired person walk safely[1]. However, the former has a narrow range in recognizing surrounding objects, and the latter has the problem of insufficient numbers of dogs. Therefore, a system has been proposed called MY VISION[2] to recognize surrounding pedestrians and their approaching directions, and give an alarm to the user if pedestrian approaches him/her, thereby it is possible for the user avoiding collision. However, since the MSC-HOG (Multiple-Scale-Cell Histograms of Oriented Gradients) features [3, 6] used in the literature[2] consider only a single model image (a pedestrian model), there is a problem that only one type of object (e.g. a pedestrian) can be recognized. Another problem is that, since they use an optical flow technique to detect a pedestrian, the recognition rate of approaching directions is low.

In this paper, we propose a method of recognizing multiple objects using MSC-HOG features using the models of both pedestrians and bicyclists. In addition, we also propose a method of detecting approaching passersby who are more dangerous to a visually impaired person. We verify the effectiveness of the proposed method using the images of passersby (pedestrians and bicyclist) in our own database, and the pedestrian images in INRIA Person Dataset.

2. Outline of the Proposal Method

The proposed method consists of a learning part and a discrimination part. As shown in **Fig.1**, we define three directions by which the surrounding passersby approach a user i.e., a visually impaired person. They are the front left, the front and the front right direction, expressed by the 10 o'clock, 12 o'clock, and 2 o'clock direction, respectively. In the learning part, we first train a three-class

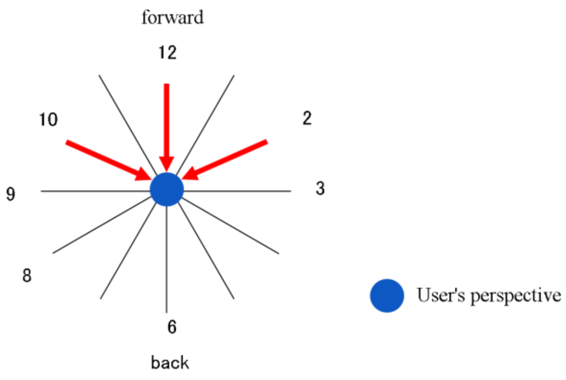


Fig.1 Defining the direction of approaching pedestrians perceived from a visually impaired person.

classifier of passersby approaching from the three directions mentioned above using MSC-HOG features. After known the direction, it is used as a risk index to give alarm to the user. For example, a pedestrian at the 2 o'clock direction walking towards the 10 o'clock direction is not a target of the recognition, whereas, a pedestrian at the 2 o'clock direction walking towards the 8 o'clock direction might be high risk to the user.

By recognizing only those pedestrians approaching from the three specific directions, we can recognize both the pedestrian and the approaching direction at the same time. In the proposed method, we create an average pedestrian model and an average bicyclist model, and then integrating the two models to obtain a model (hereafter referred to as passerby model) according to the three directions. We employ MSC-HOG feature to extract the features from a fed image using the three directional passerby model and train a Random Forest discriminator using the features. Having obtained an effective discriminator of each passerby direction, we apply it to the recognition of a passersby approaching from the three directions.

3. Proposed Method

In this section, some techniques contained in the proposed method are described.

3.1. *MSC-HOG features*

We use the MSC-HOG feature, which is an improved method from the conventional HOG (Histograms of Oriented Gradients). features[4]. In feature extraction, the position and the size of the rectangular cells of HOG are fixed, in contrast, they are variable with the MSC-HOG features.

3.1.1. *Creating a passerby model*

In MSC-HOG feature extraction, it is necessary to create a model of the recognition target in advance for feature extraction. In this paper, we propose a passerby model for recognizing both pedestrians and bicyclists. The MSC-HOG features mainly require the edge information at the cells placed on the target. So we apply Sobel Filter to grayscale images to obtain the edge images. Then, the average of each image is calculated by Eqs.(1)and(2), and a passerby model is created.

$$\overline{I_p}(x,y) = \frac{1}{N}\sum_{i=0}^{N} I_p(x,y) \quad (1)$$

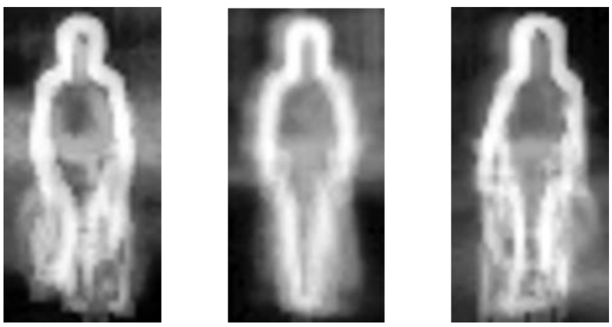
$$\overline{I_r}(x,y) = \frac{1}{N}\sum_{i=0}^{N} I_r(x,y) \quad (2)$$

Here $\overline{I_p}(x,y)$ is the pedestrian model and $I_p(x,y)$ is the differential image of the pedestrian, $\overline{I_r}(x,y)$ is the bicyclist model and $I_r(x,y)$ is the differential image of the bicyclist, and N is the number of images.

The average of each image calculated by Eqs.(1) and(2) is used to make a passerby model. Three passerby models are created, i.e., a model from 2 o'clock, a model from 12 o'clock, and a model from 10 o'clock directions. **Figure 2** shows the created models.

3.1.2. *Cell placement*

Based on the passerby model, we place cells for feature extraction. In order to obtain the edge of the target in more precisely, cells are placed densely at the locations where the edge strength of the passerby model is strong, and cells are placed sparsely at the locations where the edge strength is weak. The set of cells used for feature extraction is represented as *CELL* by Eqs(3)-(5).



(a) (b) (c)

Fig.2 Passerby Model. Passersby approaching from(a) 10 o'clock, (b) 12 o'clock, and(c) 2 o'clock direction.

$$CELL = \left\{ \begin{matrix} \left(cx, cy, w, h \middle| cell(cx, cy, w, h) \subset window(W, H) \right) \\ w = a + ci, h = b + dj, \forall i \forall j (i, j \geq 0) \end{matrix} \right\} \quad (3)$$

$$cell(cx, cy, w, h) = \left\{ \begin{matrix} (x, y) | cy - \frac{h-1}{2} \leq y < cy + \frac{h+1}{2} \\ cx - \frac{w-1}{2} \leq x < cx + \frac{w+1}{2} \end{matrix} \right\} \quad (4)$$

$$window(w, h) = \{ (x, y) | 0 \leq y < h, 0 \leq x < w \} \quad (5)$$

Here, cx, cy, w, h are the *x*- and *y*-coordinates of the center of the cell, the horizontal and vertical sizes of the cell, respectively: W, H a, b are the horizontal and vertical sizes of the detection window, the minimum values of the horizontal and vertical sizes of the cell, respectively: $c_{i,}$ and c_j are the horizontal and the vertical sizes of the cell, respectively. $c_{i,}$and c_j vary depending on the luminance of the passerby model.

3.4. *Computation of Feature Values*

In MSC-HOG features, the gradient direction and gradient intensity are calculated in the same way as in the conventional method of HOG[4] features, and a histogram is created. However, normalization is done in each cell. For the gradient vector, the gradient direction from 0° to 180° is divided into 9 directions with 20° each, and a histogram of 9 bins is created.

4. Experimental Results

We conducted two types of experiments. Experiment 1: Confirming the effectiveness of the passerby model, Experiment 2: Confirming the effectiveness of the direction of passerby.

In experiment 1, we use three direction passerby images as a positive dataset, and also use the images that do not included passersby as a negative dataset. In experiment 2, since our purpose is to recognize only the passersby approaching from the specified direction to a user, we choose one of the specified passersby directional images as a positive dataset, and we put the other two directional images as negative dataset, so that we can obtain a specified direction discriminator. In both experiments, Random Forest was used as the discriminator, and a total of three different approaching direction discriminators were created for recognizing passerby and their approaching directions.

For the passerby and directional learning images, we used 3000 images of a passerby approaching from the 10 o'clock direction, 1752 images of a passerby approaching from the 12 o'clock direction, 3000 images of a passerby approaching from the 2 o'clock direction, and 6000 negative images. For the learning images, we used our own dataset and the INRIA Person Dataset[5].

For the test images, 500 pedestrians in the 10 o'clock (referred to as group A in the experiment 1), 12 o'clock (referred to as group B in the experiment 1) and 2 o'clock (referred to as group C in the experiment 1) directions and 500 bicyclists in each direction were collected and used as a positive test data, and 1000 negative images from the INRIA Person Dataset were used as negative test data.

The results of the both experiments are shown in **Table.1** and **Table.2,** respectively. Recall, Precision, and F-values are used to evaluate the proposed method.

Table.1 Accuracy of the passerby detection (Experiment 1)

	group A	group B	group C
$Recall$	0.961	0.962	0.933
$Precision$	0.980	0.999	0.987
F	0.970	0.980	0.959

Table.2 Accuracy of the passerby approaching direction (Experiment2)

	Directions	10 o'clock	12 o'clock	2 o'clock
Pedestrian	$Recall$	0.788	0.670	0.776
	$Precision$	0.999	1.000	0.979
	F	0.881	0.802	0.866
Bicyclists	$Recall$	0.636	0.922	0.774
	$Precision$	0.998	1.000	0.979
	F	0.777	0.959	0.865
Average	$Recall$	0.712	0.796	0.775
	$Precision$	0.999	1.000	0.979
	F	0.831	0.886	0.865

5. Discussion

In this study, the goal is to help visually impaired people walk safely. It is therefore, important to avoid failing of recognition to approaching passersby. For this reason, among the three evaluation indexes shown in Table1&2, the Recall value is the most important index that needs to

be noticed. From Table.1, the minimum Recall value was 0.93 with the respect to the data group A, B, and C. This indicates the effectiveness of the proposed MSC-HOG feature and the passerby model.
From Table.2 the precision results are high, whereas, the value of Recall of a passerby is 0.76 in average.

The recall of the bicyclist was 92.2[%] in the 12 o'clock direction, but 63.6 [%] and 77.4 [%] in the 10 o'clock and 2 o'clock directions, respectively. One of the reasons of this lower accuracy in direction 10 and 2 o'clock may be the symmetric figure of a bicycle. On the other hand, the recall of pedestrians approaching directions were 67[%], 78.8[%], and77.6[%] in 12,10, and 2 o'clock, direction, respectively. Adversely, the 12 o'clock direction obtained a lower accuracy compared to the two other directions. Since the direction changes along time, the recall rate of pedestrian and bicyclist directions detection recall rate can be improved by adding another characteristics of the direction of movement such as Optical Flow.

6. Conclusion

In this paper, we have proposed a method for recognizing pedestrians that including bicyclist and their approaching directions using the MY VISION system. Unlike conventional method[2], we considered and detected not only pedestrians but also bicyclists and their approaching directions. For this purpose, MSC-HOG with a passerby model was introduced. We also proposed a method of recognizing the direction of approaching passersby using three different directions discriminators.

The effectiveness of the propose was confirmed by two experiments that focused on the recognition of passersby and also on the recognition of three main approaching directions. Further improvement on the recognition of passerby approaching directions, especially, 10 o'clock and 2 o'clock directions of bicyclists remains for future work.

References

1. "Social Welfare Corporation Japan Braille Library official website", https://www.nittento.or.jp/news/koekake.html
2. Joo Kooi Tan, Kenta Hori, Seiji Ishikawa, "Detecting pedestrians and their walk directions using a MY VISION system", Journal of Robotics, Networking and Artificial Life, vol.7-2, pp. 137-141, 2020.
3. Y. Ehara, J.K. Tan, H. Kim S. Ishikawa, "Human detection from an image employing reduced number of weak-classifiers", Proceedings of the IIEEJ Image Electronics and Visual Computing Workshop, IEVC, Kuching, Malaysia, 2012, pp. 1–5.
4. N. Dalal, B. Triggs, C. Schmid, "Human detection using oriented histograms of flow and appearance", Proceedings of European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science, vol. 3952, Springer, Berlin, Heidelberg, 2006, pp. 428–441.
5. "INRIA Person Dataset", http://pascal.inrialpes.fr/data/human/.
6. Yuta Ono, Joo Kooi Tan, Akitoshi Sato, "Finding a rush-out human employing a human body direction detector", Proceedings of International workshop on Advance Image Technology (IWAIT 2019) , 2019, PP.1-6.

Authors Introduction

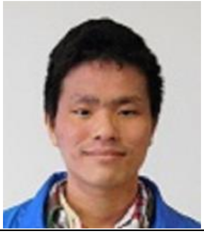
Mr. Shinya Iizumi

He received his B.E. degree in Department of Control Engineering in 2020 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 image processing, pattern recognition, and robotics.

Prof. Dr. Joo Kooi Tan

She is currently with Department of Mechanical and Control Engineering, Kyushu Institute of Technology, as Professor Her current research interests include ego-motion, three-dimensional shape and motion recovery, human detection and its motion analysis from videos. She was awarded SICE Kyushu Branch Young Author's in 1999, the AROB Young Author's Award in 2004, Young Author's Award from IPSJ of Kyushu Branch in 2004 and 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.

Mr. Yuta Ono



He received M.E. from Kyushu Institute of Technology. He is now in the Ph.D. course of the Graduate School of Engineering, Kyushu Institute of Technology. His research includes computer vision, machine learning, and rush-out pedestrian detection from videos.

Emeritus Prof. Seiji Ishikawa

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

Dr. Masuhiro Nitta

He received his PhD from Nara Institute of Science and Technology. He is now an assistant professor in Faculty of Engineering at Kyushu Institute of Technology in Japan.