World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:11, No:6, 2017

Moving Object Detection Using Histogram of Uniformly Oriented Gradient

Wei-Jong Yang, Yu-Siang Su, Pau-Choo Chung, Jar-Ferr Yang

Open Science Index, Computer and Information Engineering Vol:11, No:6, 2017 publications.waset.org/10007186.pdf

Abstract**—**Moving object detection (MOD) is an important issue in advanced driver assistance systems (ADAS). There are two important moving objects, pedestrians and scooters in ADAS. In real-world systems, there exist two important challenges for MOD, including the computational complexity and the detection accuracy. The histogram of oriented gradient (HOG) features can easily detect the edge of object without invariance to changes in illumination and shadowing. However, to reduce the execution time for real-time systems, the image size should be down sampled which would lead the outlier influence to increase. For this reason, we propose the histogram of uniformly-oriented gradient (HUG) features to get better accurate description of the contour of human body. In the testing phase, the support vector machine (SVM) with linear kernel function is involved. Experimental results show the correctness and effectiveness of the proposed method. With SVM classifiers, the real testing results show the proposed HUG features achieve better than classification performance than the HOG ones.

Keywords**—**Moving object detection, histogram of oriented gradient histogram of oriented gradient, histogram of uniformly-oriented gradient, linear support vector machine.

I. Introduction

MOD has wide application domains in computer vision. It can be used in surveillance, robotics, intelligent vehicles, etc. However, owing to the large variations in pedestrians' pose and clothing, as well as the varying background and illumination, it is still a challenging task.

In recent years, with the development of machine learning, pattern classification approaches have been shown to achieve successful results in MOD. These approaches mainly include two key points: Feature extraction and classifiers. In the feature extraction, positive and negative sub-images are densely scanned from the top left to the bottom right with sliding windows in the region of interested in capture images. The dominant features such as edges, patches and shapes are extracted from the positive and negative sub-images. Then, these features are used to train a classifier to achieve a proper classifier. During the testing phase, the entire input image in sub-image bases is scanned by feature extraction associated with the classifier to detect the moving objects.

In the past decades, several researchers concentrated on object detection in the domain of traffic surveillance. The main differences between these researches are feature selection and classification methods. For example, Haar-like features [1], [2] and HOG features [3], [4] are frequently applied. Hota et al. [5] created a cascaded classifier combining these two features. A new set of image strip features is proposed by Zheng et al. [6] to model the structural characteristics of cars within a boosting framework. Dalal and Triggs [7] present a human detection algorithm using grids of HOG descriptor and linear SVM [8]-[10] as the baseline classifier. Their detection results are excellent at that time. Later, many researchers tried to improve this approach, which combines SVM classification with HOG features. Zhu et al. [11] integrated a cascade-of-rejecters approach using AdaBoost algorithm for feature selection with modified HOG features to reduce the detection time and improve the accuracy. More recently, deep learning approaches such as convolution neural network (CNN networks) had been widely used [12]. This work learns the complex hierarchical features, saliency detection and body parts detection in one deep network. However, the CNN network is costly. In order to achieve real time detection, the researchers used an SVM classifier for both training and testing phase.

In this paper, we presented a powerful pedestrian detector by the HUG feature in down scale videos, while the linear SVM [8] is used to train the pedestrian classifier. The results presented in our dataset show that our detector is more discriminative and robust than the state-of-the-art algorithms.

The rest of the paper is organized into four sections. Section II briefly introduces the related work. Section III describes the proposed system. Section IV explains the settings of our MOD system and shows the performance for the proposed method by experiments. Finally, Section V is the conclusion.

II. HOG Feature and SVM Classifier

A. HOG

The HOG is a useful feature, as proposed in [7] for pedestrian detection, which essentially compute the gray gradient statistics of local image regions. The HOG based algorithms could be extended for face recognition [13], [14]. Assume the pixel located at (x, y) has a gray value $L(x, y)$, the gradients along x and y directions are respectively given as:

$$g_x(x,y) = L(x+1,y) - L(x-1,y) \quad (1)$$

and

$$g_y(x,y) = L(x,y+1) - L(x,y-1) \quad (2)$$

The gradient magnitude $g(x, y)$, gradient direction $\theta(x, y)$ are then respectively computed as:

W. J. Yang, Y. H. Su, P. C. Chang and J. F. Yang are with the Institute of Computer and Communication Engineering, the Department of Electrical Engineering, the National Cheng Kung University, Tainan 701, Taiwan (phone: +886-916727713; e-mail: weijong@hotmail.com, jaysu455@gmail.com, pcchung@ee.ncku.edu.tw, jefyang@mail.ncku.edu.tw).

$$g(x,y) = (g_x^2(x,y) + g_y^2(x,y))^{1/2} \tag{3}$$

and

$$\theta(x,y) = \tan^{-1}(g_x(x,y)/g_y(x,y)) \tag{4}$$

As stated in [7], the computation of gradient only refers to the neighborhood of the center pixel p in an 8×8 window. Each detection window could be divided into cells of 8×8 pixels and each group of 2×2 cells is integrated into a block in a sliding fashion. Each cell could consist of a 9-bin HOG and each block contains a concatenated vector of all its cells. Each block is thus represented by a feature vector that is normalized to an L2 unit length. Each 64×128 detection window is represented by 7×15 blocks, giving a total of 3780 features per detection window. These features are then used to train a linear SVM classifier.

B. SVM

The SVM [8]-[10] can achieve optimal classification of linearly separable data. For a linear SVM with the training samples, $\{(x_i, y_i) | 1 \le i \le N)\}$, where x_i is the i^{th} instance sample, y_i is the corresponding category labels, its decision surface equation can be expressed as:

$$\omega \cdot x + b = 0 \tag{5}$$

where x is the input vector, ω is the dynamically variable weight vector, and b is the offset. In order to find an optimal classification result, the classifier tried to find an optimal hyper plane according to (5), which can not only separate two classes correctly but also maximize the between-class distance. Accordingly, support vectors refer to the training sample points located in the classification boundaries, which are the key elements of the training sample set. Based on these theories and concepts [15]-[18], we classify the input samples by using:

$$f(x) = \operatorname{sgn}\left\{\sum_{i=1}^{k} \alpha_i^* y_i^* (x_i^* \cdot x) + b^*\right\} \tag{6}$$

where α_i^* is the weight coefficient corresponding to the support vector x_i.

III. The Proposed System

A. Overview of the Proposed System

The presented MOD system, as shown in Fig. 1, involves three major parts: 1) Selecting of regions of interest, 2) Calculating HUG, and 3) Performing the SVM for classification.

In order to reduce computation time, we first scale down the resolution of the input video from 1280×720 to 320×180 and then we choose regions of interest (ROI) which are at x direction from 40 pixels to 280 pixels and at y direction from 15 pixels to 130 pixels to search the moving objects. Also, we set the red region which restrict the detection window should be within the region and the green lines which simulate the virtual traffic lanes. Then, we define the dangerous zone which is the overlap of red region and green line, as shown in Fig. 2. If the detection window within the ROI is activated, we would treat them as moving objects. The design of ROI can remove some misdetection objects in order to reduce the false alarm rate.

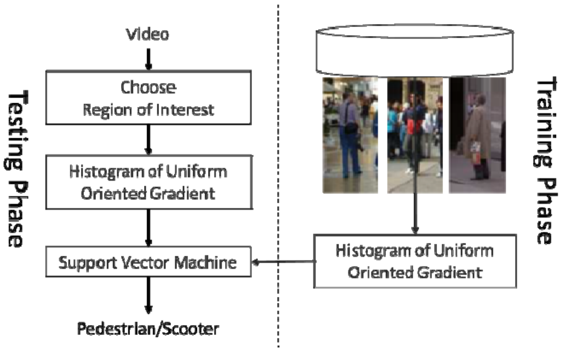

Video
Choose Region of Interest
Histogram of Uniform Oriented Gradient
Support Vector Machine
Pedestrian/Scooter
Testing Phase
Training Phase
Histogram of Uniform Oriented Gradient

Fig. 1 Flow diagram of the proposed MOD system

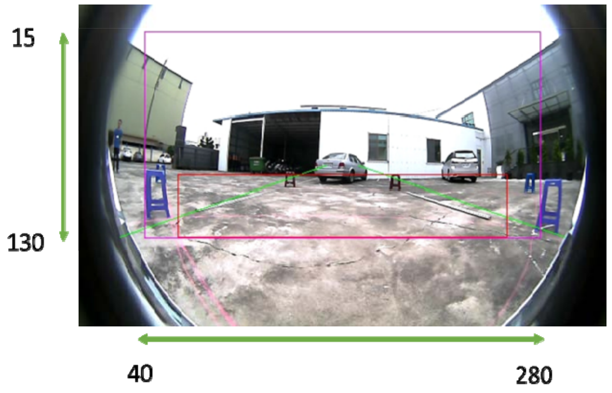


Fig. 2 ROI and dangerous zone

B. HUG

To compute the robust gradient magnitude and gradient direction, we modify the computations of direction-x and y gradients by adopting a one-dimensional center gradient operator weighted by [-0.5, -0.5, 0, 0.5, 0.5] to achieve precise horizontal gradient and vertical gradient are defined as:

$$\begin{aligned}\hat{g}_x(x,y) = {} & \tfrac{1}{2}[L(x+2,y) + L(x+1,y)] - \\ & \tfrac{1}{2}[L(x-2,y) + L(x-1,y)]\end{aligned} \tag{7}$$

and

$$\begin{aligned}\hat{g}_y(x,y) = {} & \tfrac{1}{2}[L(x,y+2) + L(x,y+1)] - \\ & \tfrac{1}{2}[L(x,y-2) + L(x,y-1)]\end{aligned} \tag{8}$$

Thus, the modified gradient magnitude $\hat{g}(x,y)$, gradient direction $\hat{\theta}(x,y)$ are then respectively computed as:

$$\hat{g}(x,y) = (\hat{g}_x^2(x,y) + \hat{g}_y^2(x,y))^{1/2} \tag{9}$$

and

$$\hat{\theta}(x,y)=\tan^{-1}(\hat{g}_x(x,y)/\hat{g}_y(x,y)) \tag{10}$$

Afterwards, for each cell consisting of 8×8 pixels the gradient magnitudes are weighted according to their gradient direction and then accumulated in their bin direction, resulting in a gradient histogram. We choose a directional range from $0\sim180°$ for HUG features of moving objects, with eight segments of $22.5°$ and correspondingly 8 bin directions. Then, four neighboring cells build up as a block. The blocks are scanned through all the ROI with sliding windows in overlapped fashions. Finally, to reduce the influence of illumination and background variation, normalization is essential. We use L2-norm for normalization:

$$v/\|v\|_2^2+\varepsilon^2 \rightarrow v \tag{11}$$

where v stands for the characterization vector and εis a small constant.

The sizes of input sub-images are 40×80, 32×64 and 24×48 separately. To compute HUG features, we adopt 50 (5×10) cells, 32 (4×8) cells and 18 (3×6) cells, respectively. After that, four consecutive cells construct of a block. A 32-element vector is used to describe the local gradient distribution. Finally, with 36 (4×9) blocks, 21 (3×7) blocks and 10 (2×5) blocks, 1152-element vector, 672-element vector and 320-element vector can be respectively built to describe the whole gradient information of the image. Concatenating the histograms from all the sub-regions gives the final HUG feature vector, as illustrated in Fig. 3.

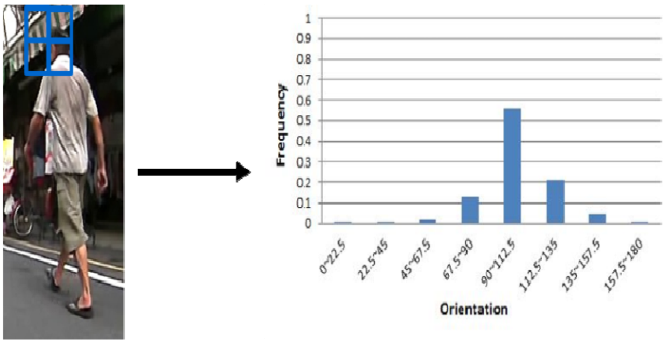


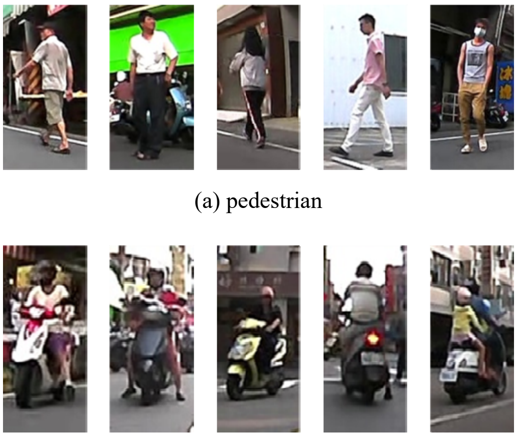
Fig. 3 Extraction of HUG feature from subregions of the image

One distinct advantage of the HUG with SVM classifier over traditional neural networks is that it achieves better generalization performance and is easily implemented into real time systems. Compared with the popular Adaboost classifiers, the SVM is slower in the test stage. However, the training speed of SVM is much faster than that of Adaboost classifiers.

IV. Experimental Results

A. Database

Our pedestrian training database contains three kinds of different scale of resolutions which are 40×80, 32×64 and 24×48. Our database contains 1044 standing pedestrian positive samples with various aspects and poses. Some of these sub-images are from the public downloadable MIT pedestrian dataset [8], while the sub-images of the scooter data set are captured by our camera systems. The negative examples are cropped from videos. For the scooter training data, we collected 235 positive samples and 136 negative samples. Some of the positive training samples are shown in Fig. 4.



(a) pedestrian

(b) scooter

Fig. 4 Some selected positive samples from training database

B. Training and Evaluation

The proposed system is developed in Visual c++ and OpenCV function library. For the hardware systems, we use the computer with Intel I-7 CPU, NVIDIA GeForce GTX 460, 8G RAM and fisheye camera.

We use the database as described in Subsection IV-B, as the training set. In our experiments, we detect the pedestrian and scooter only in ROI and evaluated the recognition rates and false alarm rates. We also compared the classification performance between HOG with SVM and HUG with SVM. For fair comparisons, their features are all extracted on blocks with same block sizes and with same cell sizes. Thus, their dimensions are 1152, 672 and 320. We tested on two different scenes as urban road and company road, as shown in Fig. 5.

The recognition rates, false alarm rates and speed are shown in Table I. We can observe that the proposed method performs better than the HOG+SVM in both of pedestrian detection (PD) and scooter detection (SD) in urban road and company road. Besides, the false alarm rate in the proposed method also achieves the best expected result. Furthermore, it can achieve real-time and can be easily implemented by the hardware system as well.

TABLE I
The Performance of PD and SD with HOG and HUG features

Detected Objects	Methods	Testing video	Recognition rate	False alarm rate	Frame rate
Pedestrian Detection	HOG	Urban Road	60.00%	4.35%	90fps
	HUG		80.00%	0.76%	90fps
	HOG	Company Road	100.00%	0.07%	90fps
	HUG		100.00%	0.00%	90fps
Scooter Detection	HOG	Urban Road	80.65%	0.69%	90fps
	HUG		80.65%	0.54%	90fps

Fig. 5 Urban road and company road

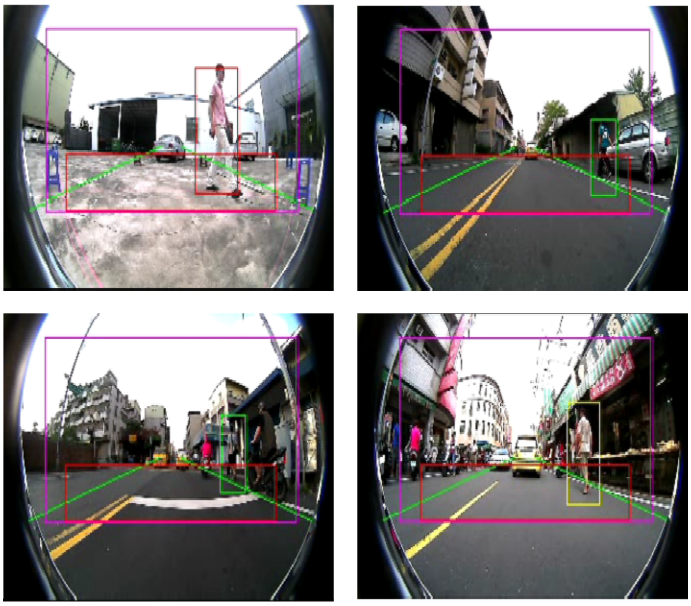


Fig. 6 Pedestrian detection results

The detection results by HUG feature with the SVM classifier, including pedestrian and scooter, are shown in Figs. 6 and 7, respectively, where we use three colors, red, yellow and green to stand for three different detection windows.

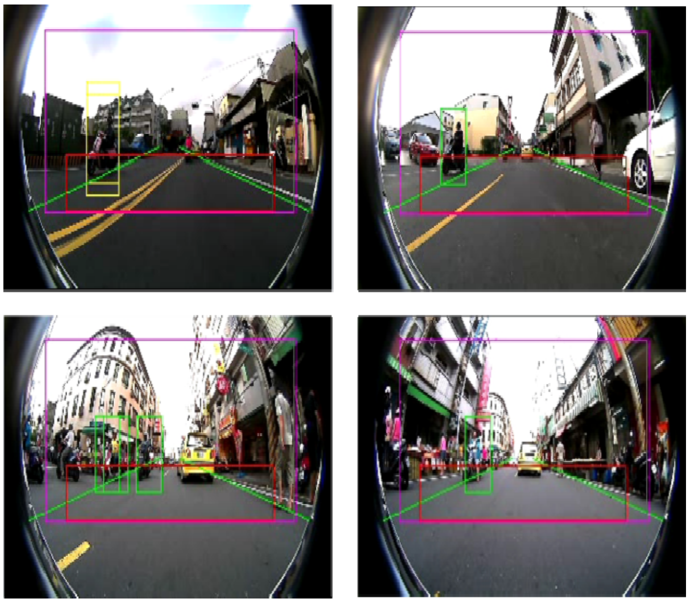


Fig. 7 Scooter detection results

V. Conclusions

In this paper, a MOD system, which could perform pedestrian detection and scooter detection, is presented. We presented the new features extraction method, HUG for down scaled video to reduce the computation. The classification is based on a linear SVM classifier. In the experimental results, we compared with the original HOG with SVM classifier. The proposed system was tested in two situations including city road and urban road. The experimental results show that the proposed system with the HUG feature achieves a favorable detection rate and false alarm rate. It also showed that the system can achieve real-time detection which is able to be implemented on hardware easily.

References

[1] S. Zhang, C.Bauckhage, and A. B. Cremers. "Informed haar-like features improve pedestrian detection." *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014.

[2] P. Viola, and M. Jones. "Rapid object detection using a boosted cascade of simple features." *Proc. of Computer Vision and Pattern Recognition, CVPR 2001,* vol. 1, 2001.

[3] F. Suaard, A. Rakotomarmonjy, A. Bensrhair, and A. Broggi, “Pedestrian detection using infrared images and histograms of oriented gradients,” *Prof. of Intelligent Vehicles Symposium*, June 2006.

[4] T. Watanabe, S. Ito, and K. Yokoi, “Co-occurrence histograms of oriented gradients for pedestrian detection,” *Proc. of Pacific-Rim Symposium on Image and Video Technology*, pp.37-47, 2009

[5] R. N. Hota, K. Jonna, and P. R. Krishna. "On-road vehicle detection by cascaded classifiers." *Proc. of the Third Annual ACM Bangalore Conference*. ACM, 2010.

[6] W. Zheng, and L.Liang. "Fast car detection using image strip features." In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009. I*IEEE, 2009.

[7] N. Dalal and B. Triggs. "Histograms of oriented gradients for human detection." In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, 2005.

[8] J. A. K. Suykens and J. Vandewalle, “Least Square Support Vector Machine Classifiers,” Neural Network Letter, vol. 9, pp.293-300, 1999.

[9] C. C. Chung, C. J. Lin, “LibSVM: A Library for Support Vector Machines,” 2001, http://www.csie.ntu.edu.tw/~cjlin/libsvm/ (access at 2016/4/20)

[10] C. Cortes and V. Vapnik, “Support vector networks,” *Machine Learning*, vol. 20, pp.273-297, 1995

[11] Q. Zhu, A. Avidan, M.-C. Yeh, and K.-T. Cheng, "Fast human detection using a cascade of histograms of oriented gradients." In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. vol. 2, 2006.

[12] P. Luo, Y. Tian, X. Wang, and X. Tang "Switchable deep network for pedestrian detection." *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014.

[13] O. Deniz, G. Bueno, J. Salido, and F. D. la Torre, “Face recognition using histrogram of oriented gradients,” *Pattern Recognition Letter*, vol. 32, no. 12 pp.159-1602, 2011.

[14] C.-Y. Su and J. F. Yang, “Histogram of gradient phase: A new local descriptor for face recognition,” *IET Computer Vision*, vol. 8, no. 6, pp.556-567, December 2014.

[15] R. E. Osuna, R. Freund, and F. Girosit. "Training support vector machines: an application to face detection." In *Proc. of IEEE Computer Vision and Pattern Recognition,* 1997.

[16] V. Vapnik. *The Nature of Statistical Learning Theory*. New York Springer-Verlag, 1995.

[17] C. Papageorgiou and T. Poggio, "A trainable system for object detection." *International Journal of Computer Vision*, vol.38.1, pp.15-33, 2000.

[18] P. Viola, M. J. Jones, and D. Snow. "Detecting pedestrians using patterns of motion and appearance." *International Journal of Computer Vision* 63.2: 153-161, 2005.