ICIC Express Letters
Part B: Applications
Volume **8**, Number **12**, December **2017**
 ISSN 2185-2766
pp. **1611–1618**

MULTI-OBJECT RECOGNITION BASED ON 2DHOG WITH VARIABLE WINDOW APPROACH

Ryo Kawanami[1] and Kousuke Matsushima[2]

[1]Advanced Engineering School
[2]Department of Control and Information Systems Engineering
National Institute of Technology, Kurume College
1-1-1, Komorino, Kurume-shi, Fukuoka 830-8555, Japan
a3302rk@std.kurume-nct.ac.jp; matsushima@kurume-nct.ac.jp

Received June 2017; accepted September 2017

Abstract. *Multi-object recognition (MOR) technologies in road scenes have received much attention due to the necessity of advanced safety vehicle (ASV) in recent years. In order to keep the safety of vulnerable road users (VRUs), it is important to recognize VRUs rapidly, correctly and automatically. According to experiments on bicycle recognition, two-dimensional histogram of oriented gradient (2DHOG) with variable window approach (VWA), one of feature extraction methods, shows high accuracy and fast processing speed. However, previous works have focused only on bicycle recognition because cyclists often face the possibility of traffic accidents compared to pedestrians on daily basis. In this study, 2DHOG was applied with VWA to the mixed data including data set of the bicycle, pedestrian, and road structure to represent actual traffic situations. Experimental results of the mixed data show that the recognition rates of the bicycle, the pedestrian, and the road structure sets were 93%, 99%, and 100%, respectively.*
Keywords: Multi-object recognition, Two-dimensional histogram of oriented gradient, Variable window approach

1. **Introduction.** The intelligent vehicles (IVs) have assisted in reducing the number of road traffic injuries in Japan. However, vulnerable road users (VRUs) have not benefited to the same extent as vehicle users. Therefore, many researchers have been studying to increase the safety using an in-vehicle camera for VRUs such as pedestrians, bicycles, and pedestrians.

For instance, vision-based method is proposed, which combines the use of a pedestrian model as well as the walking rhythm of pedestrians to detect and track walking pedestrians [1]. Recent research shows that there emerge two dominant pedestrian detection methods, namely deep learning incurred convolutional neural network [2,3] and classifiers with hand crafted features [4-7] (e.g., histogram of oriented gradient (HOG) [8], local binary patterns (LBPs) [9-11], scale invariant feature transform (SIFT) [12], and speed up robust features (SURF) [13]). Jung et al. also proposed an improved HOG named multiple size cell HOG (MSC-HOG) [14] and used the real AdaBoost algorithm [15] uniting weak classifier to detect the bicycle. Watanabe et al. used the CoHOG feature [16] only on the lower part of the blob because of the expensive computational cost to verify the bicyclist detection, and proposed a semi-supervised learning to estimate the orientation [17].

Actually, various kinds of objects exist in traffic situations and have to be detected in vision sensor approach [18-20]. In addition, we must recognize multi-object such as bicycles, pedestrians, and other road structures respectively because the danger is different for each object. For example, although various road objects are in road traffic scenes, pedestrians and cyclists are similar to each other. However, cyclists are faster and more dangerous than pedestrians. Thus, multi-object recognition (MOR) is important to

distinguish each object. In various feature extraction methods, 2DHOG [21] is the effective way for high accuracy recognition. Although the processing time is inferior to other feature extraction methods, the defect has already been improved using variable window approach (VWA) [22]. In this study, three objects composed of bicycles, pedestrians, and other road structures were recognized using 2DHOG with VWA.

In Section 2, 2DHOG with VWA is discussed. Section 3 covers the mask processing. MOR is discussed in Section 4. Section 5 shows the experiment and the results. Section 6 concludes the paper.

2. 2DHOG with VWA.

2DHOG with VWA, feature extraction method is processed according to the following procedure.

2.1. Computation of gradient.

Firstly, the two kinds of information on the gradient are needed to extract features. The gradient is composed of gradient magnitude $m(x, y)$ and gradient orientation $\theta(x, y)$. These are calculated by (1) and (2), respectively.

$$m(x,y) = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2} \tag{1}$$

$$\theta(x,y) = \tan^{-1}\left(\frac{\partial I}{\partial y} \Big/ \frac{\partial I}{\partial x}\right) \tag{2}$$

where symbol I means an input image and (x, y) expresses coordinate of the image. These Equations (1) and (2) are called parallel shift-invariant gradient (PS-I Gradient).

2.2. VWA.

VWA makes it possible to place the polar coordinate areas in an input image randomly. The polar coordinate areas are described later. The steps of approach are as the following:

- Generate unbiased random numbers by using Mersenne Twister method [23]
- Assign random numbers to focused polar coordinates
- Decide the appropriate number of polar coordinates by experiment which is described later, and place it in the image

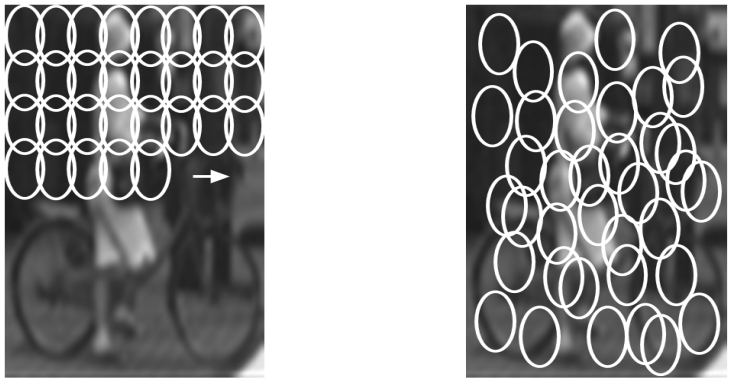
In the case of VWA, the polar coordinate area was arranged from the upper left to the lower right by raster scan as shown in Figure 1(a). However, when introducing VWA, the polar coordinate area is randomly arranged in the input image as shown in Figure 1(b). Furthermore, this approach makes it possible to determine the number of polar coordinate areas arbitrarily and processing time is shortened by deciding an appropriate number since arranged number is proportional to the processing time.

2.3. Polar coordinate area.

The polar coordinate area is used for calculating a two-dimensional histogram to be described later. In this research, elliptical areas as shown in Figure 2(a) are used as polar coordinate areas. Assuming that a focused pixel is (x, y), the surrounding pixels $S(u, v)$ are calculated by (3).

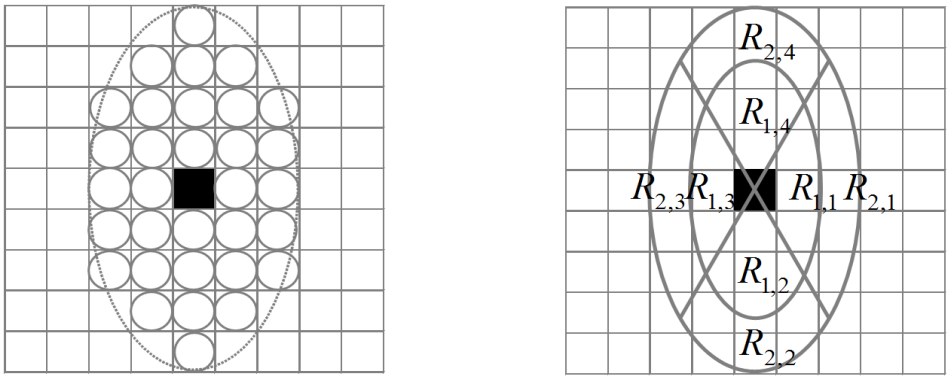
$$S(x,y) = \left\{ (u,v) \middle| \left(\frac{x-u}{a^2+2^2}\right)^2 + \left(\frac{y-v}{b^2+2^2}\right)^2 < 1,\ (u,v) \neq (x,y) \right\} \tag{3}$$

The symbol of long diameter shows $a = 2$ and the symbol of short diameter shows $b = 4$. Furthermore, the polar coordinate areas are divided into $R_{r,d}$ ($r = 1, 2, \ldots, m$, $d = 1, 2, \ldots, n$) areas based on the distance and direction around the focused pixels as shown in Figure 2(b). By using $R_{r,d}$ region, the object in the image becomes robust against local rotation changes. Symbol each of r and d is the label representing the distance and the direction which are divided centering on the focused pixel.



(a) Raster scan (b) VWA

FIGURE 1. Arrangement of polar coordinate areas



(a) Elliptical area (b) Polar coordinate area

FIGURE 2. Area for two-dimensional histogram

2.4. **Two-dimensional histograms.** Two-dimensional histograms $H_{r,d}(o_0, o_1)$ are calculated for each polar coordinate area using gradient information as the following

$$H_{r,d}(o_0, o_1) = \sum_{\boldsymbol{X} \in \boldsymbol{I}} \sum_{\boldsymbol{P} \in S(\boldsymbol{X})} \{w | m(\boldsymbol{X}), m(\boldsymbol{P}) | t_{r,d}(\boldsymbol{P}) \cdot f_{o_0}(\boldsymbol{X}) \cdot f_{o_1}(\boldsymbol{P})\} \tag{4}$$

where each of $\boldsymbol{X} = (x, y)$ expresses the focused pixel set of the polar coordinate area. $\boldsymbol{X} = (x, y)$ is the factor of $\boldsymbol{I} = \{(x_\alpha, y_\beta) | \alpha = 1, 2, \ldots, X, \beta = 1, 2, \ldots, Y\}$, the vector sets concerning the input images which have the width X and the height Y. $\boldsymbol{P} = (u, v)$ is also the factor of $S(\boldsymbol{X}) = S(x, y)$, the surrounding pixels as shown in (3). In addition, the symbol o_0, $o_1 = 1, 2, \ldots, D$ denotes quantization parameters obtained from the relation between $\boldsymbol{X}$ and $\boldsymbol{P}$. Moreover, $t_{r,d}$ expresses the weight to function w, which consists of f_{o_0} and f_{o_1}. Thus, $H_{r,d}$ is set of element w which satisfies the above condition.

2.5. **Block normalization.** An input image is divided into pixel sets called block as shown in Figure 3. In this research, we use rectangular blocks, which are used for normalization. Each of rectangular width and height is two and four, respectively. Two-dimensional histograms are divided into blocks as shown in Figure 3 and then are normalized by L2-Hys.

3. **Mask Processing.** In order to more effectively extract feature quantities from images, we applied mask processing to both learning image and testing image. In this study, we used this processing between bicycles, pedestrians. Because the images of road structures contain backgrounds and buildings, mask processing was not applied only to road structures. Mask processing is performed in the following steps and Figure 4.

- Extract ROI from an image and then create original image
- Calculate the parallax from left and right images shot using a stereo camera (stereo matching)
- Create a depth image from original image using the derived parallax
- Using a depth image, create a mask image by filling outside the bicycle in the original image

The mask image is created following this process and used as an input image such as Figure 5 in experiments.

FIGURE 3. Block division

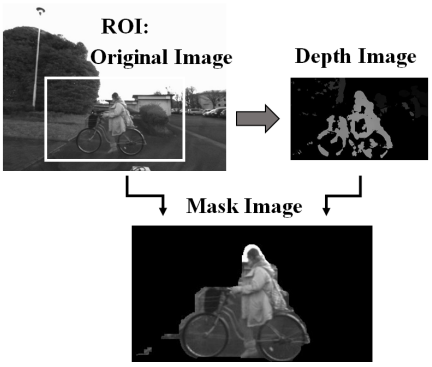


FIGURE 4. Mask processing

(a) Bicycle (b) Pedestrian (c) Road structure

FIGURE 5. Mask image (except road structure)

4. **MOR.** We have recognized bicycles by combining VWA with 2DHOG in the previous study, and the result showed enough accuracy and processing time. The reason why we have recognized only bicycles is that cyclists are faster and more dangerous than pedestrians and others. In this study, we propose the method that 2DHOG with VWA is applied to MOR. Flowchart of MOR is shown in Figure 6. Input images acquired from stereo vision are recognized with the classifier for bicycle and pedestrian recognition. As a result, it shows any of bicycles, pedestrians, and other road structures.

5. **Experiments.** In this research, we conducted the following two experiments.

5.1. **Number of appropriate polar coordinate areas.** While we need to increase the number of coordinate areas to improve recognition accuracy, we must reduce the number of areas to increase processing speed. Thus, we have to determine the number of polar coordinate areas which are appropriate for each of the bicycle and the pedestrian. In this experiment, we evaluate whether the number is appropriate or not by calculating the recognition rate about the bicycle and the pedestrian using the various number of polar coordinate areas. We show the experimental environment in Table 1. The number of

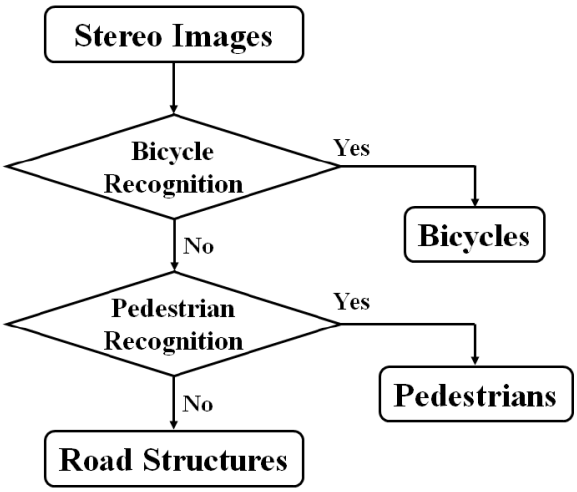


FIGURE 6. Flowchart of multi-object recognition

TABLE 1. Experimental environment

Item	Performance
OS	Windows 7 Home Premium
CPU	Intel (R) Pentium (R) CPU G2020 @2.90 GHz
Memory	4.00 GB

TABLE 2. Number of image in experiment 5.1

Object	Train Image (pcs)		Test Image (pcs)	
	Positive	***Negative***	***Positive***	***Negative***
bicycle	30	30	286	99
pedestrian	500	500	2500	2500

images used in this experiment is shown in Table 2, the recognition rate of the bicycle is shown in Figure 7, and the recognition rate of pedestrian is shown in Figure 8. In Table 2, "positive" represents an image of the target itself, and "negative" represents an image other than the target. For example, if it is "positive", it is a bicycle on "bicycle", if it is "negative", it is a pedestrian or other road structure. All images have widths and heights of 64 and 128 pixels. According to Figure 7, the recognition rate of the bicycle shows almost the same in 3000 or later. In addition, according to Figure 8, the recognition rate of the pedestrian hardly changes after 5000. Therefore, in this experiment, we decided the number of polar cordite areas arranged in the image as 5000. The calculated feature quantities are classified by support vector machine (SVM) [24].

5.2. **Recognition experiment of bicycles and pedestrians.** In this experiment, we assume the actual traffic situation and use the testing image set in which bicycle, pedestrian, and other road structures are mixed as the test image. As described above, mask processing is applied to the test image. The number of images used in this experiment is shown in Table 3, and the result of the experiment is shown in Table 4. As shown in Table 4, there were only seven images that misrecognized a bicycle as a pedestrian despite the similarity of the bicycle and the pedestrian. In addition, there is only one image that misrecognized a pedestrian as another road structure, which is thought to be caused by pillars and buildings in the image.

6. **Conclusions.** When a number of elliptical polar coordinate areas are decided 5000, experiment to recognize the 100 images mixed bicycle, pedestrian, and road structure

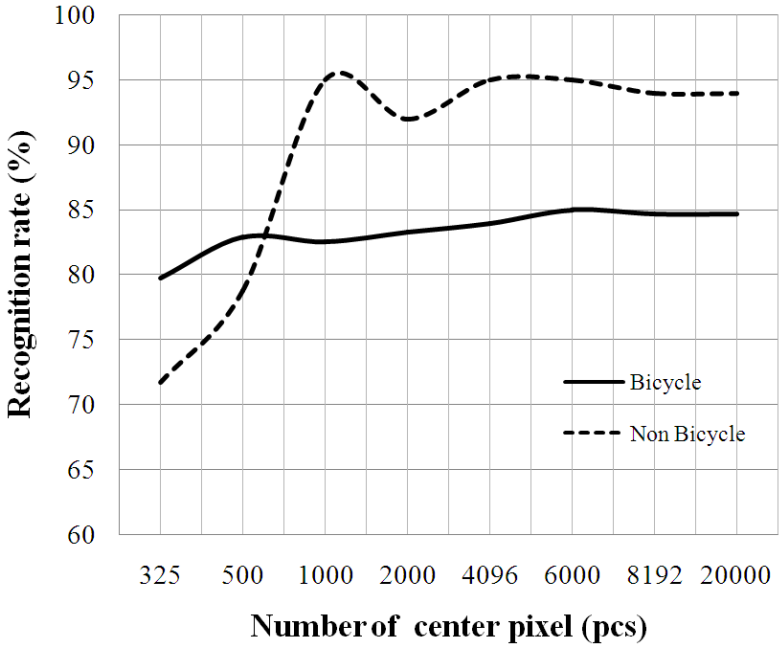


FIGURE 7. Comparison of recognition rate about bicycles

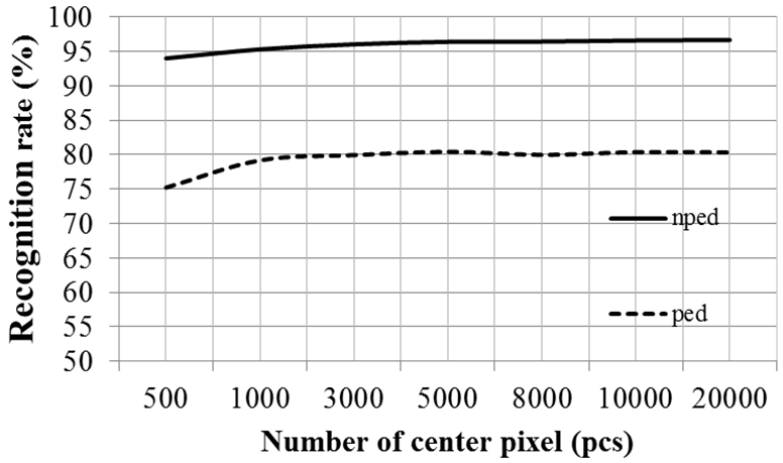


FIGURE 8. Comparison of recognition rate about pedestrians

TABLE 3. Number of images in experiment 5.2

Object	Training Image (pcs)	Testing Image (pcs)
bicycle	30	100
pedestrian	30	100
road structure	30	100

TABLE 4. Recognition result in experiment 5.2

Object	Recognition Result (pcs)			Rate (%)
	Bicycle	Pedestrian	Road object	
bicycle	93	7	0	93
pedestrian	0	99	1	99
road structure	0	0	100	100

was performed with 2DHOG using VWA. This method showed high accuracy using two discriminators which are for bicycle recognition and for pedestrian recognition. As a result, the recognition rate showed as the following: (bicycle: 93%, pedestrian: 99%, road object: 100%). Bicycles, pedestrians, and other road structures are mixed in actual traffic conditions. Therefore, we can conclude that the feature quantity obtained by using this method is effective also when multiple objects are actually recognized on the road. In the future, we will apply MOR to not only bicycles, pedestrians, and road objects but also motorcycles, car, and so on.

REFERENCES

[1] C. J. Paia, H. R. Tyanb, Y. M. Liang et al., Pedestrian detection and tracking at crossroads, *Pattern Recognition*, vol.37, no.5, pp.1025-1034, 2004.

[2] J. Hosang, M. Omran, R. Benenson et al., Taking a deeper look at pedestrians, *The IEEE Conference on Computer Vision and Pattern Recognition*, pp.4073-4082, 2015.

[3] B. Yang, J. Yan, Z. Lei et al., Convolutional channel features, *International Conference on Computer Vision*, pp.82-90, 2015.

[4] S. Zhang, R. Benenson and B. Schiele, Filtered channel features for pedestrian detection, *The IEEE Conference on Computer Vision and Pattern Recognition*, pp.1751-1760, 2015.

[5] P. Dollar, R. Appel, S. Belongie et al., Fast feature pyramids for object detection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.36, no.8, pp.1532-1545, 2014.

[6] H. Ren and Z. N. Li, Object detection using boosted local binaries, *Pattern Recognition*, vol.60, pp.793-801, 2016.

[7] Q. Liu, J. Zhuang and J. Ma, Robust and fast pedestrian detection method for far-infrared automotive driving assistance systems, *Infrared Physics and Technology*, vol.60, pp.288-299, 2013.

[8] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, *Proc. of Computer Vision and Pattern Recognition*, vol.1, pp.886-893, 2005.

[9] T. Ojala, M. Pietikäinen and T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.24, no.7, pp.971-987, 2002.

[10] T. Ahonen, A. Hadid and M. Pietikäinen, Face recognition with local binary patterns, *Proc. of European Conference on Computer Vision*, 2004.

[11] B. Jun, I. Choi and D. Kim, Local transform features and hybridization for accurate face and human detection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.35, no.6, pp.1423-1436, 2013.

[12] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, vol.60, no.6, pp.91-110, 2004.

[13] H. Bay, A. Ess, T. Tuytelaars et al., Speeded-up robust features (SURF), *Computer Vision and Image Understanding*, vol.110, no.3, pp.346-359, 2008.

[14] H. Jung, Y. Ehara, J. K. Tan, H. Kim and S. Ishikawa, Detection of a bicycle in video images using MSC-HOG feature, *International Journal of Innovative Computing, Information and Control*, vol.10, no.2, pp.521-533, 2014.

[15] R. E. Schapire and Y. Singer, Improved boosting algorithms using confidence-rated predictions, *Machine Learning*, vol.37, no.3, pp.297-336, 1999.

[16] T. Watanabe, S. Ito and K. Yokoi, Co-occurrence histograms of oriented gradients for pedestrian detection, *IPSJ Trans. Computer Vision and Applications*, vol.2, pp.39-47, 2010.

[17] G. Yanlei and S. Kamijo, Bicyclist recognition and orientation estimation from on-board vision system, *International Journal of Automotive Engineering*, vol.6, no.2, pp.67-73, 2015.

[18] D. Singh and C. K. Mohan, Graph formulation of video activities for abnormal activity recognition, *Pattern Recognition*, vol.65, pp.265-272, 2017.

[19] M. H. Zaki and T. Sayed, A framework for automated road-users classification using movement trajectories, *Transportation Research Part C: Emerging Technologies*, vol.33, pp.50-73, 2013.

[20] A. Alahia, M. Bierlaireb and P. Vandergheynst, Robust real-time pedestrians detection in urban environments with low-resolution cameras, *Transportation Research Part C: Emerging Technologies*, vol.39, pp.113-128, 2014.

[21] K. Fukushima and K. Matsushima, Robust multi-directional bicycle recognition on the rotation using the stereo vision, *IEEE International Conference on Vehicular Electronics and Safety*, pp.36-40, 2015.

[22] R. Kawanami and K. Matsushima, Bicycle detection by variable window approach of 2DHOG descriptors, *International Conference on Electronics, Information, and Communication*, vol.S04, pp.89-92, 2017.
[23] M. Matsumoto and T. Nishimura, Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator, *ACM Trans. Modeling and Computer Simulation*, vol.8, no.1, pp.3-30, 1998.
[24] E. Osuna, R. Freund and F. Girosi, Training support vector machines: An application to face detection, *Proc. of Computer Vision and Pattern Recognition*, pp.130-137, 1997.