Nighttime Pedestrian Detection Using Local Oriented Shape Context Descriptor

Guoliang Li, Yong Zhao, Daimeng Wei, Ruzhong Cheng
The key laboratory of Integrated Microsystems
Shenzhen Graduate School, Peking University
Shenzhen, China
yongzhao@pkusz.edu.cn

***Abstract*—In this paper, we propose a novel feature based on local oriented shape context (LOSC) descriptor for nighttime pedestrian detection. Shape context descriptor is widely used in object recognition but not making a good performance in complex situations because it does not consider the edge's orientation. To address this limitation, our method adds orientation information to the shape context descriptor. We first compute the image gradient in nine directions and then extract shape context descriptor in each direction. Finally we put the feature vector to linear SVM for training. We tested this descriptor's performance on our nighttime pedestrian samples captured by normal night-vision camera. The experiment results show that our method achieved high detection rate and had fewer dimensions than the HOG descriptor.**

Keywords-Nighttime Pedestiran Detection; Local Oriented Shape Context Descriptor; HOG; SVM; Night-Vision Camera

I. INTRODUCTION

Traffic accidents happen very frequently every year, and about 30 percent of them are related to pedestrian collision. This is especially true in the evenings when drivers' concentration drops and ignores the obstacles ahead at the night/dark environment. In Europe and America, many universities and research institutes have been doing relevant research. Some car manufactures have also been developing ADAS (Advanced Driver Assistant System) over the last ten years [1] and some night-time pedestrian detection devices have been installed on the car such as BMW 7 series.

Most prior pedestrian detection studies were based on the daytime environment [2, 3]. The main algorithm is about finding a feature of pedestrian that has better performance and is robustness. Compared with the daytime, nighttime pedestrian detection is more difficult as a result of low contrast, image blur, and image noise.

Most night-vision-based pedestrian detection uses NIR (Near-Infrared) camera or FIR (Far-Infrared) camera [4, 5, 6], while others are based on the thermal image. In general, the process of nighttime pedestrian detection includes two stages: ROI (Region-Of-Interest) segmentation and candidate verification. The purpose of segmentation is to reduce the scanned area for next the stage. In candidate verification stage, the approaches of detection can be divided into appearance feature extraction and template matching. Nanda and Davis [7] introduced a simple yet effective idea of probabilistic templates for pedestrian detection. Just like daytime pedestrian detection, HOG (Histograms of Oriented Gradients) feature [8, 9] is also used for nighttime pedestrian detection very often. Cao [10] propose a modified LBP (Local Binary Pattern) feature extraction method for pedestrian detection in night/dark environment. Lin and Chan [11] combined two significant features, HOG and contour, and used SVM to build up a reliable classification system. Sun [12] uses Haar-like features to discriminate infrared pedestrians. Cerri [13] proposed ad-hoc features for both daytime and nighttime pedestrian detection.

These methods are very complex and have a large number of dimensions of feature vector. In this paper, we propose a new descriptor for nighttime pedestrian detection inspired by the shape context and the HOG descriptor. In contrast to prior methods, our detector uses a simpler architecture with a single detection window.

The rest of this paper is organized as follows: Section 2 describes the whole system structure; In section 3, we present our algorithm of nighttime pedestrian detection using local oriented shape context in detail; Section 4 displays the results of our experiments and analysis of the effects of different methods; In the last section, we summarize the conclusions and discuss further research directions.

II. OVERVIEW

In this section, we introduce the whole system structure. In most nighttime pedestrian detecting systems, the process is divided into four main parts as shown in Fig.1: 1. Image preprocess; 2.ROI (region of interesting) selection; 3. Feature extraction; 4.Candidate verification. In this paper, we focus on how to extract the features for nighttime pedestrian.

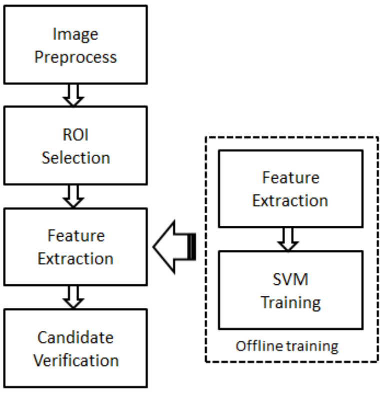


Figure 1. The whole system structrue.

Published by Atlantis Press, Paris, France.

Offline training uses linear SVM classifier to train our proposed feature. The nighttime pedestrian samples were collected by normal night-vision camera. The camera has two built-in star power near-infrared lamps.

At the ROI Selection Stage, we used a simple adaptive dual-threshold segmentation algorithm proposed by Ge [14] for our original images.

III. Feature Extraction

A. *LOSC Descriptor Extraction*

The object's appearance and shape can be well characterized by shape context [15] descriptor but it doesn't have orientation information of edge energy because the basic shape context descriptor does not include gradient orientation information. Inspired by HOG feature [8], we propose a new descriptor based on the basic shape context that we find a way to add gradient orientation information to it. And we call this descriptor as local oriented shape context. The feature is extracted in four steps: 1.Gradient computation; 2.Voting for nine orientation bins; 3. Shape context extraction; 4. Contrast normalization.

1) Compute gradients

The way of gradient computation is sensitive to the detector performance. We tested different kinds of mask and found that the simplest mask [-1, 0, 1] turned out to be the best mask for gradient computation in our feature. The gradient is computed as:

$$E = \sqrt{dx^2 + dy^2} \quad (1)$$

$$\theta = \arctan(dy/dx) \quad (2)$$

Where E is for edge energy and θ is for edge direction. We compute dx and dy by scanning all pixels in image by mask

$$S_x = [-1 \quad 0 \quad 1] \quad (3)$$

$$S_y = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \quad (4)$$

Where S_x is horizontal mask and S_y is vertical mask.

2) Vote for nine orientation bins

The next step is essential of the whole descriptor extraction. We first sliding compute an orientation histogram of 8×8 pixel block with a 4 pixel stride (hence 4 pixel fold coverage of each block) in 64×128 pixel image. There are nine orientation bins from 0° to 180° in each histogram. In every block, each pixel calculated a weighted vote for an edge orientation histogram channel, and the votes were summed together into its orientation bin. Here we have nine 15×31 image as showing in Fig.2. The pixel value of each image is the accumulated votes of the orientation bin.

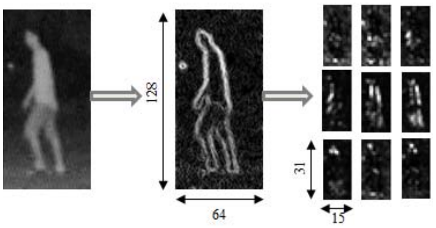


Figure 2. Multi-orientaion edge energy.

We used the orientation range from 0° to 180° (ignore 'sign') instead of orientation range from 0° to 360° because signed gradients decrease the performance. Increasing the number of orientation bins improve the performance up to 9 bins. The value of pixels in each block voted to its bin is the edge energy at this point. And each pixel votes two bins by linear interpolation between the neighboring bin centers in both orientation and position.

We test the different size of scanning block with different stride of block (see Fig. 7(c)). The results show that 8×8 pixel block with a 4 pixel stride is the best one to choose.

3) Extract shape context

As shown in Fig.3, for each 8×8 block in nine 15×31 orientation bins we extracted a shape context descriptor. In this descriptor the distribution of locally normalized gradient orientations was captured in a log-polar histogram. The log-polar bin was tolerant to small changes in the rotation of the body parts. We used 2 bins for location and 4 bins for gradient orientation, which generates a 2×4 =8 dimensions descriptor for each block. We ignored the sign of the gradient as we found this can improve generalization.

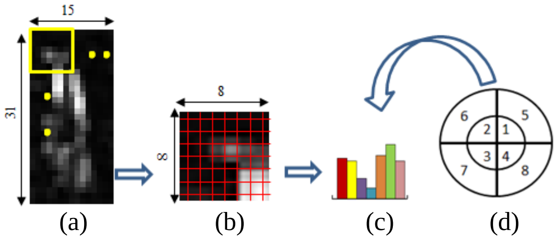


Figure 3. Procedure for extract shape context in one orientation bin: (a) one orientation bin; (b) one 8×8 block; (c) a 8-D vecotor; (d) eight bins of gradient orientaion and location.

Here we used trilinear interpolation while voting to eight bins. As shown in Fig.4 (a) every pixel in block first was assigned to four orientations by bilinear interpolation like HOG [8, 16], then the votes for each orientation interpolated linearly (by the distance to the center) to the two bins (see Fig.4 (b)). First we know the distribution of point 'o' on four orientations (o1,o2,o3,o4), then point 'o' voting for bin 'B1' and 'B2' is $o1 \times r/R$ and $o1 \times (R - r)/R$, where R is the radius of bigger location bins and r is the distance of point 'o' to the block center.

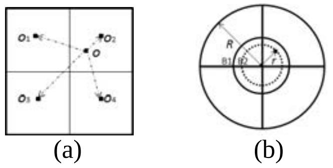


Figure 4. Trilinear interpolation. (a)Bilnear interpolation to four orientations. (b) Linear interpolation to two same orientation bins.

We tested the different stride of block, experiment results show that 4 pixel improve the performance significantly. As a result, we got a 21×8 =168 dimensions vector for each orientation bin, and total 168×9 =1512 dimensions descriptor for a 64×128 image.

4) Normalization

Published by Atlantis Press, Paris, France.
© the authors

Normalization was used to reduce the illumination variability. After extracting a shape context descriptor, we evaluated three different block normalization schemes for our LOSC descriptor. The L1-norm and L2-norm schemes are showed as follows:

$$p \to p/(\|p\|_1 + \xi) \tag{5}$$

$$p \to p/\sqrt{\|p\|_2^2 + \xi^2} \tag{6}$$

Where p is the unnormalized descriptor vector, $\|p\|_k$ is k-norm for k = 1, 2, and ξ is a small constant. The L2-Hys method is based on L2-norm followed by clipping and renormalization. The experiment result (see Fig.7(b)) shows that using L2-norm method resulted in a better performance than the others.

B. Candidate Verification

After extracting LOSC descriptor, we do the dot product between the feature vector and the weight vector trained by SVM. The corresponding result is determined as

$$R = \begin{cases} 1, & \text{If } I > T \quad \text{pedestrian} \\ 0, & \text{If } I < T \quad \text{non-pedestrian} \end{cases} \tag{7}$$

Where I is the dot product result and T is the threshold value. If there are many result windows detected in an image, we used mean shift [17] method to merge all these windows. Fig.5 shows the final detection result in our system.

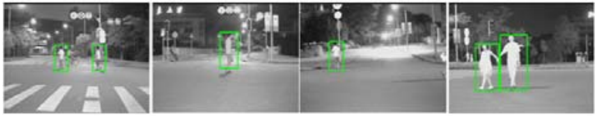


Figure 5. Detection result of some image after windows-merging.

After merging the neighboring windows, the total number of detection windows decreased a lot.

IV. EXPERIMENT

In this section, we introduce our dataset at first, and then list our experiment process, present the experiment results and analyze the result data.

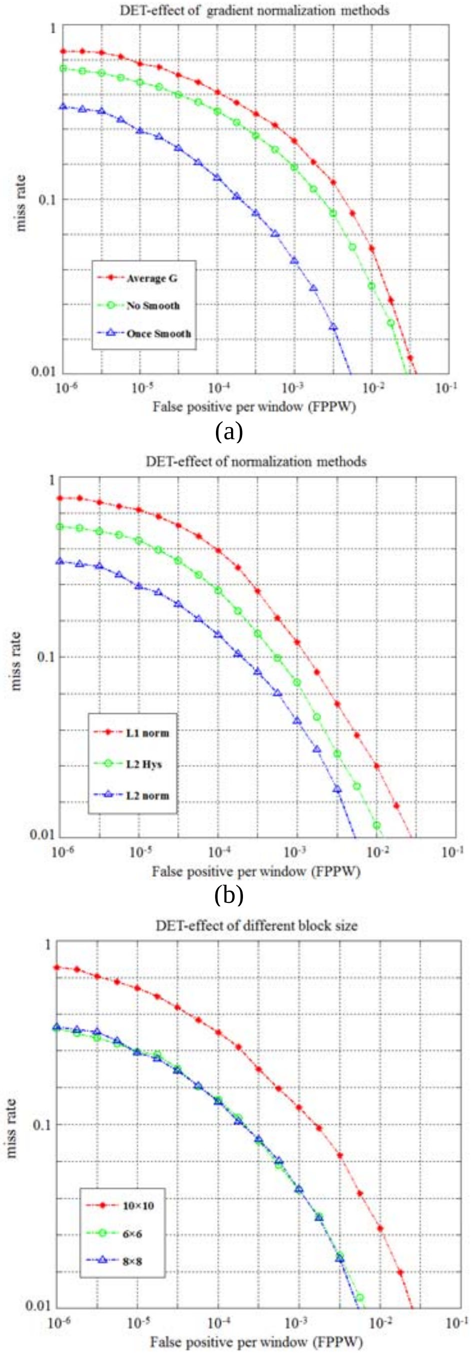
A. Dataset

Because there is no pubic dataset for nighttime pedestrian image, we tested our descriptor using our collecting image. These images were captured from a night-vision camera in different scenes at night time, in total 19 scenarios. At the training stage, we used 1,248 64×128 positive image samples and 16,313 64×128 negative image samples. And we used the rest 3,072 samples and 2×10^6 negative samples for testing. Fig.6 shows some pedestrian samples for training and testing.

Figure 6. Some seletecd nighttime pedestiran samples

B. Result

To control for confounding factors, all of our experiments used the same training samples and testing samples. At training stage, we used a soft (c=0.01) linear SVM trained with LIBLINEAR SVM [18]. When computing gradients, we tested four different mask for our descriptor and find that mask [-1,0,1] turn out to be the better one than the others. This mask reserves more details than other masks. At the descriptor extraction stage, we did lots of experiments by changing some parameters including bock size, step size, normalization methods, and smooth methods.



(a)

(b)

(c)

Figure 7. The performace comparision with different methods: (a) different gradient normalization methods at orientation bin's voting stage; (b) different normalization methods at shape context extracting stage; (c) different block size when scaning the input image at gradient computing stage.

As shown in Fig.7 (a), we tested three gradient normalization methods when voting for nine orientation bins. Using average gradient (sum of voting of each bin divide the total number which is the sum of pixels voted for this bin) decreases the performance about 28% at 10^{-4} FPPW, while every pixel just votes only one bin (no smooth) but decreases the performance about 19% at 10^{-4} FPPW.

At the descriptor normalization stage, we compared three different normalization methods that have been described in section 3. Fig.7 (b) shows that using L2-norm is 10% and 26% better than L2-Hys and L1-norm separately at 10^{-4} FPPW. And the L2-Hys normalization method is greatly influenced by the clipping variable parameter.

Fig. 7 (c) shows the result of using different block size when sliding computing an orientation histogram in 64×128 pixel image. Using an 8×8 pixel block with a 4 pixel stride is 18% better than using a 10×10 pixel block with a 5 pixel stride. It has a near similar performance with using a 6×6 pixel block with a 3 pixel stride but creates less-dimension vector than it.

V. Conclusion

As the experiment results showed, the local oriented shape context descriptor (LOSC) we proposed has a good performance at nighttime pedestrian detection. And the number of dimension of feature vector is smaller than HOG and other complex features. We studied the influence of various descriptor parameters to find strong gradient normalization, local contrast normalization and better block size for good performance. We used a simple adaptive dual-threshold method for ROI selection and mean shift method for merging detection windows. The results show that the system achieved a high rate of pedestrian detection at nighttime.

Our descriptor just a simple descriptor and the system algorithm have not done very well at the daytime environment. So we will try to use multi-features base on combining our proposed LOSC descriptor for a better performance at both daytime and nighttime in further work. And we try to use adaboost and other multi-classifiers for speeding up the detection rate while keeping a high detection rate. The situations that some pedestrians may be occluded by other pedestrians or objects are the research difficulties. How to solve this problem is also one of our research emphases in future.

References

[1] David Gero´nimo, Antonio M. Lo´pez, Angel D. Sappa, "Survey of Pedestrian Detection for ADAS," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, Vol.32, No.7, pp.1239-1258.

[2] Enzweiler M., Gavrila, D.M., "Monocular Pedestrian Detection: Survey and Experiments," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, Vol.31, No.12, pp.2179-2195.

[3] Bo Li, Qingming Yao, Kunfeng Wang, "A review on vision-based pedestrian detection in intelligent transportation systems," 2012 9th IEEE International Conference on Networking, Sensing and Control (ICNSC 2012), pp.393-398.

[4] Yun Luo, Jeffrey Remillard, Dieter Hoetzer, "Pedestrian Detection in Near-Infrared Night Vision System." IVS(Intelligent Vehicles Symposium), 2010.

[5] Q. Tian, H. Sun, et al., "Nighttime Pedestrian Detection with a Normal Camera Using SVM Classifier," Advances in neural networks (ISNN 2005), Vol.3497/2005, Dol.10.1007.

[6] Y. Fang, K. Yamada, et al., "Comparison between Infrared-image-based and Visible-image-based Approaches for Pedestrian Detection," Intelligent Vehicles Symposium, 2003, pp.505-510.

[7] Harsh Nanda and Larry Davis, "Probabilistic Template Based Pedestrian Detection in Infrared Videos," Intelligent Vehicle Symposium, 2002, IEEE, Vol.1 pp.15-20.

[8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005) , San Diego, USA, pp.886-893, June 2005.

[9] S. Chang, F. Yang, W. Wu, et al., "Nighttime Pedestrian Detection Using Thermal Imaging Based on HOG Feature," Proceeding of 2011 International Conference on System Science and Engineering, Macau, China, June 2011.

[10] Yunyun Cao, Sugiri Pranata, et al., "Local Binary Pattern features for pedestrian detection at night/dark environment," 2011 18th IEEE International Conference on Image Processing, pp. 2053-2056.

[11] Yu-Chun Lin, Yi-Ming Chan, et al., "Near-Infrared Based Nighttime Pedestrian Detection by Combining Multiple Features," 2011 4th International IEEE Conference on intelligent Transportation Systems Washington, DC, USA, October 5-7, 2011.

[12] H. Sun, C. Wang, B. Wang, "Night vision pedestrian detection using a forward-looking infrared camera," Multi-Platform/Multi-Sensor Remote Sensing and Mapping (M2RSM), 2011 International Workshop.

[13] Pietro Cerri, Luca gatti, et al., "Day and Night Pedestrian Detection Using Cascade AdaBoost System," 2010 13th international IEEE Annual Conference on Intelligent Transportation Systems, Madeira Island, Portugal, September 19-22, 2010.

[14] J. Ge, Y. Luo, G. Tei, "Real-Time Pedestrian Detection and Tracking at nighttime for Driver-Assistance Systems," IEEE Transactions on Intelligent Transportation Systems, Vol.10, No.2, June 2009.

[15] S.Belongie, J. Malik and J. Puzicha, "Shape Context: A new descriptor for shape matching and object recognition," NIPS 2000.

[16] S. Maji, A. Berg and J. Malik, "Classification using intersection kernel support vector machines is efficient," CVPR 2008.

[17] F. Keinosuke, L. D. Hostetler, "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition," IEEE Transactions on Information Theory, 1975.

[18] R. E. Fan, K. W. Chang, et al., "LIBLINEAR: A library for large linear classification," Journal of Machine Learning Research, 2008, pp.1871-1874