The Seventeenth International Symposium on Artificial Life and Robotics 2012 (AROB 17th '12),
B-Con Plaza, Beppu, Oita, Japan, January 19-21, 2012

Human Detection Employing the HOG Feature based on Multiple Scale Cells

Yusuke Ehara, Joo Kooi Tan, Seiji Ishikawa, and Takashi Morie

Kyushu Institute of Technology, Kitakyushu 804-8550, Japan

Tel: 093-884-3088, Email: ehara@ss10.cntl.kyutech.ac.jp

Abstract: In this paper, we propose a new human detection method based on local descriptors and machine learning. The HOG feature and RealAdaBoost are well-known methods for detecting humans. However, this technique has shown its limitation because of long processing time due to fixed number of weak-classifiers. To overcome this problem, this paper proposes a HOG based framework using multiple-scale cells for the RealAdaBoost with a variable number of weak-classifiers. The proposed method is faster than existent methods, because it describes more comprehensive intensity gradients and classifies them using low dimension feature vectors. Experimental results show that the proposed method is effective in accuracy and processing time.

Keywords: HOG, Multiple Scale Cells, RealAdaBoost, human detection, machine learning.

1. Introduction

Recently, various techniques for human detection (especially pedestrians) from an image have been developed. These techniques are the main topics in the field of ITS (Intelligent Transport systems) and robot vision. For pedestrian detection, there exist many approaches such as background subtraction, motion based techniques [1], etc. that have been proposed. However, these methods have restrictions in the scenes in which they can be used. For example, some methods can be used only for stationary camera images or only for a moving object. Therefore, using local descriptors and machine learning schemes which can respond to a moving camera and do not have restriction in the detection scene has been gaining spotlights. More especially, the Histograms of Oriented Gradients (HOG) feature proposed by Dalal and Triggs [2] is one of the well-known methods for human detection. The method has robustness to illumination and small shape changes.

There exist a lot of methods based on the HOG feature [3, 4]. For example, Nakashima has proposed Multiple-HOG [5]. Multiple-HOG method is based on using a variable number of divisions on the direction of intensity gradients. Better results have been obtained as compared with conventional methods because of its flexible as well as precise description of the gradients in a cell. However, since the size of a cell is fixed, the amount of information included in one feature vector is limited. Moreover, these methods have the same problem, i.e., long processing time, because HOG needs normalization of every block in an image and the number of weak-classifiers is large and fixed in the RealAdaBoost. If the number of weak-classifiers is reduced, processing time can be faster. Instead, we may might lose high accuracy.

In this paper, we propose a novel human detection method employing the HOG feature based on multiple-scale cells for RealAdaBoost with a variable number of weak-classifiers. By this method, since the size of the edges can be captured more comprehensively, one feature vector contains more information. Therefore, this method can perform at higher speed using few weak-classifiers. Experimental results show that the proposed method is effective in accuracy and in processing time.

2. HOG feature based on Multiple-scale cells

When calculating the HOG feature employing the method proposed by Dalal et al., a cell size is fixed. In this method, we can acquire only the information on a small cell (usually 5×5 pixels) as shown in **Fig. 1(a)**. However, a pedestrian has generally large and strong vertical edges. We propose Multiple-scale-cell HOG feature in this paper. The idea of Multiple-scale-cell HOG is illustrated in Fig. 1(b).

(a) (b)

Fig. 1. Arrangement of a cell in the case of (a) HOG, and (b) Multiple-scale-cell HOG.

The Multiple-size-cell HOG has a variable size of cells. For example, 5×5, 5×20, 20×5 and so on. Moreover a cell-overlap is permitted. Then we can describe the size and the location of an edge comprehensively. The cell size is defined as follows.

$$S_x = \alpha n + C, \qquad S_y = \beta k + C \tag{1}$$

Here S_x is the cell size in the horizontal axis direction, and S_y is the cell size in the vertical direction.

When calculating the HOG feature, block normalization consumes a large part of the processing time, although it assures robustness to lighting condition change. Therefore normalization with every block is not performed in the proposed method. As a substitute, cell overlap is permitted which leads to the use of a larger number of cells.

3. RealAdaBoost with variable number of weak-classifiers

Boosting obtains classifiers of good performance (a strong classifier) by combining many weak-classifiers. The number of weak-classifiers is related to accuracy and processing time. In this paper, we also propose a technique for making the number of weak-classifiers variable. If an image is simple enough to recognize it, such as a road or the sky, we use a small number of weak-classifiers. On the other hand, if an image is complicated, such as roadside trees, we use a large number of weak-classifiers.

First, we train provided data using RealAdaBoost. In RealAdaBoost, the output of a weak-classifier becomes large (positive) when it is a positive sample, and it becomes small (negative) when the sample represents the negative class. Using this characterization during the learning stage of RealAdaBoost, the threshold value is determined by the following formula.

$$th_i = \min\{h_{i,1}, h_{i,2}, \cdots h_{i,N_p}\} \tag{2}$$

Here i is the round number, th_i is the threshold in round, h_i is the output value of a weak-classifier in i round about a positive sample. N_p is the number of positive samples.

Discrimination is performed using the computed threshold value. The idea of discrimination is illustrated in **Fig. 2**. In Fig. 2, WC_i are weak-classifiers in round i. Even if i have not reached, if the output of weak-classifiers is smaller than a threshold value, calculation will stop at the time. During the image search of a human, the number of weak-classifiers is made variable for every search window as shown in **Fig. 3**.

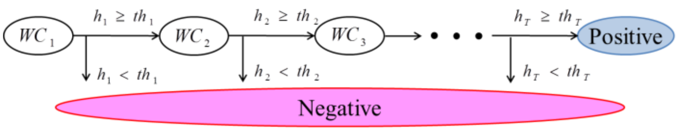


Fig. 2. Variable number of weak-classifiers.

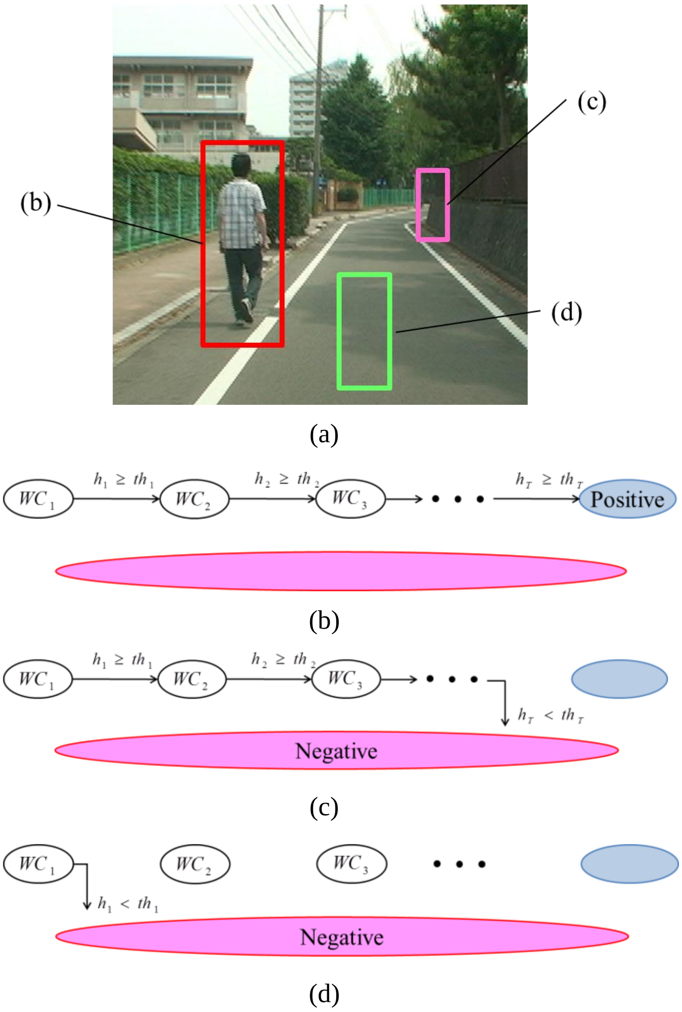


Fig. 3. The number of weak-classifiers for every search: (a) Input image, (b) positive class, (c) difficult negative class, and (d) easy negative class.

4 EXPERIMENTAL RESULTS

We evaluated the proposed method for human detection by using INRIA Person Dataset. Training data contains 2416/6000(positive/negative) images and test data includes 1126/3000(positive/negative) images. The image size is 30×60 pixels. We compared the proposed method with the conventional method using the HOG feature [1] and RealAdaBoost, and using the Multiple-HOG feature[5] and RealAdaBoost.

The combination of the method in the experiment is shown in a **Table 1**. We compared two previous methods to our proposed schemes. In [1], [5] and our proposed method-1, the number of weak-classifiers is 500. First, the accuracy evaluation experiment of the proposed method (Multiple-scale-cell HOG) was conducted. The DET curve is shown in **Fig. 4**. The DET curve is plotted: True Negative rate is the vertical axis, whereas False Positive rate is the horizontal axis. The method that has its plot in the lowest and the left most area is considered to be the best. Moreover, the accuracy in the case of 500 weak-classifiers is shown in **Table 2**.

Next, we conducted the evaluation experiment of RealAdaBoost with variable number of weak-classifiers. The relationship between the number of weak-classifiers and the rate of an error is shown in **Fig. 5**. The vertical axis represents the error rate. The method that has the lowest plot is the best, which the proposed method-2.

Finally, processing time is compared in **Table 3**. Since processing time differs in a Positive class and a Negative class sample, in the proposed method-2 the processing time of each class is shown separately.

Table 1. The combination of the method in

previous method-1[1]	HOG and RealAdaBoost
previous method-2[5]	Multiple-HOG and RealAdaBoost
proposed method-1	Multiple-scale-cell HOG and RealAdaBoost
proposed method-2	Multiple-scale-cell HOG and RealAdaBoost with variable number of weak-classifiers

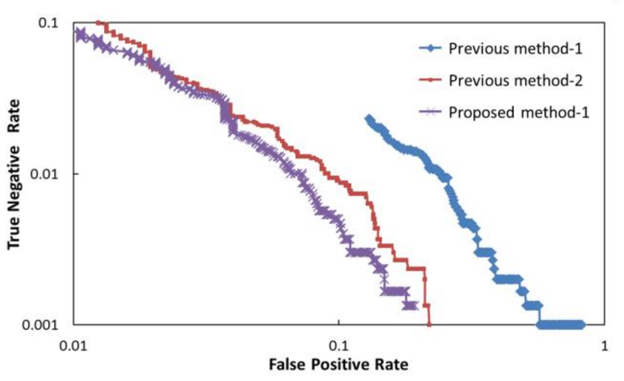


Fig. 4. Accuracy evaluation experiment of Multiple-scale-cell HOG.

Table 2. Detection rate

	Detection Rate[%]
previous method-1	94.15
previous method-2	96.24
proposed method-1	96.32

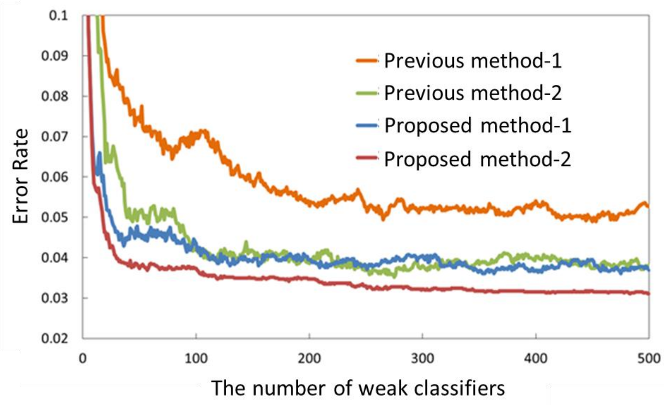


Fig. 5. The relation between the number of weak-classifiers and the error rate.

Table 3. Processing time

	Positive [micro sec]	Negative [micro sec]
previous method-1	152.51	
previous method-2	40.24	
proposed method-1	28.96	
proposed method-2	27.80	0.99

5 DISCUSSION

Fig.4. shows that the proposed method achieved satisfactory accuracy compared with the previous methods. Table 2 also supports this fact. Moreover, the best accuracy was obtained by using Multiple-scale-cell HOG and RealAdaBoost with variable number of weak-classifiers from as shown in Fig.5. It is thought that a satisfactory result is obtained because the probability of FPR was reduced by making the number of weak-classifiers variable.

Moreover, Table 3 showed that the proposed method-2 performed the specified task at a higher speed. This is due to the skipping of the block normalization during the feature vector computation stage. Moreover, during the discrimination stage, it turns out that the negative class can be eliminated at a high speed.

Next, the plot showing the relationship between the number of weak-classifiers and the rate of an error in the first 100 rounds is shown **in Fig. 6**. As shown in Fig. 5, When the number of weak-classifiers is high (~500), the difference in the accuracy of previous methods and Multiple-scale-cell HOG (proposed method-1) is small. However, as shown in Fig. 6, in the first 100 rounds, there is a big difference in the accuracy of two methods. For this reason, within the first 100 rounds, the proposed method can eliminate the negative class samples with high accuracy. This is considered to be based on the contribution of a large scale cell. The amount of information which one feature vector has by using a large scale cell becomes large. Therefore, good accuracy can be acquired when few weak classifiers are used. This is understood also from the feature selection process. The selected edge in three first rounds is shown in **Fig. 7**. It turns out that a large edge is chosen in the first 100 rounds. On the contrary, when the number of weak-classifiers is high, a small edge is chosen. That is, first, its attention is paid to big edge and easy portions, such as a road domain, are eliminated. Next, as Round becomes high, difficult objects, such as a roadside tree, are eliminated using small edge.

6 CONCLUSIONS

We have proposed a method of human detection employing the Multiple-scale-cell HOG and RealAdaBoost with variable number of weak classifiers.

In the experiment, it was checked that it performed higher-speed processing and a better precision than the previous methods. Performance of the method was satisfactory. We are going to apply the method to ITS and robot vision.

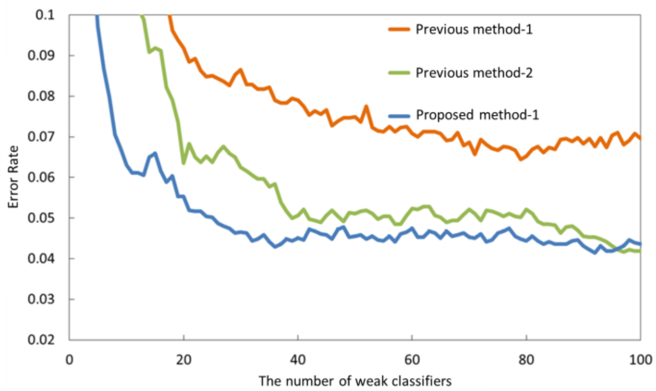
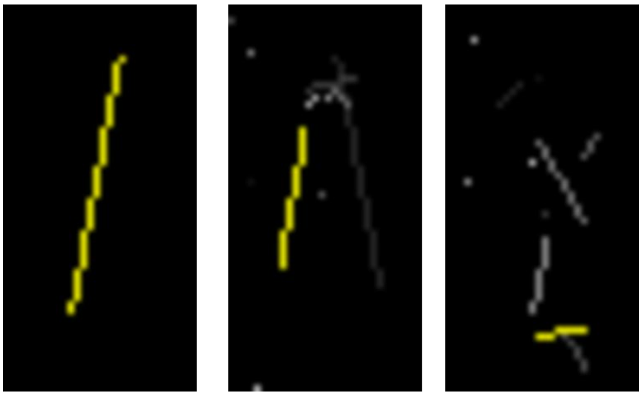


Fig. 6. The relation between the number of weak-classifiers and the error rate in the first 100 rounds.



(a) (b) (c)

Fig. 7. Selected edges:(a) Round 1, (b) round 240, and (c) round 499.

ACKNOWLEDGMENT

This work was supported by grant of Regional Innovation Cluster Program and JSPS KAKENHI (22510177).

REFERENCES

[1] Y. Zhang, S. J. Kiselewich, W. A. Bauson, R. Hammoud (2006) Robust moving object detection at distance in the visible spectrum and beyond using a moving camera, Proc. Computer Vision and Pattern Recognition Workshop (CVPRW), 131-139.

[2] N. Dalal, B. Triggs (2005), Histograms of oriented gradients for human detection, Proc. IEEE Conf. on Computer Vision & Pattern Recognition, 886-893.

[3] Q. Zhu, et al. (2006), Fast human detection using a cascade of histograms of oriented gradients, Proc. IEEE Conf. on Computer Vision & Pattern recognition.

[4] Y. Yamauchi, H. Fujiyoshi, T. Yamashita (2008), Human detection by boosting-based co-occurrence feature representation, Meeting on Image Recognition and Understanding.

[5] Y. Nakashima, J. K. Tan, S. Ishikawa, T. Morie (2011), Detecting a human body direction using multiple-HOG, FAST-zero'11.