An Enhanced Histogram of Oriented Gradients for Pedestrian Detection

***Abstract*—**The outstanding Histogram-of-Oriented-Gradients (HOG) feature proposed by Dalal and Triggs is a state-of-art technique for pedestrian detection, and it is usually applied with a linear support vector machine (SVM) in a sliding-window framework. Most other algorithms for pedestrian detection use HOG as the basic feature, and combine other features with HOG to form the feature set. Hence, the HOG feature is actually the most efficient and fundamental feature for pedestrian detection. However, the HOG feature cannot adequately handle scale variation of pedestrians. In addition, simply downsampling an image into a different scale, or decomposing via wavelet into multi-resolution subimages, calculating their HOG feature and combining them cannot enhance performance. Therefore, in this paper, based on the idea of multi-resolution feature descriptors, we propose a new robust edge feature referred to as Enhanced HOG (eHOG). It is a complementary descriptor for**

Yong Zhao
The Key Lab of Integrated Microsystems, Peking University, Shenzhen Graduate School, Shenzhen, Guangdong Province, China

Yongjun Zhang
College of Computer Science and Technology, Guizhou University, Guiyang, Guizhou Province, China
E-mail: zyj6667@126.com

Ruzhong Cheng, Daimeng Wei, and Guoliang Li
The Key Lab of Integrated Microsystems, Peking University, Shenzhen Graduate School, Shenzhen, Guangdong Province, China

This work was supported by the Joint Fund of Department of Science and Technology of Guizhou Province and guizhou university under Grant: LH [2014]7635.

Digital Object Identifier 10.1109/MITS.2015.2427366
Date of publication: 24 July 2015

1939-1390/15©2015IEEE

Accurate pedestrian detection will have immediate and far-reaching impact on applications such as video surveillance, robotics and driver assistance.

the histograms-of-oriented-gradients feature. Though the extraction process of the eHOG descriptor is derived only from HOG itself, similar to the process of extracting edge information from the downscaling image, it retains much more information for the edge gradient than that of the original HOG, without significantly increasing the complexity of computation. In the INRIA pedestrian dataset, many experiments have been conducted with eHOG and HOG, and the results show that the proposed new feature consistently improves the detection rate more than the original HOG feature detector. Particularly, eHOG with a Histogram Intersection Kernel SVM (HIKSVM) classifier has greatly improved performance. These results suggest that eHOG may be a better substitute for HOG for pedestrian detection in many applications.

I. Introduction

Object detection is a fundamental problem in computer vision with wide applications. Since pedestrian detection is one of the most important topics in object detection, it has attracted much attention, and significant research has been carried out on this technology. However, detecting humans in images/videos is a challenging task owing to their variable appearance and the wide range of poses that they can adopt. Accurate pedestrian detection will have immediate and far-reaching impact on applications such as video surveillance, robotics and driver assistance. In this paper, a new feature is proposed for this purpose, which is inspired by the multi-resolution principle. The new feature is derived from Dalal and Triggs's [1] well-known HOG feature, and is referred to as the enhanced HOG feature (denoted as eHOG). HOG is based on the description of the pedestrian contour, and only extracts features that are sensitive to scale in fixed-size blocks. The enhanced HOG feature overcomes the disadvantages of original HOG to some extent. The experimental results demonstrate that the enhanced feature effectively improves the detection rate compared to the original HOG in the commonly used INRIA database without significantly increasing complexity.

The contribution of this paper is three-fold as follows:

1) A new gradient image-processing method is introduced, which refines the original gradient image to a certain smaller scale, but retains most gradient information of the original gradient image. The method enhances the relevance of reduced images. It is superior to the general compression method.
2) The reduction method described above is used to reduce the original image to gradient images of different sizes. The eHOG feature is extracted via the fixed-size blocks and blocks of variable sizes, which can enhance the description of the contour and is more robust to the scale of the pedestrian object.
3) The performance of the system depends on the effectiveness of feature descriptors and the accuracy of classification models. We compare the performance of feature descriptors with both linear-SVM and HIKSVM, respectively. Two evaluation methods, the False Positive per Window (FPPW) and the False Positive per Image (FPPI), are used to evaluate the performance of eHOG.

II. Related Work

It is hard to outperform the HOG descriptor by using other single features. However, other features are proposed together with HOG to provide some improved performance, which will enhance the detection rate. The cascaded features are presented repeatedly in pedestrian detection after the HOG feature. Wojek and Schiele [2] proposed the MULTIFTR feature combining Haar-like [3], shapelets[4], shape context [5] and HOG descriptors. Based on these works, Walk et al. [6] proposed the MULTIFTR+MOTION feature combining HOG, local color self-similarity and histogram of flow features [1]. Wu and Nevatia et al. [7] presented a boosting learning method using HOG, edgelet and covariance features. Watanabe et al. [8] used pairs of gradient orientations as units to builds histograms (CoHOG) for pedestrian detection, which can express complex object shapes with local and global distributions of gradient orientations. Wang et al. [9] proposed the HOG-LBP joint feature combining HOG and a local binary pattern (LBP) [10]. Based on HOG and LBP descriptors, [11] proposed a variant of an LBP feature called local ternary pattern (LTP). [12] added color information and implicit division to the HOG feature and exhibited better performance compared to the original method. Dollár et al. [13] proposed a feature extraction method of Haar-like functionality on the basis of paper [3]. They extracted Haar-like features in gray-scale images, color images (LUV), gradient images and directional gradient images, and a simple method called CHNFTRS was found to compute and integrate these different features. R.M. Anwer et el. [14] fed the opponent color space (OPP) in the baseline framework of Dalal et al. for human detection (based on RGB, HOG and linear SVM), They obtained better detection performance than by using RGB space. M.A. Rao et al. [15] challenged RGB

space with OPP, which compute the HOG on top of OPP, and then train and test the part-based human classier in possible differences among types of scenarios: indoor, urban and countryside. Their experiments demonstrated that the benefits of OPP with respect to RGB mainly come for indoor and countryside scenarios. Y. Socarrás et el. [16] used the higher-level information coming from image segmentation, for re-weighting the HOG descriptor for each one of the cells while it is computed, which will enhance human silhouette orientations, without explicitly computing such silhouettes, but using information not as local as the own gradient magnitude. J. Marin et al. [17] proposed a novel ensemble of local experts by means of a Random Forest ensemble. This method works with rich block-based representations such as HOG and LBP, in such a way that the same features are reused by the multiple local experts, so that no extra computational cost is needed with respect to a holistic method. [18] posited an improved method of pyramid scanning called FPDW that can estimate the neighboring scales feature by using a certain scale feature already extracted. In this way, it reduces the number of scanning scales and increases the detection speed.

The HOG feature has demonstrated good performance for pedestrian detection. Compared with variable-size blocks, fixed-size blocks might miss the information that is contained in other different size blocks. As a remedy, we add other size of blocks information with our downscaling method to the original HOG feature.

III. eHOG Formulation

The eHOG feature extraction process is derived from HOG. During the extraction process of the HOG feature, we propose a feature extraction method after edge detection. The method includes the original HOG feature and the enhanced HOG feature. As shown in Fig. 1(a) and (b), Fig. 1(b) is a further development of (a), and the extraction procedure of eHOG feature contains two parts: the extraction of the original HOG feature and the extraction of the enhanced HOG feature. Based on the transformation

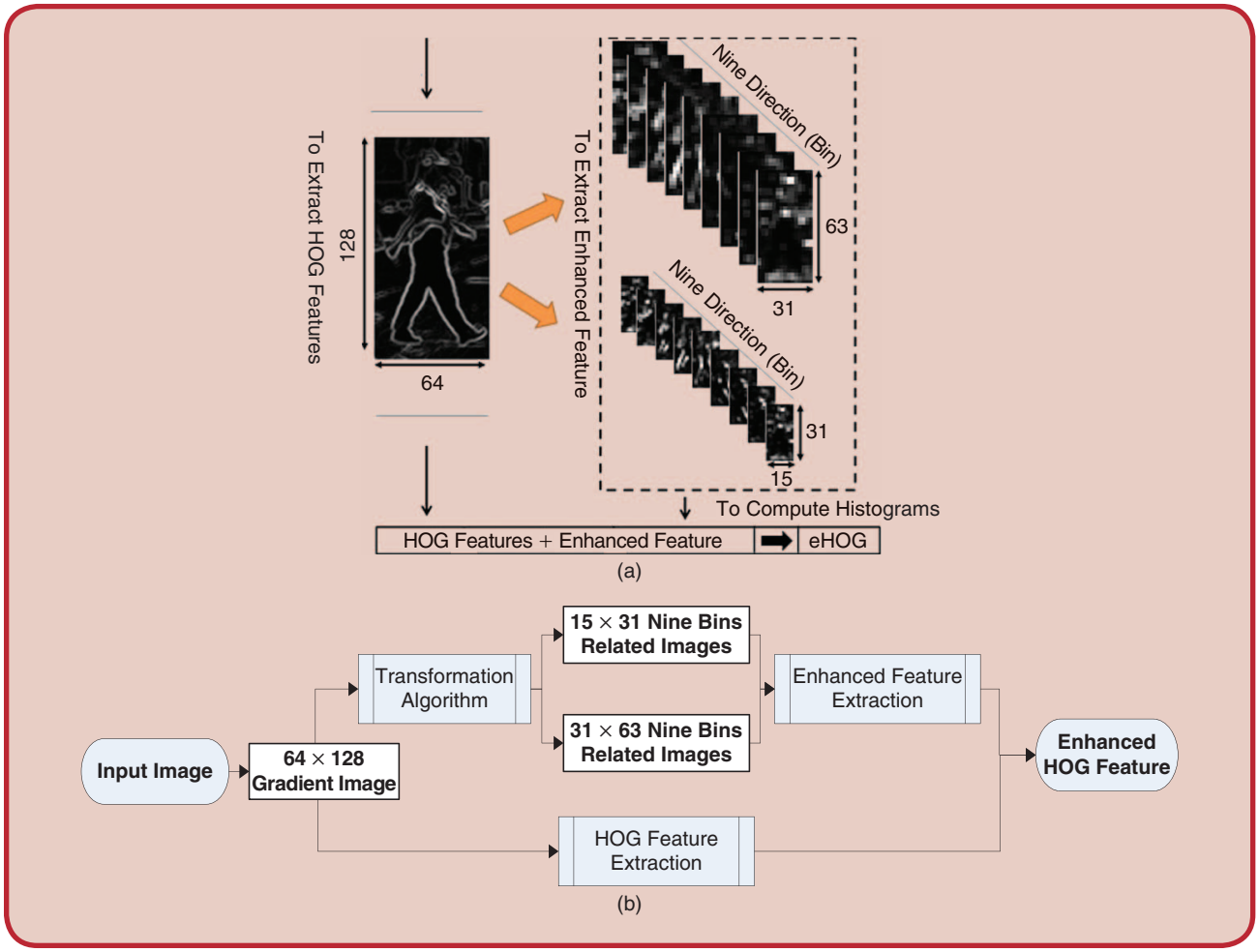


FIG 1 The process of extracting the eHOG feature: (a) The original edge image is transformed into two groups of small edge image, with each group corresponding to nine direction bins. (b) The process of extracting the eHOG feature, which is a further development of (a).

The HOG feature is based on pedestrian contour information: it lies only in the junction of the pedestrian and the background, so the pedestrian contour information is relatively small.

algorithm of gradient images (see section Fig. 1(b)), it will generate two groups of the nine bin-related images used to extract the enhanced feature (see section Fig. 1(b)); one group is 31×63 pixel and the other group is 15×31 pixel. Finally, we combine the enhanced feature with the HOG feature to form the eHOG feature. Details will be given in the following paragraphs.

A. Gradient Image Transformation Algorithm

The HOG feature is based on pedestrian contour information: it lies only in the junction of the pedestrian and the background, so the pedestrian contour information is relatively small, compared to the entire pedestrian image. To enhance the contour information, without adding redundant information, we propose the eHOG feature. In the extraction of HOG, overlapping is used between adjacent blocks to enhance the correlation. The HOG feature is based on 16×16 pixel blocks, and the sliding step is 8 pixels, which is equivalent to a three-quarters overlapping region between adjacent blocks. Overlapping makes it possible to use edge information over a larger range and enhances edge information. Similarly, we apply the overlapping method in the extraction of the eHOG feature. We use the overlapping method not only in eHOG feature extraction, but also in the gradient image transformation process. This is equivalent to using the overlapping method twice, so it further enhances the correlation.

For example, to obtain a group of 15×31 pixel uni-orientation gradient images, the gradient image transformation process is shown in Fig. 2. For a 64×128 pixel gradient image, each pixel contains gradient magnitude and angle information. The gradient image transformation process is as follows:

1) Dividing Gradient Image into Fixed-Size Blocks

Using the same overlapping method as HOG feature extraction with an 8×8 block via a four-pixel sliding step, we can obtain 15×31 blocks (see Fig. 2(a)).

2) Computing Gradient Orientation Histogram

Similar to HOG feature extraction, each block is mapped into nine bins. It will produce a $(15 \times 31) \times 9$-D gradient orientation histogram (see Fig. 2(b)).

3) Forming Nine Uni-Orientation Gradient Images

Taking one bin of each block, we can get 15×31 bins that are used to form one uni-orientation gradient image of 15×31. Similarly, taking one bin of each block in sequence, with nine bins we can obtain nine uni-orientation gradient images (see Fig. 2(b) to (c)).

Likewise, using 4×4 pixel blocks with two-pixel sliding step, we can obtain a group of 31×63 pixel uni-orientation gradient images. Every group contains nine uni-orientation gradient images.

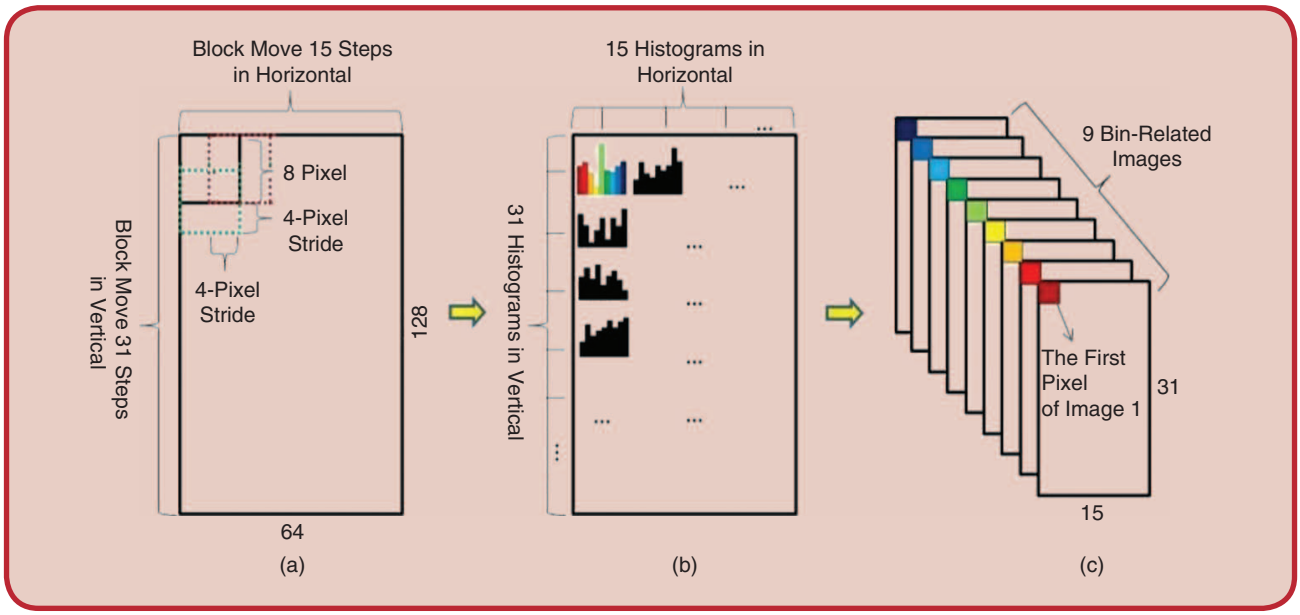


FIG 2 Gradient image transformation: (a) Gradient image. (b) The nine bins that form a uni-orientation gradient image. (c) Nine bin-related images.

B. Enhanced Hog Feature Extraction

The enhanced HOG feature is based on every group of uni-orientation gradient images. Using a method similar to HOG extraction, we computed the vector in each uni-orientation gradient image of the group, and then cascaded it into the final eHOG feature. For example, for the 15×31 pixel uni-orientation gradient image, the process of extraction is described in the following steps, also shown in Fig. 3 (a):

1) Dividing each 15×31 pixel uni-orientation gradient image of the group into $7\times3=21$ blocks; every block is 8×8 pixel and sliding step is four pixels.
2) Dividing the 8×8 block above into four 4×4 pixel cells.
3) The histogram of every 8×8 block is formed by four bins via bilinear interpolation similar to that of HOG extraction. By using the corresponding weight, the value of a pixel not only contributes to the cell to which the pixel belongs, but also to that of neighboring cells. This can be seen from Fig. 3 (b). Every pixel (i, j) in the block has a weight $w_N(i,j)$ corresponding to the bin N(N = 1, 2, 3, 4) in this block. The weights are related to the spatial position of pixels; they are calculated as follow:

$$w_1(i,j)=(1-t_i)\times(1-t_j) \tag{1}$$

$$w_2(i,j)=(1-t_i)\times t_j \tag{2}$$

$$w_3(i,j)=t_i\times(1-t_j) \tag{3}$$

$$w_4(i,j)=t_i\times t_j \tag{4}$$

$$t_i=i/7,\ t_j=j/7 \tag{5}$$

where $w_N(i,j)$ is the weight shown in Fig. 3 (b), t_i and t_j are the distance variations in the horizontal and vertical directions.

4) We block normalization, which is different from the usual normalization approach, such as L1-norm, L1-sqrt or L2-norm. Our normalization method is as follows:

$$W=\sum I(i,j)^2 \tag{6}$$

$$v_i=\frac{v_i}{\varepsilon+\sqrt{W}} \tag{7}$$

where W is the sum of the squares of the pixel values, $I(i,j)$ is the intensity of point (i,j), $\{v_i \mid i=1,2,3,4\}$ denotes four dimensional vectors of the i-th block, and ε is a small correction factor, used to avoid a zero as the denominator.

As a result, each uni-orientation gradient image of the group can be described by a $(7\times3)\times4=84$ dimensional vector, and the group of 15×31 pixel uni-orientation gradient images can be described by $84\times9=756$ dimensional vectors, which will be used as enhanced vectors together with original HOG features for pedestrian detection.

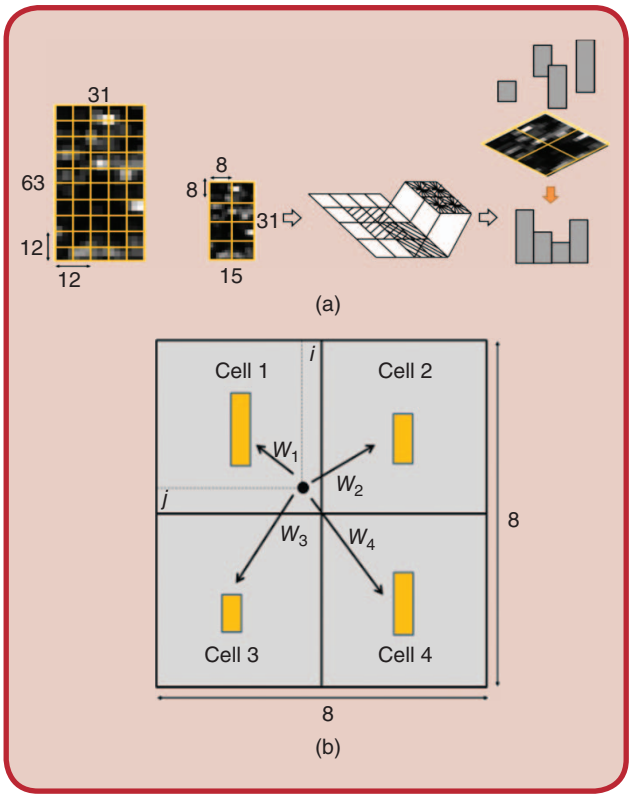


FIG 3 (a) eHOG feature computation process. (b) Bilinear interpolation process.

Similarly, for the 31×63 pixel uni-orientation gradient images, we select the block size as 12×12 and sliding step as six pixels. Thus, nine 31×63 pixel uni-orientation gradient images of the group can be described by a $9\times((4\times9)\times4)=1296$ dimensional vector.

After the computation of the enhanced features, the enhanced features are concatenated with the HOG feature to form the eHOG feature, which has a very low $3780+756+1296=5832$-dimensional vector.

IV. Optimization of eHOG Feature

In fact, the optimal parameters of eHOG are dependent on the optimal parameters of the enhanced features. The computation of eHOG consists of two parts: one is gradient image transformation, and the other is feature extraction in a group of uni-orientation gradient images, varying with different block sizes and sliding steps used. Hence, the detector performance varies according to block size and sliding step. In our work, we always set the sliding step equal to half of the block size, to reduce the number of combinations. To study the effect of various parameter combinations on detector performance, we use only linear SVM [19] for training and classification, and use the INRIA dataset [1] (available at [20]). Each evaluation result is described by a detection error tradeoff (DET) curve, which is a graphical plot of error rates for binary classification systems, plotting false reject rate vs. false accept rate. For convenience, we

use the representation of "a-b × c" to denote performance curve, where "a" denotes the sliding step, and "b × c" denotes that the block size is b × c pixels when extraction is performed using the enhanced HOG feature.

If we use an 8 × 8 pixel block with a four-pixel sliding step in the transformation process of the gradient image, a group of 15 × 31 pixel uni-orientation gradient images is the result, and we extract using the enhanced HOG feature by block size: 4 × 4, 6 × 6 or 8 × 8 pixels. As shown in Fig. 4 (a), "4-4 × 4", "4-6 × 6" and "4-8 × 8" give close results, with the detection rate at 90.7%, 91.1% and 90.5% at 10^{-4} FPPW, respectively.

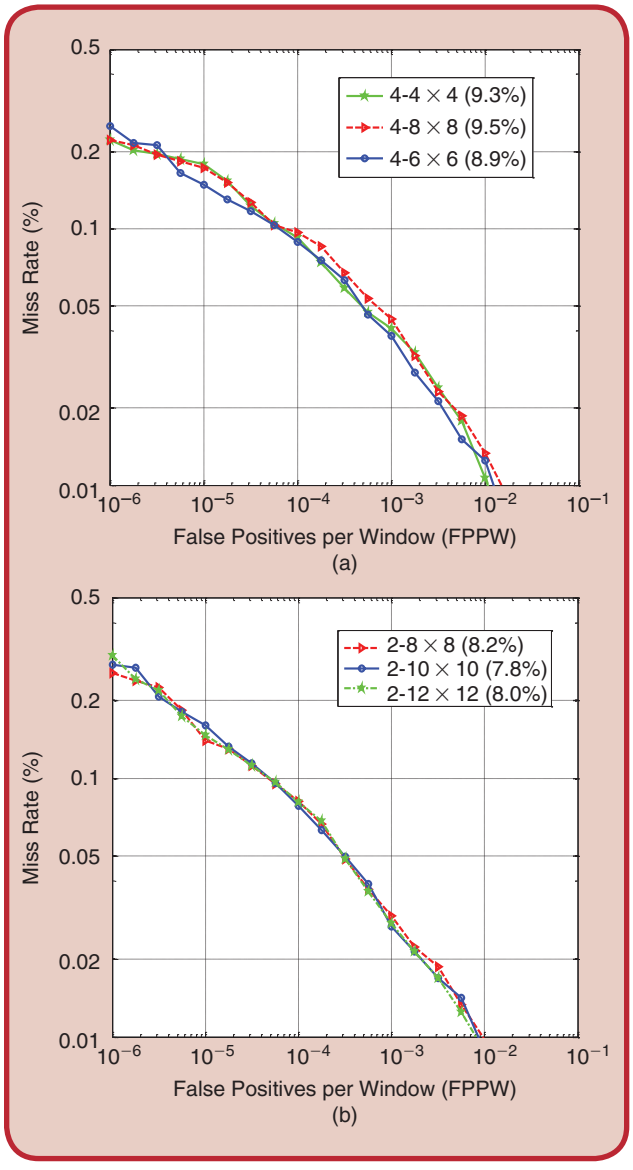


FIG 4 For convenience, we use the representation of "a-b × c" to denote performance curve, where "a" denotes the sliding step, and "b × c" denotes that the block size is b × c pixels when extraction is performed using the enhanced HOG feature: (a) The effect of block size on performance in 15 × 31 pixel gradient image; (b)The effect of block size on performance in 31 × 63 pixel gradient image.

If we use a 4 × 4 pixel block with a two-pixel sliding step in the transformation process of gradient images, we can compute the enhanced HOG feature in a group of 31 × 63 pixel uni-orientation gradient images, and then we can have 8 × 8, 10 × 10 and 12 × 12 pixel blocks. According to Fig. 4 (b), the detector performance of "2-8 × 8", "2-10 × 10" and "2-12 × 12" pixel blocks are nearly identical, with a detection rate of 91.8%, 92.2% and 92.0% at 10^{-4} FPPW, respectively.

If we use a 6 × 6 pixel block with a three-pixel sliding step in the transformation process of gradient image, we can obtain the enhanced HOG feature in a group of 20 × 41 pixels of a uni-orientation gradient image. Then, we can use blocks of various sizes. Fig. 5 describes the detector performance using 6 × 6, 8 × 8 and 10 × 10 pixel blocks, and the detection rates are 90.6%, 91.4% and 90.3% at 10^{-4} FPPW, respectively.

Table 1 shows the performance of various block sizes and sliding steps. Considering the additional dimension, information redundancy and detection rate, we add the eHOG feature at 2-12 × 12 and 4-8 × 8 to the HOG feature to form the eHOG feature. The eHOG feature describes not only the information about the 16 × 16 pixel block (HOG), but also 24 × 24 (eHOG at 2-12 × 12) and 32 × 32 (eHOG at 4-8 × 8) pixel blocks. The length of our eHOG vector is 3780 + 1296 + 756 = 5832, which achieves the best performance, as shown in section V.

V. Results and Analysis

A. Experimental Results and Analysis on FPPW

The INRIA Person dataset is widely used for evaluating pedestrian detection, and it has been used for evaluating

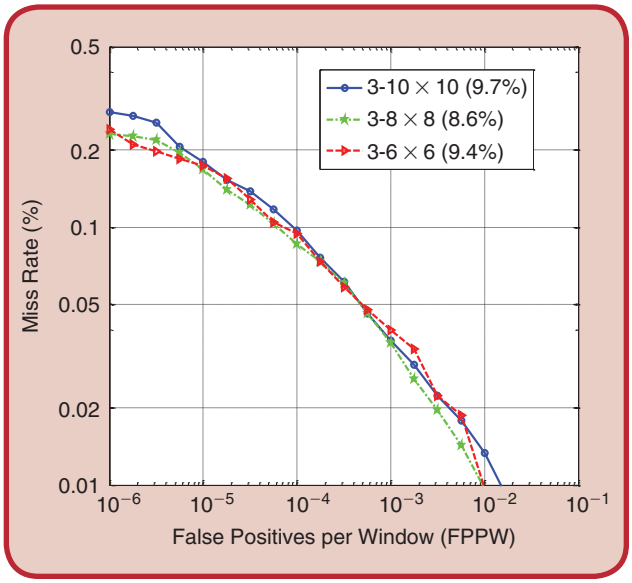


FIG 5 The effect of block size on performance in 20×41 pixel gradient image.

Table 1. The effect of various block sizes and sliding steps on the result.

Block Size When Transforming	Uni-Orientation Gradient Image Size	Block Size with Extraction Feature	Equivalent Block Size in Original Image	Length of eHOG Vector	Additional Dimensions	Detection Rate at 10^{-4} FPPW
8×8	15×31	4×4	16×16	7560	3780	90.70%
		6×6	24×24	5076	1296	91.10%
		8×8	32×32	4536	756	90.50%
6×6	20×41	6×6	18×18	6588	2808	90.60%
		8×8	24×24	5076	1296	91.40%
		10×10	30×30	4536	756	90.30%
4×4	31×63	8×8	16×16	7560	3780	91.80%
		10×10	20×20	5760	1980	92.20%
		12×12	24×24	5076	1296	92.00%

the HOG features as well. Hence, we used the INRIA Person dataset for our experiment in this paper. We use 1208 mirror images from 1208 initial positive samples in the INRIA dataset, and make them the positive samples, i.e., we have 2416 positive samples. Then, we cut each negative sample into 10 images of 64×128 pixels for the 1218 negative samples; hence, we get 12180 initial negative images for the first training. Again, for the positive samples, we get image pairs of 64×128 pixels from the centre zone of the initial positive samples with 96×160 pixels. In this paper, using FPPW-MissRate, we compare the performance of eHOG and HOG with both linear-SVM and Histogram Intersection Kernel SVM (HIKSVM) [21], respectively. We also compare these two features combined with the color self-similarity feature (termed CSS) [6]. The CSS feature is based on self-similarity of lowlevel features, which captures pairwise statistics of spatially localized color distributions, in particular color histograms from different sub-regions within the detector window.

1) Comparison for Single Feature of eHOG and HOG

As shown in Fig. 6 (a), we will evaluate the performance of HOG, eHOG and CoHOG using linear SVM or HIKSVM, where "-Linear" and "-HIK" denote that the linear SVM and HIKSVM is used. As can be seen, compared with the method proposed by Dalal, the performance of the eHOG feature with linear SVM classifiers has been greatly improved. The experiment using the INRIA database shows that eHOG improves the detection rate by approximately 4.5% at 10^{-4} FPPW, in comparison with Dalal's results. In addition, increasing the detection rate without significantly increase the number of dimensions only adds 2053 dimensions. Furthermore, we compare our detector to CoHOG detectors whose experimental results involved the use of an SVM classifier. Although CoHOG-Linear is about 2.8% higher than eHOG-Linear and about 1.8% higher than eHOG-HIK, the dimension of the CoHOG feature is 138,816 on the INRIA dataset (see Table 2), which is not conducive to fast pedestrian detection.

Through the optimization of the evaluation of these features, we see that HIKSVM and linear SVM require the same time in which to judge the classification. For the original HOG feature, we compare using HIKSVM with using linear SVM, and gain an approximate 2% increase in the detection rate. It shows that, compared to using linear SVM, HIKSVM can effectively improve the detection rate without increasing the computation time. Using HIKSVM as a classifier, and by combining eHOG features, detection efficiency has been greatly improved. Experiment results show that the detection rate based on HIKSVM and eHOG (about 94.2% at 10^{-4} FPPW) increases about 6% to initial HOG, and about 1.5% for linear-SVM combined with eHOG. So, HIKSVM is more efficient than linear SVM. Since eHOG is also a histogram feature, it is more suitable to use HIKSVM as the final classifier.

The results above show that eHOG is a much more efficient alternative to HOG and improves the detection rate considerably. eHOG improves the detection rate by approximately 4.5% over HOG for the same classifier.

In order to compare the detection speed of HOG and eHOG, we selected 500 images from the INRIA database. On an AMD Phenom TM II X4 965 Processor, 3.4-GHz computer with 16-GB RAM, the average processing time of the eHOG feature was about 0.53s while that of HOG was 0.28s. The eHOG feature can achieve high performance without requiring too much processing time.

2) Comparison for the Combined Features of eHOG and HOG

We have compared HOG and eHOG as a single feature above, and now we will evaluate them when combined with

other features; for example, HOG+CSS and eHOG+CSS, as shown in Fig. 6 (b).

We mainly consider the following four cases: i.e., HOG+CSS with HIKSVM, HOG+CSS with linear SVM, eHOG+CSS with HIKSVM and eHOG+CSS with linear SVM. Using HIKSVM enhances performance after fusion compared to using linear SVM for detection rate (improved about 1.5% at 10^{-4} FPPW), which is similar to the former comparison. When linear SVM is used, the performance for using eHOG+CSS improves by about 2.6% to HOG+CSS, and about 3.8% by using HIKSVM. Experimental results show that, in comparison to the original HOG, eHOG can express the gradient information more effectively and hence increase the detection rate. Therefore, eHOG is an efficient complement to the original HOG, as the latter is more sensitive to the difference in scales. Table 2 shows the comparison results for FPPW. It also illustrates that the detection rate is improved constantly with eHOG after combination with the CSS feature at 10^{-4} FPPW.

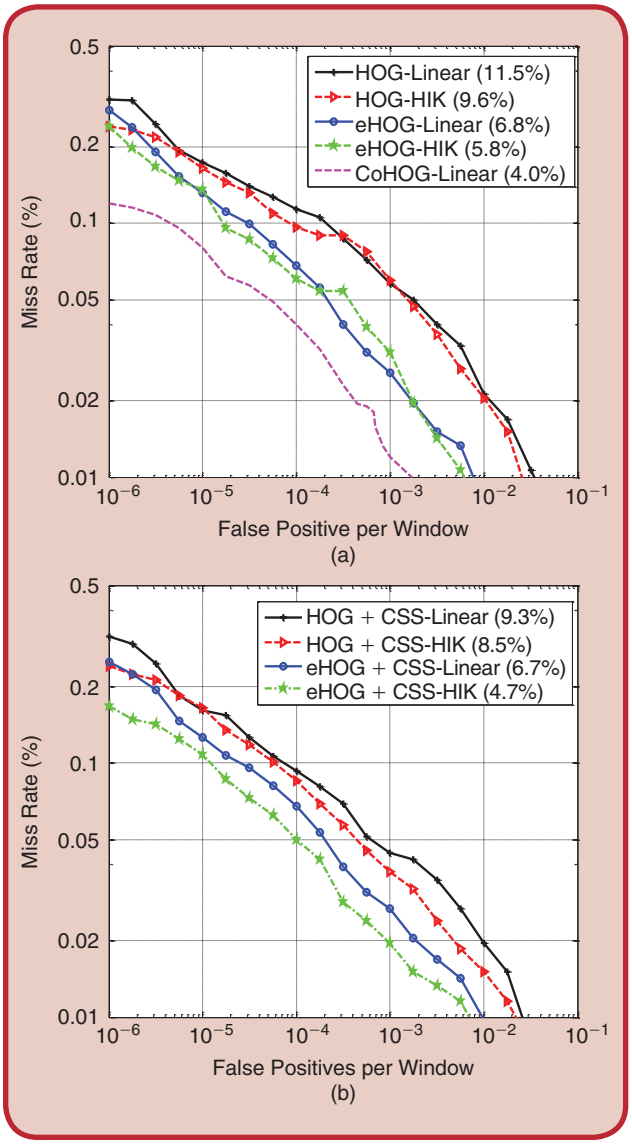


FIG 6 (a) Performance Comparison for eHOG and HOG on FPPW. (b) Performance Comparison for eHOG+CSS and HOG+CSS on FPPW.

B. Experimental Results and Analysis on FPPI

The FPPI-MissRate evaluation method, which is widely used in object detection, evaluates performance based on the detected images. Compared with FPPW-MissRate, this approach offers a more appropriate evaluation method for error detection.

1) Comparison for Single Feature of eHOG and HOG

We used linear SVM and HIKSVM for eHOG and HOG and conducted the experiments, respectively, with FPPW. When we used linear SVM, as we see from Fig. 7 (a), the detection rate improves by about 6% (at 10^{-1} FPPI) for eHOG to HOG and about 8% by using HIKSVM. These experimental results fully demonstrate that eHOG performance is far superior to HOG in the two evaluation methods, and eHOG can effectively improve the detection rate.

2) Comparison for the Combined Features of eHOG and HOG

We also evaluate the performance with FPPI for eHOG+CSS and HOG+CSS. We compare the detection performance of eHOG + CSS and HOG + CSS with HIKSVM and linear SVM classifier at 10^{-1} FPPI, as shown in Fig. 7 (b). For HOG + CSS, using HIKSVM improves by about 6% over the use of linear SVM, and about 7% for eHOG + CSS. By using linear SVM, eHOG + CSS increases the detection rate by about 2% over HOG+CSS, and about 3% by using HIKSVM. Hence, after fusion with other features, eHOG can still consistently improve the detection rate. Table 3 shows the comparison results of FPPI, while it illustrates that the detection rate is greatly improved by combining the eHOG feature at 10^{-1} FPPI.

C. Some Result Images

We evaluate our method for pedestrian detection using the INRIA Person dataset. The liblinear and libHIK packages are used for SVM training. We use the same initial positive and negative samples that Dalal and Triggs used in their training. During training, we apply several rounds of

Table 2. Experiment results based on FPPW.

Feature	Single Feature			Integrate CSS	
	Linear SVM	HIKSVM	Dimension	Linear SVM	HIKSVM
HOG	88.5%	90.4%	3,780	90.7%	91.5%
CoHOG	96%	/	138,816	/	/
eHOG	93.2%	94.2%	5,832	93.3%	95.3%

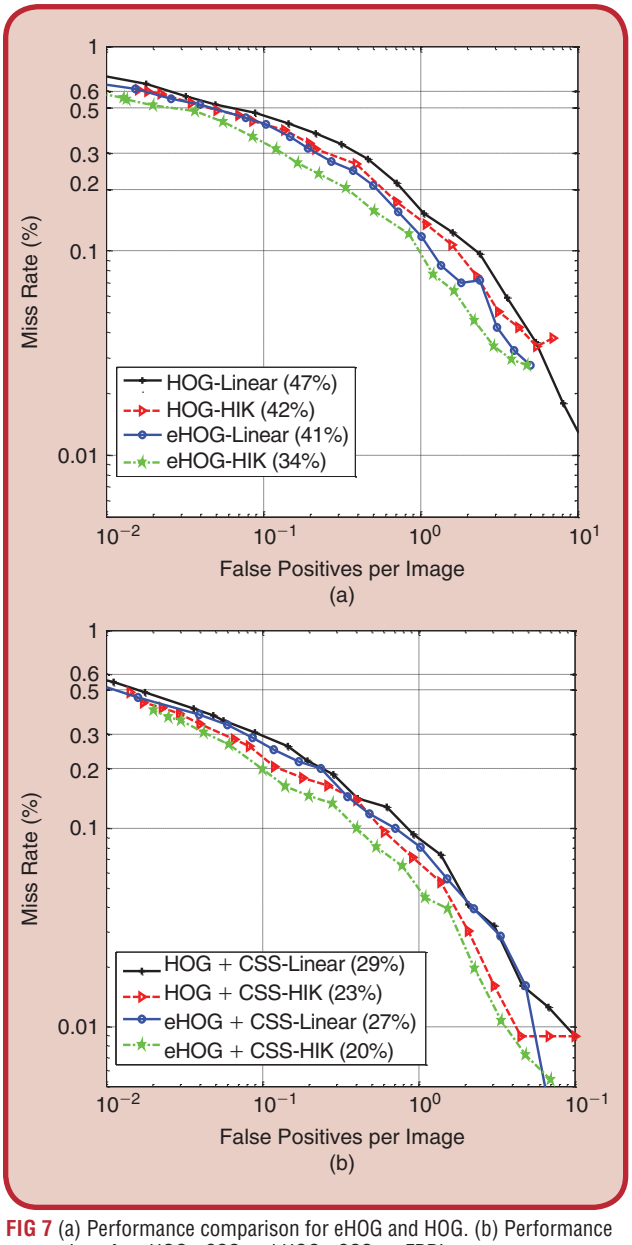


FIG 7 (a) Performance comparison for eHOG and HOG. (b) Performance comparison for eHOG+CSS and HOG+CSS on FPPI.

bootstrapping to collect false positive examples. In every bootstrapping exercise, we use meanshift [22] to reduce the false positive windows in the neighborhood. Meanshift could also be used in detection.

Fig. 8 shows some result images demonstrating the performance of our detector by using meanshift to merge windows, in addition to the sample detections of the detector (eHOG) trained on HIKSVM on images from Test Set 2 of the INRIA Person dataset. Set 2 contains challenging images taken of people in different poses, and most images contained people standing or walking. Moreover, some images show people running, going downhill, bicycling or playing.

Table 3. Experiment results based on FPPI.

Feature	Single Feature		Integrate CSS	
	Linear SVM	HIKSVM	Linear SVM	HIKSVM
HOG	53%	58%	71%	77%
eHOG	59%	66%	73%	80%
Improved rate	6%	8%	2%	3%

VI. Conclusion

The HOG feature is a well-known feature for pedestrian detection. However, the HOG feature is not sophisticated enough to handle a scale variation of pedestrians. Besides, simply downsampling an image into a different scale, or decomposing via wavelet into multi-resolution subimages, and by calculating their HOG feature and combining them cannot improve performance. Therefore, in this paper, based on the idea of multi-resolution feature descriptors, we propose a new and robust edge feature referred to as Enhanced HOG (eHOG) for pedestrian detection. The extraction process of the eHOG descriptor is similar to the process of extracting edge information of a downscaling image, while it retains most of the edge gradient information. To evaluate performance, the eHOG feature and HOG feature are used, respectively, for experiments with the Linear SVM classifier or HIKSVM classifier. In addition, we also compare the performance between HOG and eHOG features with the coalesced CSS feature, respectively. The experimental results, according to two evaluation methods - FPPW-Miss-Rate and FPPI-MissRate, show that the performance of the eHOG feature is significantly higher than that of the current HOG feature, whether it is used with a single feature or coalesced features and regardless of whether it used with the Linear SVM classifier or HIKSVM classifier. The detection rate of eHOG outperforms HOG consistently.

FIG 8 Some result images on the INRIA dataset.

Acknowledgment

We would like to thank our colleagues for helping to this paper and the anonymous reviewers and the associate editor for their valuable suggestions.

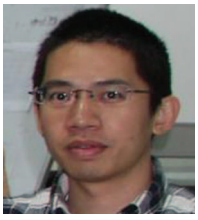
About the Authors

Yong Zhao received the M.S. degree in Electrical Engineering from Northwestern Polytechnic University and the Ph.D. degree from Southeast University in 1989 and 1991, respectively. He has worked for Honeywell in Canada from 2000 to 2004. He joined the Department of Electronics Engineering of Shenzhen Graduate School of Peking University in 2004, where he is currently an associate professor, and also an external professor employed by Guizhou University. His research interests mainly focus on the embedded application of intelligence image algorithms, such as scene target detection, extraction, tracking, recognition, and behavior analysis.

Yongjun Zhang received the M.S. degree in college of computer science and information in 2010 from Guizhou University, Guiyang, China. Now he is an associate professor in Guizhou University. He is currently a joint training doctoral student of Peking University and Guizhou University, and studying in the key laboratory of Integrated Microsystems of Shenzhen Graduate School of Peking University. His main research interest is computer vision.

Ruzhong Cheng received the M.S. degree in Aerospace Engineering and Mechanics in 2003 from Harbin Institute of Technology (HIT), and received Ph.D. degree in Electrical Engineering from Peking University in 2013. He had worked in the MEMS Center of HIT from 2004 to 2005. He is currently working in the Department of Electronic Engineering and Computer Science in Peking University Shenzhen Campus as a postdoctor. His main research interest is computer vision.

Daimeng Wei received his Bachelor of Engineering degree from Chongqing University in 2009, China. He has studied as a graduate student in the key laboratory of Integrated Microsystems of Shenzhen Graduate School of Peking University from 2009 to 2013. His main research interest is computer vision.



Guoliang Li received his Bachelor of Engineering degree from Wuhan University in 2009, China. He has studied as a graduate student in the key laboratory of Integrated Microsystems of Shenzhen Graduate School of Peking University from 2009 to 2013. His main research interest is computer vision.

References

[1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. Conf. Computer Vision Pattern Recognition*, San Diego, CA, 2005, vol. 1, pp. 886–893.

[2] C. Wojek and B. Schiele, "A performance evaluation of single and multi-feature people detection," in *Pattern Recognition*, vol. 5096. Berlin, Heidelberg, Germany: Springer, 2008, pp. 82–91.

[3] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," *Int. J. Comput. Vis.*, vol. 63, no. 2, pp. 153–161, July 2005.

[4] P. Sabzmeydani and G. Mori, "Detecting pedestrians by learning shapelet features," in *Proc. Conf. Computer Vision Pattern Recognition*, Minneapolis, MN, 2007, pp. 1–8.

[5] S. Belongie, J. Malik, and J. Puzicha, "Matching shapes," in *Proc. 18th Int. Conf. Computer Vision*, Vancouver, CA, 2001, vol. 1, pp. 454–461.

[6] S. Walk, N. Majer, K. Schindler, and B. Schiele, "New features and insights for pedestrian detection," in *Proc. Conf. Computer Vision Pattern Recognition*, San Francisco, CA, 2010, pp. 1030–1037.

[7] B. Wu and R. Nevatia, "Optimizing discrimination-efficiency tradeoff in integrating heterogeneous local features for object detection," in *Proc. Conf. Computer Vision Pattern Recognition*, Anchorage, AK, 2008, pp. 1–8.

[8] T. Watanabe, S. Ito, and K. Yokoi, "Co-occurrence Histograms of Oriented Gradients for pedestrian detection," *Adv. Image Video Technol.*, vol. 5414, pp. 37–47, 2009.

[9] X. Y. Wang, T. X. Han, and S. C. Yan, "An HOG-LBP human detector with partial occlusion handling," in *Proc. 12th Int. Conf. Computer Vision*, Kyoto, Japan, 2009, pp. 32–39.

[10] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, July 2002.

[11] S.-U. Hussain and W. Triggs, "Feature sets and dimensionality reduction for visual object detection," in *Proc. British Machine Vision Conf.*, Wales, U.K., 2010.

[12] P. Ott and M. Everingham, "Implicit color segmentation features for pedestrian and object detection," in *Proc. Conf. Computer Vision*, Kyoto, Japan, 2009, pp. 723–730.

[13] P. Dollár, Z. W. Tu, P. Perona, and S. Belongie, "Integral channel features," in *Proc. British Machine Vision Conf.*, London, U.K., 2009.

[14] R. M. Anwer, D. Vázquez, and A. M. López, "Opponent colors for human detection," in *Pattern Recognition and Image Analysis*, vol. 6669. Berlin Heidelberg, Germany: Springer, 2011, pp. 363–370.

[15] M. A. Rao, D. Vázquez, and A. M. López, "Color contribution to part-based person detection in different types of scenarios," in *Computer Analysis of Images and Patterns*, vol. 6855. Berlin Heidelberg, Germany: Springer, 2011, pp. 463–470.

[16] Y. Socarrás, D. Vázquez, A. M. López, D. Gerónimo, T. Gevers, "Improving HOG with image segmentation: Application to human detection," in *Proc. Conf. Advanced Concepts Intelligent Vision Systems*, 2012, vol. 7517, pp. 178–189.

[17] J. Marin, D. Vázquez, A. M. López, J. Amores, and B. Leibe, "Random forests of local experts for pedestrian detection," in *Proc. Conf. Computer Vision*, Sydney, NSW, 2013, pp. 2592–2599.

[18] P. Dollár, S. Belongie, and P. Perona, "The fastest pedestrian detector in the west," in *Proc. British Machine Vision Conf.*, Wales, U.K., 2010.

[19] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Li, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, 2008.

[20] [Online]. Available: http://pascal.inrialpes.fr/data/human/

[21] S. Maji, A. C. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in *Proc. Conf. Computer Vision Pattern Recognition*, Anchorage, AK, 2008, pp. 1–8.

[22] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Trans. Inform. Theory*, vol. 21, no. 1, pp. 32–40, 1975.

ITS