A Pedestrian Detection Method Using the Extension of the HOG Feature

Yuuki Nakashima, Joo Kooi Tan, Hyoungseop Kim, Seiji Ishikawa
Department of Mechanical and Control Engineering
Kyushu Institute of Technology
Kitakyushu, Japan
{nakashima, etheltan, ishikawa}@ss10.cntl.kyutech.ac.jp

***Abstract*— Development of an ITS (Intelligent Transport System) has drawn much attention from computer vision community in recent years. In particular, various techniques for detecting pedestrians automatically have been proposed by many researchers. Among them, the HOG feature proposed by Dalal & Triggs has gained much interest in the pedestrian detection. However, previous methods including the original HOG feature have not achieved satisfactory detection rates.**

In this paper, we propose an extension of the HOG feature, i.e., flexible choice of the number of bins and automatic definition of a cell size and a block size by parameterizing their scales. By comparative experiments, it was confirmed that the proposed method outperforms the previous methods in the performance of pedestrian detection.

Keywords—pedestrian detection, HOG, Real-Adaboost

I. Introduction

ITS (Intelligent Transport System) has attracted much attention in recent years. This is a generic name of a system for the sensors mounted on a vehicle to realize safety of traffic environments. Many researchers have developed and proposed road lane detection methods [1], vehicle detection methods [2,3] and traffic sign recognition methods [4,5] by introducing various image processing techniques into the ITS problems. In particular, techniques for detecting pedestrians automatically [6,7,8] are important to protect drivers and pedestrians from traffic accidents.

The pedestrian detection is a difficult problem because of a variety of poses, clothes, backgrounds, illumination change in a real environment, etc. For solving this problem, a novel feature called HOG (Histograms of Oriented Gradients) feature has been proposed by Dalal & Triggs [8]. Since the HOG feature well represents the shape of a human, it has been improved in various ways by many researchers [9, 10, 11, 12, 13, 14]. Hou et al. proposed EHOG(Extend HOG)[9] which integrates some values of bins. Zhu et al. [10] proposed the HOG feature based on variable block sizes. Zeng et al. [11] proposed multi-level HOG feature which employs a variable cell size. Mitsui et al. [12] proposed Joint-HOG that uses co-occurrence of low-level HOG feature selected by Real-AdaBoost. Watanabe et al. also proposed Co-HOG[13] feature which is co-occurrence feature based on the HOG feature. The method using the HOG feature with Lab color space was proposed as well by Gall et al. [14].

However, it cannot be claimed that these previous methods are able to select effective parameters employed in the HOG feature. The HOG feature has the following 3 main parameters.

- **The cell size**: A cell, on which the HOG is defined, is a rectangular region containing some pixels on an image and is specified by the top-left position (x,y) and the width w and the height h of the rectangle. In [11], some rectangular sizes are employed, but it does not state how to optimize the size automatically.
- **The block size**: A block is a rectangle composed of a set of cells, and all the histograms the cells provide are normalized in the block to reduce the negative influence of illumination change. Its location and the size are defined by (x',y') and (w',h'), respectively. In [10], several sizes of blocks are used independently for the normalization.

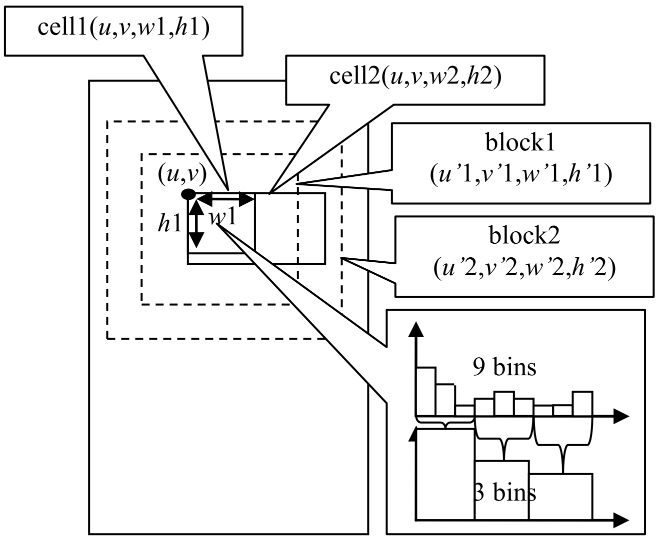


Fig. 1. The extension of the HOG feature.

- **The number of bins**: Only a fixed number of bins is considered for a histogram representation in the original HOG feature such as 9 bins. On the other hand, some other number of bins are employed in [9], such as a 3-bin histogram whose bin is a sum of three successive bins in 9-bin representation, in addition to a 9-bin histogram.

Those previous methods [9,10,11,12,13,14] consider variations in the above parameter sizes/number, but they concentrate in a single parameter and not all the parameters are taken into account simultaneously.

In this paper, we introduce the formulation which includes all extensions (Fig.1) of these topics on the parameters. The proposed method selects effective features for detecting a pedestrian from high dimensional feature. Experimental results show effectiveness of the proposed method compared to previous ones which employ part of the extension. The proposed method realizes high accuracy pedestrian detection rate.

II. Machine Learning for pedestrian detection

In this section we describe how to calculate the extension of the HOG feature and select effective features from those extended features. A single region of a cell (e.g., 5x5[pixel]), a single block for normalization (e.g., 3x3[cell]), a single number of bins (e.g., 9 bins) are introduced in the original HOG feature. The classifier is built using the original HOG feature and a linear SVM (Support Vector Machine). However it may be reasonable to consider the optimal cell size, the optimal block size and the optimal number of bins with each cell. This optimization is taken into account in the proposed method.

A. The extension of the HOG feature

The idea of the extension of the HOG feature is shown in Fig.1. Let $f(u,v)$ denote the pixel intensity at coordinates (u, v).

$$\begin{aligned} f_u(u,v) &= f(u+1,v) - f(u-1,v), \\ f_v(u,v) &= f(u,v+1) - f(u,v-1). \end{aligned} \tag{1}$$

Here f_u and f_v denotes the u and v components of the image gradient. The magnitude $m(u,v)$ and the orientation $\theta(u,v)$ of the gradient at coordinates (u,v) are calculated by

$$\begin{aligned} m(u,v) &= \sqrt{f_u(u,v)^2 + f_v(u,v)^2}, \\ \theta(u,v) &= \tan^{-1}\left(f_v(u,v) / f_u(u,v)\right). \end{aligned} \tag{2}$$

By letting B (e.g., B=1,3,9), $Cell_{wh}$ (e.g., w[pixel]$\times h$[pixel] =5×5, 5×10, 10×5), R_{pq} (e.g. p[cell]$\times q$[cell] = 1×1, 2×2, 3×3) and l be the number of bins, the cell size at the coordinate (u,v), the rectangular size of a block and the index of the bin, respectively, the extension of the HOG feature $x^{(u,v,w,h,p,q,l)}$ is calculated by the following equation;

$$x^{(u,v,w,h,p,q,l)} = \frac{\sum_{(i,j)\in Cell_{wh}} m(i,j)}{\sum_{(i,j)\in R_{pq}} m(i,j)}, \quad \text{if } (l-1)\pi / B < \theta(i,j) \le l\pi / B. \tag{3}$$

The center pixel of the block is located at the center pixel of the cell. These features are enumerated into a feature vector $\boldsymbol{x}$. If we set these parameters such as these examples to 30×60 [pixel] image, the number of features is 154,362={(30-5+1) × (60-5+1)+(30-5+1) × (60-10+1)+(30-10+1) × (60-5+1)}$\times3\times$(1+3+9).

B. Real-AdaBoost

The classifier is built using Real-AdaBoost [15]. The method can select effective features for detecting a pedestrian. The dimensional reduction of a feature space gives a positive influence for generalization ability.

Given N training samples, the feature vector of the nth training sample $\boldsymbol{x}_n$ and its class label (-1 or 1) y_n, a strong classifier $H(\boldsymbol{x})$ is trained in the following way.

1) Initialization of the sample weight D_0

Sample weight $D_0(n)$ is initialized by

$$D_0(n) = 1/N. \tag{4}$$

2) Selection of weak classifiers

The weak classifier h_t at a training step t (t=0,... ,T) is selected in the following way.

a) PDF(Probability Density Function) calculation

The PDF of positive samples (class label=1) W^+ and that of negative samples (class label= -1) W^- are calculated by

$$\begin{aligned} W^+(i) &= \sum_{\mathrm{BIN}(x)=i \wedge y_n=1} D(n), \\ W^-(i) &= \sum_{\mathrm{BIN}(x)=i \wedge y_n=-1} D(n). \end{aligned} \tag{5}$$

Here BIN(x) is a translation function to the bin number i .

b) A weak classifier evaluation

Evaluation value z is calculated by

$$z = \sum_i \sqrt{W^+(i) W^-(i)}. \tag{6}$$

The smaller the evaluation value z is, the more separable the weak classifier is. The algorithm selects h_t having the smallest z.

c) A selected weak classifier calculation

The output of the selected weak classifier h_t is calculated by

$$h_t(x_t) = \frac{1}{2}\ln\frac{W^{+}(BIN(x_t)) + \varepsilon}{W^{-}(BIN(x_t)) + \varepsilon}. \tag{7}$$

A constant ε denotes a small positive real value (ε <<1) to avoid zero division.

3) Update of sample weight D_t

With the output of the weak classifier h_t, the sample weight D_t is calculated and normalized by

$$D_{t+1}(n) = D_t(n)\exp(-y_n h_t(x_{nt})),$$
$$D_{t+1}(n) = D_{t+1}(n) \Big/ \sum_{n=0}^{N} D_{t+1}(n). \tag{8}$$

4) Constrcution of a strong classifer

The processes 1), 2) and 3) are iterated T times. Finally, $H(\boldsymbol{x})$ is defined as

$$H(\boldsymbol{x}) = \mathrm{sign}(\sum_{t=1}^{T} h_t(x_t)). \tag{9}$$

III. Experimental Results

A. *Effectiveness of the extension*

We compared the different extensions (See Table I). The method denoted by Comp_1 adopts three kinds of bins: The method Comp_2 further includes three cell sizes: Finally the proposed method further employs three block sizes. These three methods are compared to the original HOG method which employs unique numbers of bins, cell sizes and block sizes.

In this experiment, Real-AdaBoost is employed to the 4 methods (Table I) because of confirming the effectiveness given by the feature extension. We used 64 bins to the PDF calculation and 1,000 iterations on Real-AdaBoost. We performed the experiment on the INRIA person dataset. Dataset setting is shown in TABLE II. The number of positive samples is less than that of negative samples. In machine learning, the more samples we use, the better results we will achieve. Positive samples are already clipped in INRIA Person data set. On the other hand, negative samples are given as 1 frame image (for example 320 x 240). For this reason, negative samples are clipped at random. We evaluated the error rate [%] vs. the number of weak classifiers.

The experimental result is given in Fig. 2. One can see that the methods Comp_1 (the yellow line in Fig. 2) and Comp_2 (the red line) indicate better results compared to the original HOG method (the brown line). But the proposed method (the blue line) shows much better performance than the other three methods with every number of weak classifiers.

B. *Comparsion with the previous method*

We compared the proposed method with the original HOG feature+Linear-SVM. Experimental setting is the same as TABLE I and II.

The experimental result (See Fig. 3 and TABLE III) shows that the proposed method is superior to the previous one in terms of detection rate. The result is provided by DET (Detection Error Trade-off) curves as shown in Fig. 3 and AUC (Area Under the DET Curve) as shown in TABLE III. The numbers 750 and 300 are the number of weak classifiers.

IV. Discussion

The experimental result given in Fig. 2 shows that the more kinds of extension we use, the better result we get. Selected features per topic of extensions are shown in Fig.4. All of the number of bins, cell sizes and block sizes are selected. This shows that every extension has effective features for detecting a pedestrian. We think that the detection rate has been improved for this reason. In particular, Fig.4 (a)

TABLE I. Parameters Setting

Method index	***Extensions***		
	Number of bins	Cell size [pixel]	Block size [cell]
Original HOG	9	5×5	3×3
Comp_1	1,3,9	5×5	1×1
Comp_2	1,3,9	5×5, 5×10,10×5	1×1
Proposed method	1,3,9	5×5, 5×10,10×5	1×1,2×2,3×3

TABLE II. Dataset

Dataset	INRIA Person Dataset	
	Training	Test
Positive samples	2416	1126
Negative samples	6000 (clipped at random)	3000 (clipped at random)
Image size[pixel]	30 × 60	

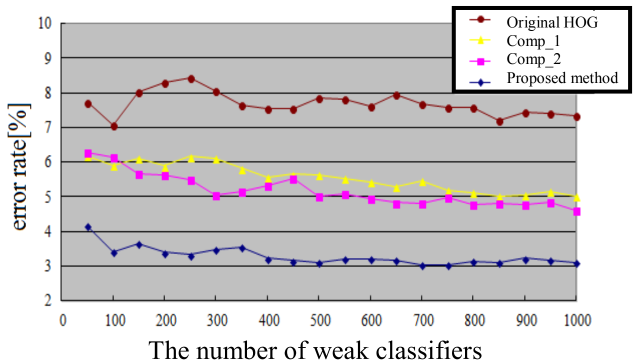


Fig. 2. Experimental results.

TABLE III. AUC(AREA UNDER THE DET CURVE)

Method	AUC
Proposed method (750)	0.007473
Original HOG	0.009785
Proposed method (300)	0.010385

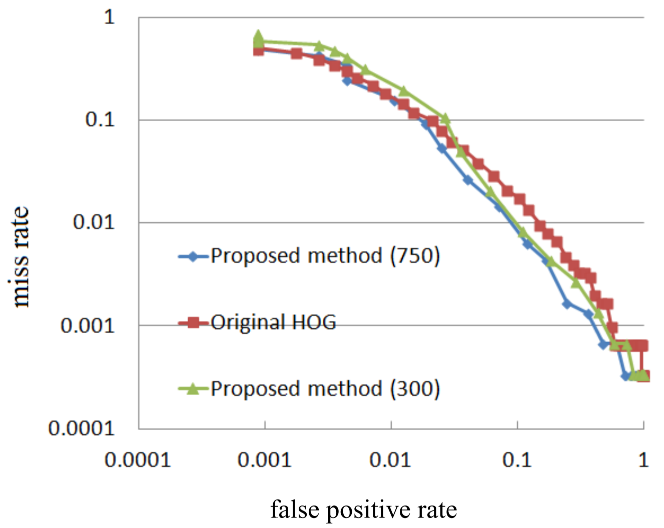


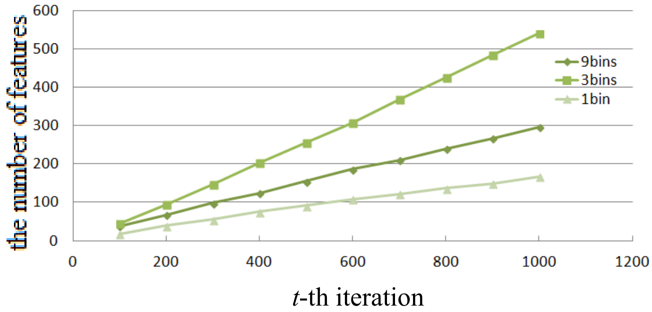
Fig. 3. Results on comparision with the previous method.

is interesting. The experimental result in [8] shows that the effective number of bins is 9. However, the best selected number of bins is 3 with the proposed method as shown in Fig.4(a). The proposed extension of the HOG feature has brought the change in the effective number of bins.

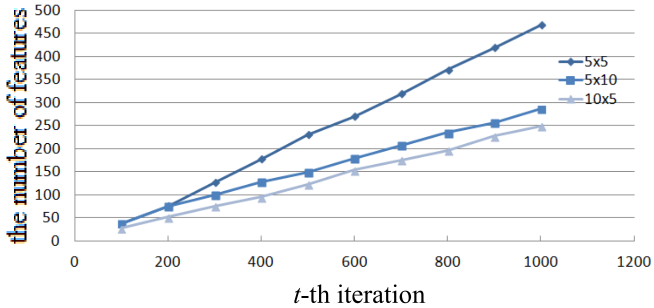
In the following discussion, we focus on the dimensionality reduction. The original HOG feature had 3240-dimensional feature vector in the performed experiment. On the contrary, Fig.3 and TABLE III show that the proposed method with 300-dimensional feature vector achieved the miss rate similar to that of the original HOG feature. In this way, 90% reduction of the dimension of the feature vector has been achieved to get a similar AUC. Even the proposed method employing 750 dimensional feature vector achieves less number of the feature dimension.

TABLE IV. CALCULATITON TIME

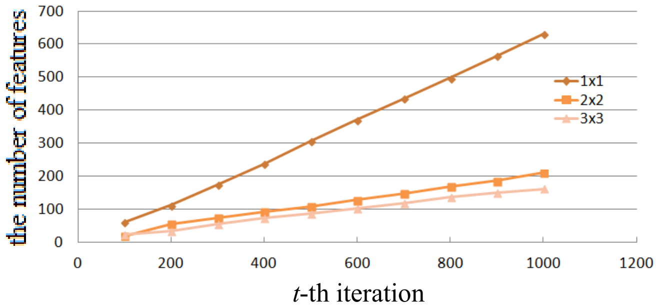
Method	Micro sec
Proposed method (750)	34
Original HOG	150



(a)



(b)



(c)

Fig. 4. The number of features: (a) The number of bins, (b) The cell sizes, (c): The block sizes.

We discuss the processing time (TABLE IV). The computer settings are shown in TABLE V. The result shows that the proposed method is roughly 4 times faster than the previous one. The proposed method can reduce the number of feature dimensions d. Both of RealAdaBoost and Liner-SVM are executed in O(d) calculation time. Feature vector calculation is also executed in O(d). For this reason, if we reduce d, the calculation time of classifying are reduced in O(d).

TABLE V. COMPUTER SETTINGS

CPU	Core i7 3.5GHz
Memory	32GB
OS	Windows7 Pro
Compiler	VC++11

V. CONCLUSION

We proposed a pedestrian detection method using the extension of the HOG features and Real-AdaBoost. Experimental results show that the proposed method is superior to the previous method. We are going to apply the proposed method to an ITS.

ACKNOWLEDGMENT

This study is supported by JSPS KAKENHI under GRANT-in-Aid for scientific Research(25350477)

REFERENCES

[1] Masakatsu Higashikubo, Yasuo Ogiuchi, Yuuki Ono, Takio Kurita, Kenji Nishida, Hiroaki Inayoshi, Reiji Arata, "Measurement of Vehicles and Motorcycles with Detection by Support Vector Machines and Pair-Pixel Feature Tracking", Symposium on Sensing via Image Information(SSII), CD-R, IS4-02, 2010.

[2] F. Ren, Jinshendg Huang, Ruyi Jiang, and Reinhard Klette,"Lane detection on the iPhone", Multimedia Imaging Report 30(The University of Auckland), 2008.

[3] Yue Wang, Dinggang Shen and Eam Khwang Teoh,"Lane detection using catmull-rom spline", Proc. of the 1998 IEEE International Conference on Intelligent Vehicles, Vol. 1, pp. 51-57, 1998.

[4] Masanari Takagi, Hironobu Fujiyoshi, "Road sign recognition using SIFT feature", Symposium on Sensing via Image Information(SSII), CD-R, LD2-06, 2007.

[5] D. M. Gavrila,"Multi-feature hierarchical template matching using distance transforms", Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 439-444, 1998.

[6] Kousuke Matsushima Zhencheng Hu, Keiichi Uchimura,"Pedestrian recognition using stereo sensor", Information Processing Society of Japan(IPSJ) SIG Technical Reports ITS, pp. 49-54, 2006.

[7] D. M. Gavrila,"Pedestrian detection from moving vehicle", Proc. European Conference on Computer Vision (ECCV), pp.37-49, 2000.

[8] N. Dalal and Bill Triggs,"Histograms of oriented gradients for human detection", Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 886-893, 2005.

[9] C. Hou, H. Ai, Shiong Lao,"Multiview Pedestrian Detection Based on Vetor Boosting", Proc. Asian Conference on Computer Vision(ACCV), pp. 210-219, 2007.

[10] Qiang Zhu, Shai Avidan, Mei-Chen Yeh, Kwang-Ting Cheng,"Fast human detection using a cascade of histograms of oriented gradients", Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1491-1498, 2005.

[11] Chengbin Zeng, Huadong Ma.,"Robust head-shoulder detection by PCA-based multilevel HOG-LBP detector for people counting", Proc. IEEE Conference on International Conference on Pattern Recognition, (ICPR), pp. 2069-2072, 2010.

[12] Tomokazu Mitsui, Yuji Yamauchi, Hironobu Fujiyoshi, "Object detection by Two-Stage boosting with joint features", Jornal on The Institute of Electronics, Information and Communication Engineers(IEICE), Vol. J92-D, No. 9, pp. 1591-1601, 2009.

[13] Tomoki Watanabe, Satoshi Ito and Kentaro Yokoi,"Co-occurrence histograms of oriented gradients for human detection, IPSJ Transactions on Computer Vision and Application Vol. 2, pp. 39-47, 2010.

[14] Juergen Gall, Victor Lempitsky, "Class-specific hough forests for object detection", Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 20-25, 2009.

[15] R. E.Schapire, Y. Singer: "Improved boosting algorithms using confidence-rated predictions", Journal on Machine Learning, Vol. 37, pp.297-336,1999.