International Journal of Innovative Computing, Information and Control
Volume **11**, Number **2**, April **2015**

ICIC International ©2015 ISSN 1349-4198
pp. 599–614

EFFICIENT AND REAL-TIME PEDESTRIAN DETECTION AT NIGHT-TIME ENVIRONMENTS

Yongjun Zhang[1], Yong Zhao[2,*], Guoliang Li[2], Daimeng Wei[2] and Ruzhong Cheng[3]

[1]College of Computer Science and Information
Guizhou University
Huaxi Dist., Guiyang 550025, P. R. China
zyj6667@126.com

[2]The Key Lab of Integrated Microsystems
Shenzhen Graduate School of Peking University
Nanshan Dist., Shenzhen 518055, P. R. China
*Corresponding author: yongzhao@pkusz.edu.cn

[3]PKU-HKUST ShenZhen-HongKong Institution
Nanshan Dist., Shenzhen 518055, P. R. China
chengruzhong@gmail.com

Received March 2014; revised August 2014

Abstract. *This paper presents a method for night-time pedestrian detection using a single near-infrared (NIR) camera; a support vector machine (SVM) is employed to select the most class-relevant infrared pedestrian features. First, for effective sub-windows scanning, a region of interest (ROI) selection strategy based on the dual-threshold segmentation method is performed to obtain candidate windows. Second, to maximize the capacity for pedestrian information, a novel feature extraction method is proposed to form a new feature called Modified HOG (MHOG), which contains a larger range of information than the Histograms of Oriented Gradients (HOG) feature used widely for pedestrian detection. The feature extraction method is evaluated on night-time/day-time pedestrian datasets. At 10^{-4} FPPW, the experimental result shows that the detection rate of MHOG can increase by approximately 5.35%, compared with HOG on our night-time pedestrian dataset (NTPD dataset), and by about 2.22% compared with HOG on the day-time pedestrian dataset (INRIA dataset). Finally, the proposed approach is tested on both urban and suburban scenes. Results indicate that the algorithm can produce an average recall rate of 90% and requires an average processing time of approximately 26ms per frame. Our findings indicate that the proposed system is robust and efficient, can run in real-time, and will be useful for practical applications.*

Keywords: Night-time pedestrian detection, ROI, Feature extraction, HOG, MHOG

1. **Introduction.** Pedestrians are the most vulnerable traffic participants; in traffic accidents, they are much more likely to be injured than individuals in motor vehicles. Accurate and reliable pedestrian detection could help drivers to prevent such accidents. Intelligent vehicles are the future trend in the automobile industry, due to their human-like user interfaces and high safety levels while driving. Pedestrian detection is one of the most critical problems for active safety systems in future intelligent vehicles [1].

Most pedestrian detection research has focused on the day-time environment, and the key problem of pedestrian detection is finding robust features characteristic of the pedestrian. Many methods have been proposed for pedestrian detection in day-time environments [2,3]. Dalal and Triggs proposed an excellent feature, called Histograms of Oriented Gradients (HOG) [4], which has demonstrated good performance in terms of accuracy;

however, its efficiency has not been sufficient for real-time applications. On the basis of HOG, Zhu et al. used AdaBoost with the HOG features to speed up the computation, while maintaining similar accuracy to HOG [5]. Maji et al. improved the detectors by using multilevel edge energy features and applying an intersection kernel SVM as a learning method [6]. Ouyang and Wang proposed a probabilistic pedestrian detection framework to handle the shortcomings of part detectors [7].

Present night-time pedestrian detection typically uses Near-Infrared (NIR) cameras or Far-Infrared (FIR) cameras [8,9], while some research has utilized cameras based on thermal images [10]. Pedestrians captured by IR cameras at night often appear as a vague figure or outline. A pedestrian detection system for night-time environments typically requires this system to overcome existing problems of low contrast level, image blur, and image noise. The popular HOG feature is good in terms of accuracy, but the performance of HOG is shown to degrade when the condition of the imaging data is poor. Nanda and Davis introduced a simple yet effective method in Infrared Videos that uses probabilistic templates for pedestrian detection [11]. Cao et al. proposed a modified Local Binary Pattern (LBP) feature extraction method for pedestrian detection in night/dark environments [12]. Lin et al. combined HOG and contour features to build a reliable classification system [13]. The large number of dimensions of the feature vectors used in these methods compromise their efficiency for practical applications, and the performance of some methods highly dependents on costly image acquisition devices. Therefore, to capture more information of the detected image and to increase the accuracy of pedestrian detection in similar cases, we propose a Modified Histogram of Oriented Gradients (MHOG), which can express shapes in more details than HOG.

As in the case of building night-time pedestrian detection systems: Tian et al. present an approach for pedestrian detection in night-time with a normal camera using an SVM classifier. Objects in the video are extracted with an adaptive threshold segmentation method [14]. Sun et al. describe a night vision pedestrian detection system for autonomous vehicles using an onboard forward-looking infrared (FLIR) camera, and Haar-like features are used to discriminate infrared pedestrians [15]. To address the non-rigid nature of human appearance on the road, Xu et al. proposed a two-step detection/tracking method for pedestrian detection and tracking using a single night-vision video camera installed on the vehicle [16]. Bertozzi et al. present a stereo system for the detection of pedestrians using far-infrared cameras, which exploits three different detection approaches: warm area detection, edge-based detection, and disparity computation [17]. Ge et al. proposed a monocular vision system for real-time pedestrian detection and tracking during night-time driving with a near-infrared (NIR) camera [18].

In this paper, we use a monocular Near-Infrared (NIR) camera to capture the images of pedestrians, and we aim to develop a robust real-time pedestrian detection system in night-time scenes. This paper makes two main contributions to the field. The first is to introduce an adaptive dual-threshold segmentation method for ROI generation, which can be regarded as a rough classifier for candidate generation. The second is to propose a novel feature extraction method. It applies the 8×8 block with 4 pixels stride to compute the orientation histograms in a 64×128 detection window, and the corresponding bins of these orientation histograms are extracted to form nine bin-related images. Sequentially, building the circular HOG (C-HOG) feature of these bin-related images will produce a new feature called the enhanced feature. Finally, we concatenate the original HOG feature and the enhanced feature as a modified HOG (MHOG) feature, which contains a larger range of information than HOG.

The rest of the paper is organized as follows. Section 2 describes the architecture of the system structure and the details of system implementation, which includes the

ROI selection, features extraction and candidate verification. Section 3 introduces the pedestrian datasets applied in this study. Section 4 discusses further details about the experiments; and the final section presents the conclusion and future work.

2. **System Implementation.** We adopt the procedures of the night-time pedestrian detection as shown in Figure 1. These steps include image preprocessing, Region of Interest (ROI) selection, feature extraction, and candidate verification. In order to achieve a more-defined image edge, in the preprocessing step, we apply a Gaussian smoothing filter to reduce noise. Region of Interest (ROI) selection is applied for effective sub-windows scanning. Feature extraction is the key part of the procedure, for which we propose a novel feature extraction method, which can express shapes in more details than HOG. During the training procedure and the candidate verification, we select key features and object classification by using the support vector machine (SVM) algorithm.

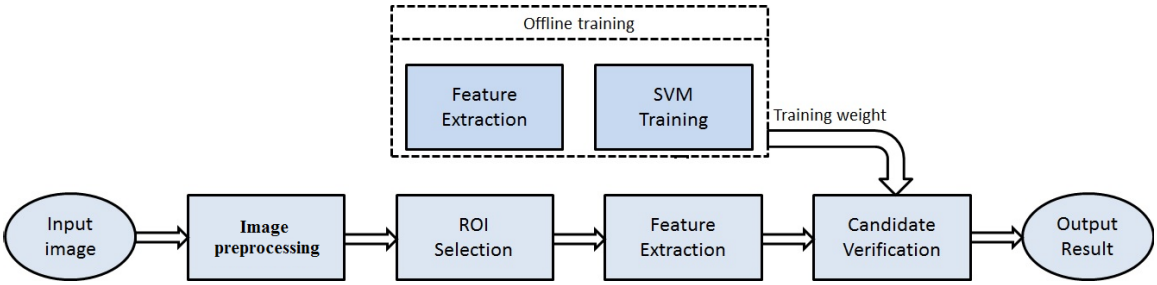


FIGURE 1. An overview of the procedures of the night-time pedestrian detection

2.1. **ROI selection.** Selecting the region of interest (ROI) can reduce the scanning area for the next process. There are many ways to select ROI. Edge-based and greyscale-based threshold segmentation methods are two popular types. A fast and robust edge-based segmentation for pedestrian detection was proposed by Kancharla et al. [19], and Ge et al. proposed efficient adaptive dual-threshold segmentation for candidate generation [18,20,21]. In the work of Ge et al., the dual-threshold segmentation is simplified and applied to ROI selection. The high threshold value, the low threshold value, and the segmentation result are computed as follows:

$$T_L(i,j) = \frac{\sum_{x=i-\omega}^{x=i+\omega} I(x,j)}{2 \times \omega + 1} + \alpha \tag{1}$$

$$T_H(i,j) = T_L(i,j) + \beta \tag{2}$$

$$S(i,j) = \begin{cases} 1, \text{ If } I(i,j) > T_H(i,j) \\ 0, \text{ If } I(i,j) < T_L(i,j) \\ 1, \ I(i,j) \in \text{ Others } \& \ I(i,j) = 1 \\ 0, \ I(i,j) \in \text{ Others } \& \ I(i,j) = 0 \end{cases} \tag{3}$$

where $I(x,j)$ is the intensity of the pixel (i,j), according to the width distribution of pedestrians (α and β), and ω can be adjusted and varied by experiments and by the different sample datasets. $T_L(i,j)$ is low threshold and $T_H(i,j)$ is high threshold. For a given pixel $I(x,j)$ in the image, $S(i,j)$ is the corresponding segmentation result. Erosion and dilatation operations are also applied to the binary image after the dual-threshold process. Figure 2 shows the segmentation results. The regions (yellow rectangle) contain a certain number of white pixels, which are selected as the ROI.

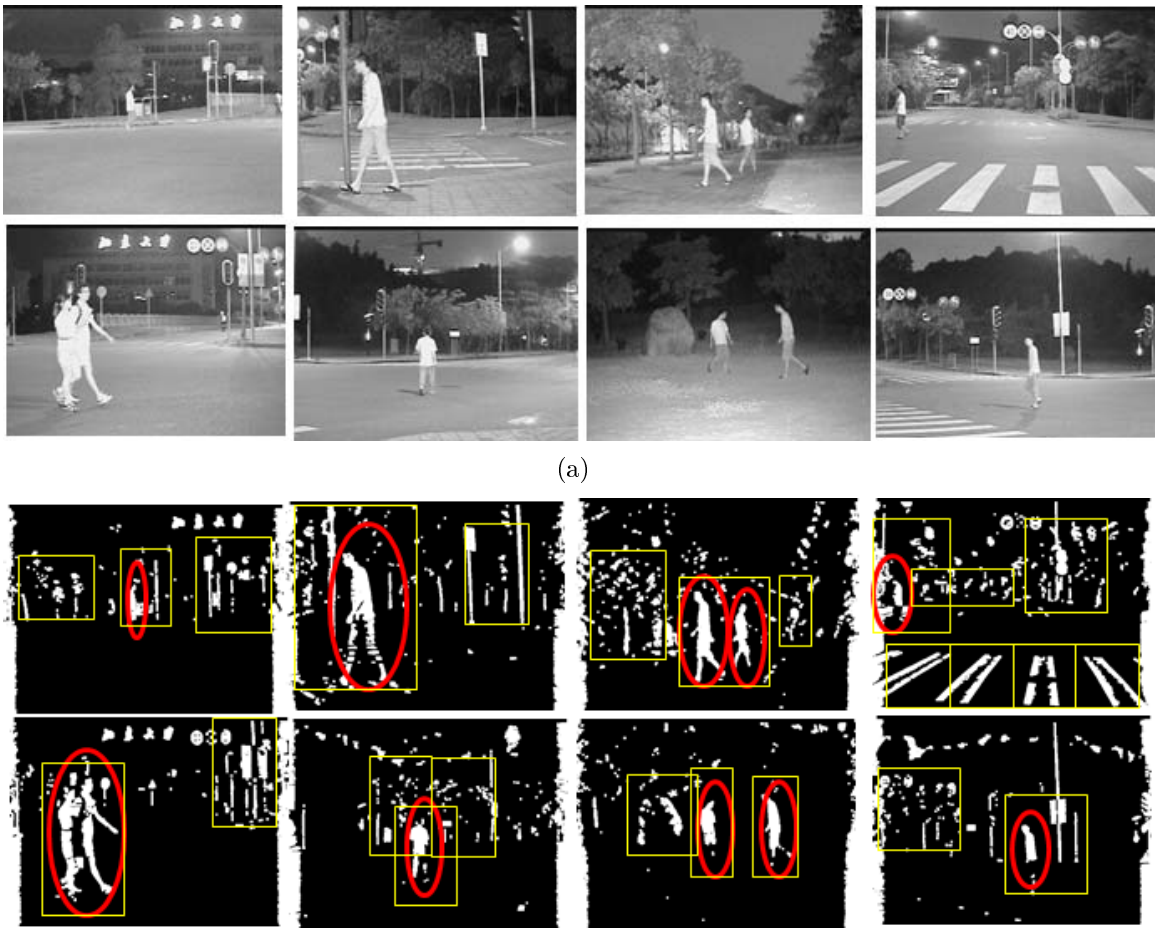


FIGURE 2. Results of segmentation: (a) input images; (b) the segmentation results after dual-threshold method

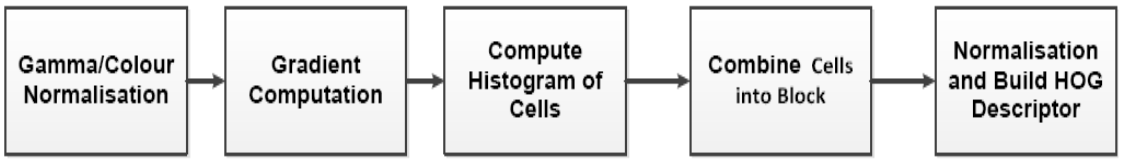


FIGURE 3. The process of calculating the HOG feature

2.2. **Feature extraction.** The HOG feature [4] is one of the most popular descriptors in pedestrian detection; it is used to describe the image gradient information. HOG originated from the Scale-Invariant Feature Transform (SIFT) descriptor [22], and it has many similarities with Edge Orientation Histograms (EOH) [23] and the Shape Context Descriptor [24]. The distinguishing characteristic of the HOG feature is that it is calculated in a uniform cell size, with dense grids. Because it is less affected by light, it is a good descriptor to detect people having a diversity of clothing and poses. Figure 3 shows the process of calculating the HOG feature.

The difficulty of night-time vision has been that the appearance of pedestrians is not distinct from their background environment, compared with that of day-time images. In order to capture more information, we add an enhanced feature to HOG to generate the MHOG feature [25]. Figure 4 shows an overview of the feature extraction process for

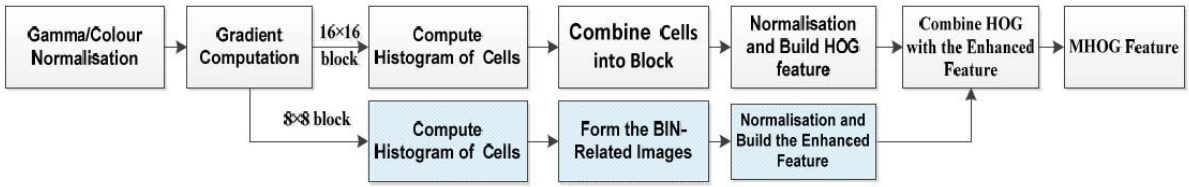


FIGURE 4. The feature extraction process for MHOG

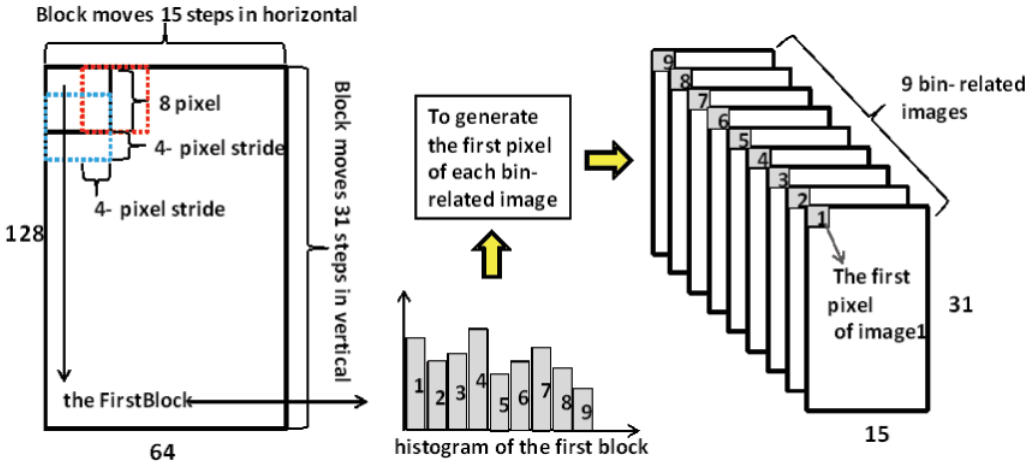


FIGURE 5. Procedure of extracting bin-related images

MHOG. The feature extraction process of the enhanced feature is shown in light blue rectangles.

The following steps describe the feature extraction process of the enhanced feature.

- The first stage shares the gradient image with HOG.
- The second stage computes the histograms of the gradient image that is similar to that of HOG. The difference is that we use the 8×8 block with 4 pixels stride instead of the 16×16 block with 8 pixels stride to compute the orientation histogram. The 8×8 block is not partitioned, and the whole block is a cell.
- The third stage aims to form nine bin-related images. The corresponding bins of these histograms are extracted to form nine images called bin-related images. Because the histogram of each cell has nine bins, the corresponding bins of all histograms are collected to generate nine bin-related images. The size of the bin-related image is 15×31 (block moves 15 and 31 steps in the horizontal and vertical directions, respectively, so that a total of 15×31 blocks will be encountered). Figure 5 describes the forming process of the first pixel of nine bin-related images.

  From Figure 5, it can be seen that the histogram of the first block will generate the first pixel of the nine bin-related images. Similarly, the second pixel of each bin-related image is obtained from the histogram of the second block; circulating this process in the horizontal and vertical directions will produce nine bin-related images. The procedure from the origin image to nine bin-related images is shown in Figure 6, and panel (c) of this figure shows the nine bin-related images with 15×31 pixels.
- The fourth stage extracts the enhanced feature using the circular HOG (C-HOG) descriptor [4]. The C-HOG layout has four spatial parameters: the numbers of

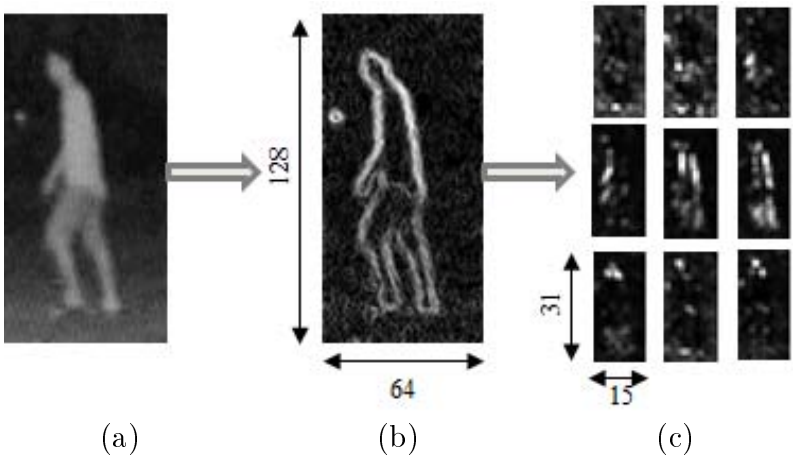


FIGURE 6. Processing the input image to yield nine bin-related images: (a) origin image; (b) gradient image; (c) nine bin-related images

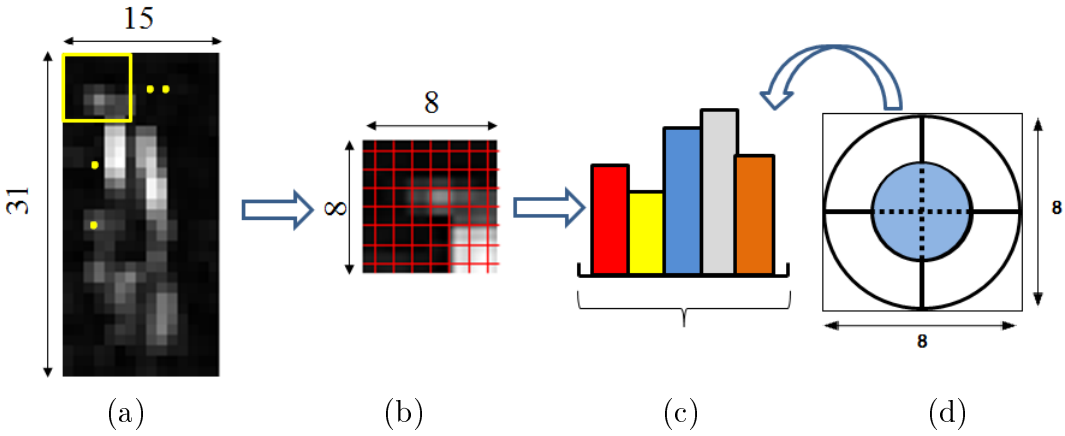


FIGURE 7. Procedure for extracting features of the bin-related image: (a) one bin-related image; (b) one 8×8 block; (c) 5-bin histogram; (d) circle block in an 8×8 block

angular and radial bins; the radius of the central bin in pixels; and the expansion factor for subsequent radii. The C-HOG descriptor has two radial bins (a centre and a surround) and four angular bins (quartering); the C-HOG descriptor with a single central cell, no central cell, and the central cell divided into four angular sectors can correspondingly generate a 5-, 4-, or 8-bin histogram in a block. As shown in Figure 7, each circle block is in an 8×8 block (Figure 7(d)), and two pixels are the best radius of the central bin. In this paper, we use a 5-bin histogram to represent one circle block, for the reason that the 5-bin histogram achieves the best detection rate (see Section 4). Each bin is created by accumulating all gradients within a circle block. Given an $n \times n$ block in a bin-related image I, every pixel $I(i,j)$ in the block has a weight $w_N(i,j)$ to the bin N $(N = 1,2,3,4,5)$ in this block. The weights are related to the spatial position of pixels; they are calculated as follows:

$$w_1(i,j) = w_h \times w_v \times w_r, \tag{4}$$

$$w_2(i,j) = w_h \times (1 - w_v) \times w_r, \tag{5}$$

$$w_3(i,j) = (1 - w_h) \times w_v \times w_r, \tag{6}$$

$$w_4(i,j) = (1 - w_h) \times (1 - w_v) \times w_r, \tag{7}$$

$$w_5(i,j) = 1 - w_r, \tag{8}$$

where

$$w_h(i,j) = i/(n-1), \tag{9}$$

$$w_v(i,j) = j/(n-1), \tag{10}$$

if $\sqrt{((i-(n-1)/2)^2 + (j-(n-1)/2)^2)} \geq n/2$

$$w_r(i,j) = 1 \tag{11}$$

else

$$w_r(i,j) = \sqrt{((i-(n-1)/2)^2 + (j-(n-1)/2)^2)}/(n/2) \tag{12}$$

By Equations (4)-(12), the feature extraction process for C-HOG does not significantly increase the computation costs, because all weights can be precomputed. By using the circle block with a 4-pixel step to scan one 15×31 bin-related image, it will produce a total of 3×7 blocks. One 5-bin histogram reflects one block; there is a $5\times(3\times7)\times9 = 945$-dimensional vector. Finally, we apply L2-norm to the 945-dimensional vector. (We experimented with a number of well-known functions including the L1-norm, L2-norm, and L1-sqrt to normalize vectors; L2-norm was selected as the best-performing option.)

- The final step concatenates the HOG feature and the enhanced feature to form the MHOG feature. The MHOG feature has a very small, $3780+945 = 4725$-dimensional vector.

From the above process, we have determined that the enhanced feature has a desirably small 945-dimensional feature vector. Meanwhile, the gradient images have been computed in the process of calculating the HOG feature (see the first stage) and the extraction process of the C-HOG feature is very fast (see the fourth stage), which does not require a significant increase in the computational complexity, compared with the requirements of HOG. The process of using nine bin-related images to extract the enhanced feature is similar to the process of extracting edge information of the downscaling image, which can retain most of the edge gradient information. Furthermore, the 8×8 block (used by the enhanced feature) and the 16×16 block (used by the HOG feature) are used to compute the orientation histograms in a 64×128 detection window, which has more large-scale information than the original HOG.

2.3. **Candidate verification.** After extracting the MHOG features, the dot product of the feature vector and the weight vector are obtained by Equation (13), and the corresponding result is determined by Equation (14). The equations are shown as below:

$$I = \sum_{i=1}^{K} v_i \times w_i \tag{13}$$

$$R = \begin{cases} 1, & \text{If } I > T \text{ pedestrian} \\ 0, & \text{If } I < T \text{ non-pedestrian} \end{cases} \tag{14}$$

where v_i is the value of the i-th feature vector, w_i is the value of the i-th weight vector, I is the dot product result, and T is the threshold. By Equation (13) and Equation (14), we can determine the result windows in the current pyramid layer. If there are many result windows detected in an image, the mean-shift method [26] is applied to merge all of these windows. Figure 8 shows the detection result and the windows-merging result.

FIGURE 8. Detection result of example images: (a) before windows merging; (b) after windows merging

TABLE 1. Pedestrian dataset in the experiment

	Training subset		***Testing subset***	
	Positive	***Negative***	***Positive***	***Negative***
Night-time	1998	291	2370	428
Day-time	2416	1218	1126	435

3. **Dataset.** For our evaluation, we focused on night-time pedestrian detection. Since there is no public dataset for night-time images, we built such a dataset, named Night-time Pedestrian Dataset (NTPD). The night-time images are captured by the normal Near-Infrared (NIR) camera in night-time scenes. As shown in Table 1, the training subset contains 1998 positive samples of 64×128 pixels and 291 negative samples of 640×480 pixels, and there are 2370 positive samples of 64×128 pixels and 428 negative samples of 640×480 pixels used for testing. In the meantime, the day-time dataset used in this study is the INRIA dataset [27], which contains 2416 positive samples and 1218 negative samples in the training subset, with 1126 positive samples and 435 negative samples applied for the test. Figure 9 shows some pedestrian samples and non-pedestrian samples in night-time environments.

(a)

(b)

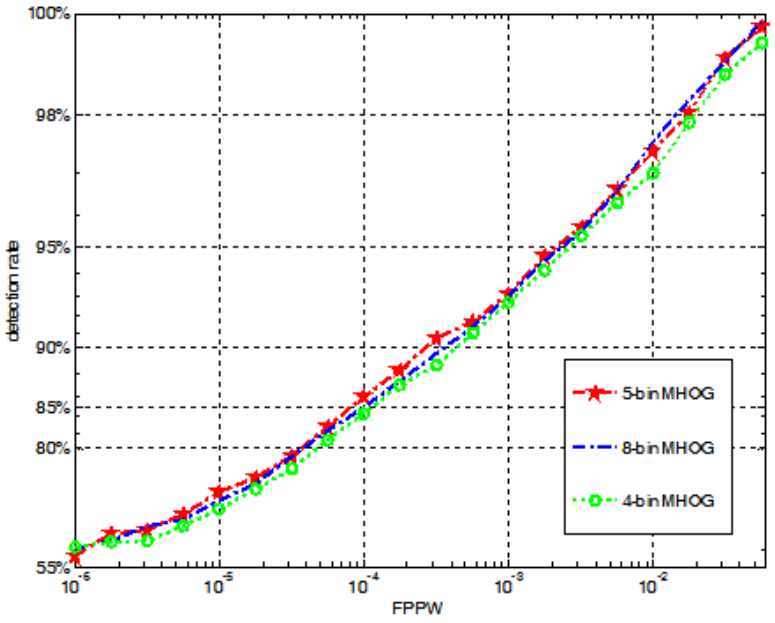
FIGURE 9. Some night-time image samples: (a) pedestrian samples; (b) non-pedestrian samples

4. **Experimental Result.**

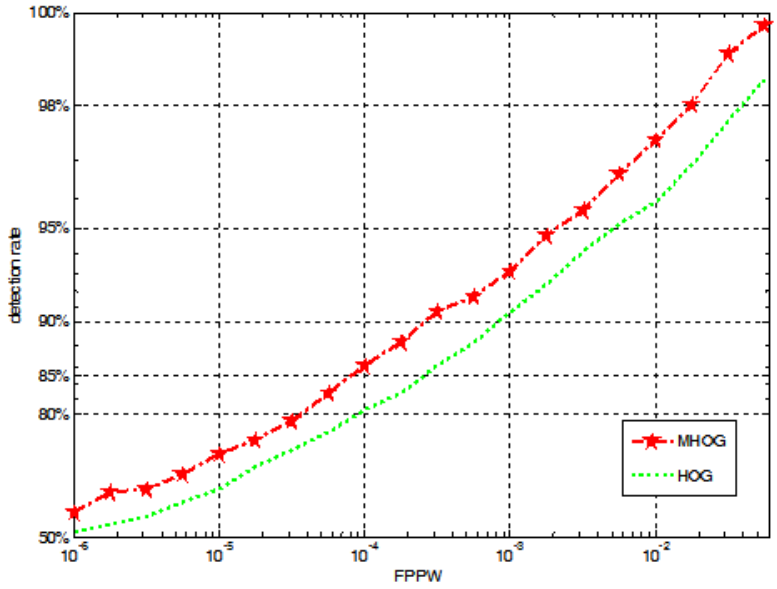
4.1. **Feature comparison.** We evaluated the performance of MHOG by applying our method to two pedestrian image datasets: our NTPD dataset and the INRIA dataset. The INRIA dataset contains human images cropped to 64×128 pixels and non-human images of various sizes. The details of these two datasets are shown in Table 1. Comparative experiments were performed on each of the two datasets, and evaluation results are described by a Receiver Operating Characteristic (ROC) curve. In this paper, a linear SVM classifier [28] is utilized.

To demonstrate that the C-HOG descriptor that uses 5-bin histogram in our method (5-bin MHOG) works best, we tested the 5-bin, 4-bin, and 8-bin MHOG on our NTPD dataset. The experimental result is shown in Figure 10(a). At the 10^{-4} FPPW, the 5-bin MHOG descriptor achieves the best detection rate of 86.03%, which outperforms the 4-bin MHOG descriptor and the 8-bin MHOG descriptor by 1.68% and 1.09%, respectively. In the following experiments, the 5-bin MHOG descriptor will be used as the default option.

Figure 10(b) illustrates the performance of the MHOG feature and the HOG feature on our NTPD dataset. The detection rate of the MHOG feature can reach 86.03% at 10^{-4} FPPW, which increases the accuracy by 5.35% more than that of the HOG descriptor. In



(a)

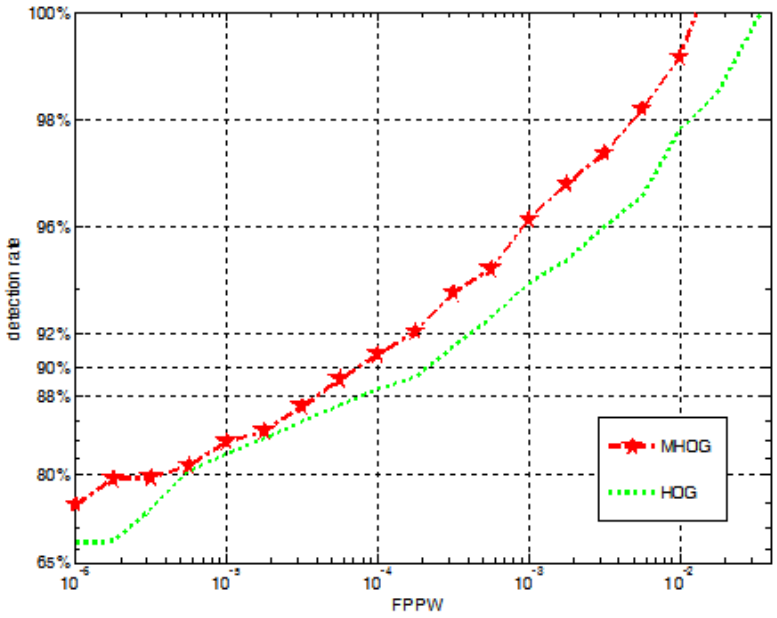


(b)

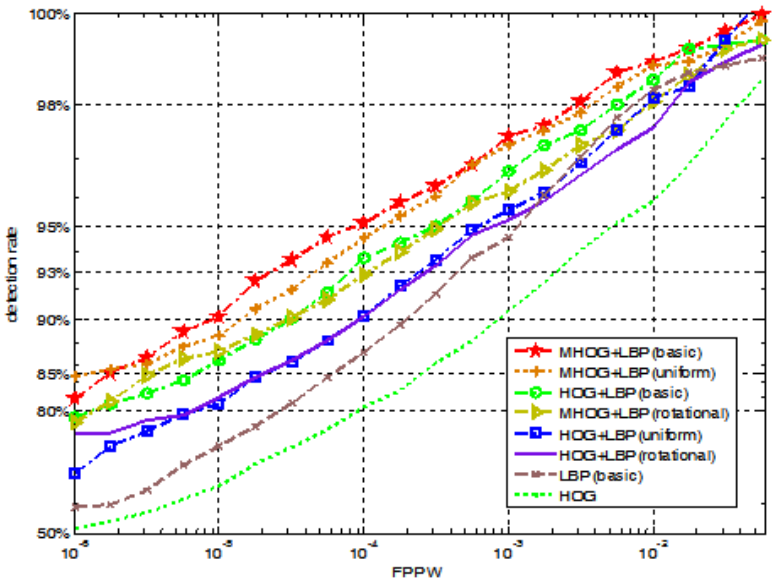
FIGURE 10. Performance comparisons of the combinational features or single feature: (a) 5-bin MHOG, 8-bin MHOG, 4-bin MHOG on the NTPD dataset; (b) MHOG and HOG on the NTPD dataset

order to ensure that the MHOG feature outperforms the original HOG feature, we also tested it in the day-time environment. Figure 11(a) shows the comparative experiment on the INRIA dataset. The MHOG descriptor achieves a 2.22% higher detection rate (90.85%) than the HOG descriptor (88.63%) at 10^{-4} FPPW.

From the evaluation results, it is known that the MHOG feature achieves a significant improvement compared to the HOG descriptor for both night-time and day-time environments (NTPD/INRIA datasets). The detection rate of the MHOG descriptor gains more improvement over HOG for the night-time environment; we hypothesize that the reason is the poor data condition of the night-time images.



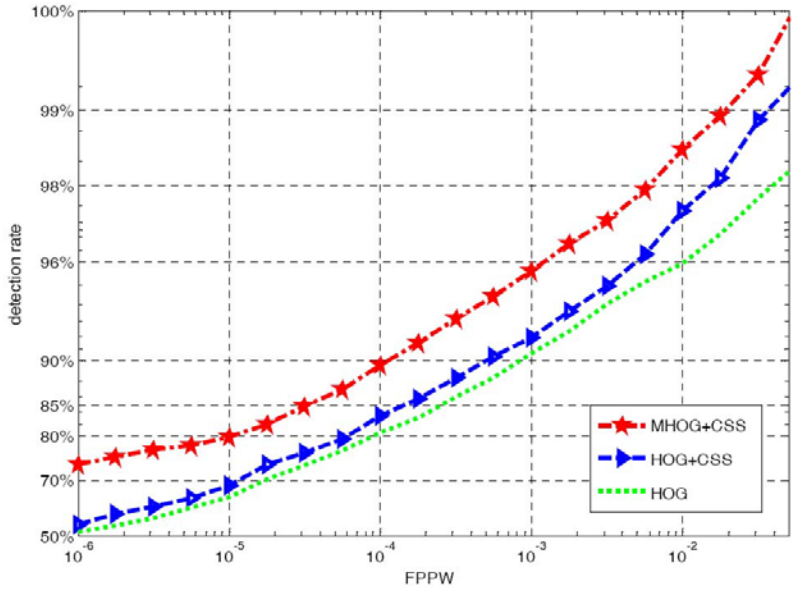
(a)



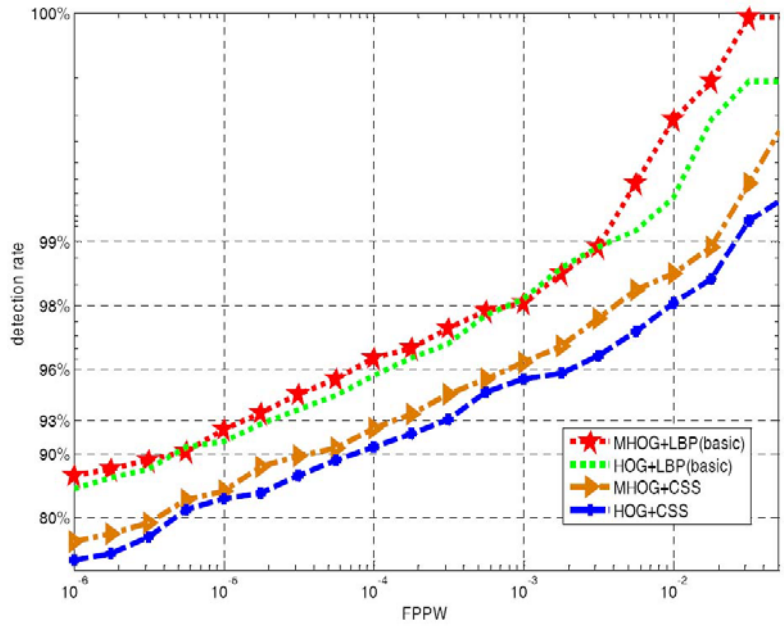
(b)

FIGURE 11. Performance comparisons of the combinational features or single feature: (a) MHOG and HOG on the INRIA; (b) adding MHOG or HOG to the three patterns of LBP on the NTPD dataset

Furthermore, to achieve the optimal detection rate, we combined the MHOG detector with the other detector on the NTPD dataset. Figure 11(b) shows the improvement gained by adding MHOG and HOG to three patterns of LBP. At the 10^{-4} FPPW, MHOG+LBP (basic pattern) achieves the highest detection rate of 95.15% and provides an improvement of 1.43% over HOG+LBP (basic pattern). MHOG+LBP (uniform pattern) can obtain a 4.43% higher detection rate compared with HOG+LBP (uniform pattern), and an improvement of 2.61%, by comparing with HOG+LBP (rotational invariance pattern) is obtained by MHOG+LBP (rotational invariance pattern). We also note that adding the color self-similarity (CSS) feature [29] provides better performance for MHOG than for



(a)



(b)

FIGURE 12. Performance comparisons of the combinational features or single feature: (a) adding CSS to MHOG or HOG on the NTPD dataset; (b) adding MHOG and HOG to CSS or LBP (basic) on the INRIA dataset

the original HOG. As shown in Figure 12(a), at 10^{-4} FPPW, MHOG+CSS consistently outperforms HOG+CSS by 6.16% and is much higher than the HOG features.

Finally, we combined MHOG and HOG with LBP (basic) or CSS, and tested them on the INRIA dataset for assessing the effects of the proposed method. Figure 12(b) presents the results for each case. At 10^{-4} FPPW, MHOG+LBP (basic) achieves the best detection rate of 96.45%, which is a small improvement (0.71%) over HOG+LBP (basic). MHOG+CSS reaches a 92.45% detection rate that produces an improvement of 2% over HOG+CSS.

4.2. **The parameters optimization and system performance evaluation.** For the ROI generation module, the parameters optimization method [18] was employed in our experiment. According to Equations (1) and (2), when the contrast of the pedestrian and background is small, false segmentation can be caused by the threshold T_H without optimization. Equation (1) causes T_H to become too great to obtain a correct segmentation. Therefore, T_H should be adaptive to the different values of T_L, which permit the object regions of the pedestrians with uniform or nonuniform brightness to be segmented properly. The final optimized T_H can be calculated as follows:

$$T_H(i,j) = \max\{T_1(i,j), T_L(i,j)\} \tag{15}$$

$$T_1(i,j) = \min\{T_2(i,j), 230\} \tag{16}$$

$$T_2(i,j) = \min\{T_3(i,j), T_L(i,j) + 8\} \tag{17}$$

$$T_3(i,j) = \max\{1.06 \times (T_L(i,j) - \alpha), T_L(i,j) + 2\} \tag{18}$$

where we set the initial values: $T_1 = 0$ and $T_2 = 0.5$. Next, on the basis that the width distribution of pedestrians is between $\alpha = 2$ and $\beta = 8$ in our experimental dataset, ω is initially set to 12 (see Equations (1) and (2)). Finally, for candidate verification, the pedestrian and non-pedestrian can be judged by Equations (13) and (14). Following extensive experimentation, the best results can be achieved when the threshold T is set to be 0.5.

After the optimization procedure, we run the resulting detector on both suburban and cluttered urban scenarios to evaluate the system performance. All the experiments are carried out on an AMD Phenom^{TM} II X4 965 Processor 3.4-GHz computer with Radeon E6760 GPU and 16-GB RAM, and the multi-threaded and GPU enabled training code are used to accelerate them.

Since there is no public dataset for night-time pedestrian detection, it is difficult to establish a baseline for comparing the various approaches from the published performance alone. The performance is influenced by many factors (e.g., training samples, test criteria, and testing data), and therefore the performance between different systems cannot be directly compared. To measure the performance of pedestrian detection in NIR videos, we are most interested in three key parameters: recall, precision, and detection speed. All of these parameters are reasonable for evaluating the performance of systems. They apparently show the performance that can be perceived by users, and the following equations are used:

$$\text{Recall} = \frac{\text{pedestrians detected}}{\text{pedestrians missed} + \text{pedestrians detected}} \tag{19}$$

$$\text{Precision} = \frac{\text{pedestrians detected}}{\text{pedestrians detected} + \text{false alarms}} \tag{20}$$

where Recall is also called the detection rate/hit rate. The proposed system detects pedestrians at 10m – 60m away from the vehicle, and is tested on video sequences under urban and suburban scenarios. In urban scenes, there are quite a lot of variations in scene, size and pose of people, as well as the cluster of the background, which results in much lower detection performance than that in a suburban environment. Some detection results from urban and suburban videos are shown in Figure 13, where we find a problem that concerns groups of pedestrians: if pedestrians are very close to each other and at the same distance from the vision system, they are often detected as a single pedestrian. Other detection failures are due to occlusions. Since processing time strongly depends on the complexity of the scene, both urban and suburban sequences are also used to evaluate temporal performance. The statistics of some of our experiments are given in Table 2. The results show that the average recall rate and precision rate is 90% and 88%

FIGURE 13. Pedestrian detection results under different scenarios: Upper row is detection results in suburban and lower row is detection results in urban

TABLE 2. Statistics of experimental results

System Performance							
Videos Pedestrian Detection Results					*Processing Time (ms)*		
Video	***Video length***	***Detected/Missed/ False alarm***	***Recall***	***Precision***	***ROI Generation***	***Classification***	***Total***
suburban	20m15s	52/4/5	0.93	0.91	14.36	8.15	22.51
urban	10m23s	218/31/37	0.87	0.85	18.25	12.09	30.34
Average			0.90	0.88	16.31	10.12	26.43

respectively, and the average processing time is about 26ms per frame, which reveals that the proposed system is robust, efficient and fast enough for real-time constraints. These results suggest that our method can meet the practical applications.

5. **Conclusion.** This paper describes a night-time pedestrian detection system. Much emphasis has been placed on feature extraction and real-time processing requirements. For effective sub-windows scanning, a region of interest (ROI) selection strategy is introduced. In order to capture more information of night-time pedestrians, we propose a novel feature extraction method that concatenates an enhanced feature and the original HOG to form the MHOG feature for night-time pedestrian detection. It applies the 8×8 block and the 16×16 block to compute the histograms in a 64×128 detection window, which has more large-scale information than the original HOG. The MHOG feature can improve the detection rate compared with the HOG descriptor not only for night-time environments (NTPD dataset), but may also be used for day-time environments (INRIA dataset). Experimental results under different night-time scenarios prove that our algorithm is robust, efficient and fast enough for real-time constraints.

For future work, we will improve the MHOG feature extraction method in terms of multi-resolution to increase the detection rate. Our study includes further performance improvement by an optimum combination of detection and tracking used to address some particularly challenging cases (e.g., pedestrians are very close to each other or occlusive), pedestrian detection based on leg movement, and the combination of a motion-based method and SVM classification.

Acknowledgment. We will thank PKU-HKUST ShenZhen-HongKong Institution and Shenzhen Mele Digital Technology Ltd. for providing the support and also thank our

colleagues in Mobile Video Network Technology Center for helping to collect the hand datasets. All works are supported by the technology development program of Shenzhen (No. CXZZ20120831104503786), China.

REFERENCES

[1] D. Geronimo, A. M. Lopez, A. D. Sappa and T. Graf, Survey of pedestrian detection for advanced driver assistance systems, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.32, no.7, pp.1239-1258, 2010.

[2] M. Enzweiler and D. M. Gavrila, Monocular pedestrian detection: Survey and experiments, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.31, no.12, pp.2179-2195, 2009.

[3] B. Li, Q. M. Yao and K. F. Wang, A review on vision-based pedestrian detection in intelligent transportation systems, *Proc. of the 9th IEEE Conf. on Networking, Sensing and Control*, pp.393-398, 2012.

[4] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp.886-893, 2005.

[5] Q. Zhu, M.-C. Yeh, K. T. Cheng and S. Avidan, Fast human detection using a cascade of histograms of oriented gradients, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp.1491-1498, 2006.

[6] S. Maji, A. C. Berg and J. Malik, Classification using intersection kernel support vector machines is efficient, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp.1-8, 2008.

[7] W. Ouyang and X. G. Wang, A discriminative deep model for pedestrian detection with occlusion handling, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp.3258-3265, 2012.

[8] Y. Luo, J. Remillard and D. Hoetzer, Pedestrian detection in near-infrared night vision system, *Proc. of the IEEE Intelligent Vehicles Symposium*, pp.51-58, 2010.

[9] Y. J. Fang, K. Yamada, Y. Ninomiya, B. Horn and I. Masaki, Comparison between infrared-image-based and visible-image-based approaches for pedestrian detection, *Proc. of the IEEE Intelligent Vehicles Symposium*, pp.505-510, 2003.

[10] S. L. Chang, F. T. Yang, W. P. Wu, Y. A. Cho and S. W. Chen, Nighttime pedestrian detection using thermal imaging based on HOG feature, *Proc. of the IEEE Conf. on System Science and Engineering*, pp.694-698, 2011.

[11] H. Nanda and L. Davis, Probabilistic template based pedestrian detection in infrared videos, *Proc. of the IEEE Intelligent Vehicle Symposium*, pp.15-20, 2002.

[12] Y. Cao, S. Pranata and H. Nishimura, Local binary pattern features for pedestrian detection at night/dark environment, *Proc. of the IEEE Conf. on Image Processing*, pp.2053-2056, 2011.

[13] Y.-C. Lin, Y.-M. Chan, L.-C. Chuang, L.-C. Fu, S.-S. Huang, P.-Y. Hsiao and M.-F. Luo, Near-infrared based nighttime pedestrian detection by combining multiple features, *Proc. of the 14th IEEE Conf. on Intelligent Transportation Systems*, pp.1549-1554, 2011.

[14] Q. Tian, H. Sun, Y. Luo and D. Hu, Nighttime pedestrian detection with a normal camera using SVM classifier, *Proc. of the 2nd International Symposium on Neural Networks*, vol.3497, pp.189-194, 2005.

[15] H. Sun, C. Wang and B. Wang, Night vision pedestrian detection using a forward-looking infrared camera, *Proc. of the International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping*, pp.1-4, 2011.

[16] F. Xu, X. Liu and K. Fujimura, Pedestrian detection and tracking with night vision, *IEEE Trans. Intelligent Transportation Systems*, vol.6, no.1, pp.63-71, 2005.

[17] M. Bertozzi, A. Broggi, C. Caraffi, M. Del Rose, M. Felisa and G. Vezzoni, Pedestrian detection by means of far-infrared stereo vision, *Comput. Vis. Image Underst.*, vol.106, no.2/3, pp.194-204, 2007.

[18] J. F. Ge, Y. P. Luo and G. M. Tei, Real-time pedestrian detection and tracking at nighttime for driver-assistance systems, *IEEE Trans. Intelligent Transportation Systems*, vol.10, no.2, pp.283-298, 2009.

[19] T. Kancharla, P. Kharade, S. Gindi, K. Kutty and V. G. Vaidya, Edge based segmentaion for pedestrian detection using NIR camera, *Proc. of the International Conf. on Image Information Processing*, pp.1-6, 2011.

[20] J. F. Ge, Y. P. Luo and D. Xiao, Adaptive hysteresis thresholding based pedestrian in nighttime using a normal camera, *Proc. of the IEEE Conf. on Vehicular Electronic and Safety*, pp.46-51, 2005.

[21] J. F. Dong, J. F. Ge and Y. P. Luo, Nighttime pedestrian detection with near infrared using cascaded classifiers, *Proc. of the IEEE Conf. on Image Processing*, pp.VI-185-VI-188, 2007.

[22] D. G. Lowe, Object recognition from local scale-invariant features, *Proc. of the IEEE Conf. on Computer Vision*, pp.1150-1157, 1999.
[23] K. Levi and Y. Weiss, Learning object detection from a small number of examples the importance of good feature, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp.II-53-II-60, 2004.
[24] S. Belongie, J. Malik and J. Puzicha, Shape context: A new descriptor for shape matching and object recognition, *Neural Information Processing Systems*, pp.831-837, 2000.
[25] D. Wei, Y. Zhao, R. Cheng and G. Li, An enhanced histogram of oriented gradient for pedestrian detection, *Proc. of the International Conf. on Intelligent Control and Information Processing*, pp.459-463, 2013.
[26] F. Keinosuke and L. Hostetler, The estimation of the gradient of a density function, with applications in pattern recognition, *IEEE Trans. Information Theory*, vol.IT-21, no.1, pp.32-40, 1975.
[27] *http://pascal.inrialpes.fr/data/human/*.
[28] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang and C. J. Lin, LIBLINEAR: A library for large linear classification, *Journal of Machine Learning Research*, vol.9, pp.1871-1874, 2008.
[29] S. Walk, N. Majer, K. Schindler et al., New features and insights for pedestrian detection, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp.1030-1037, 2010.