World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:10, No:1, 2016

Fused Structure and Texture (FST) Features for Improved Pedestrian Detection

Hussin K. Ragb, Vijayan K. Asari

***Abstract*—**In this paper, we present a pedestrian detection descriptor called Fused Structure and Texture (FST) features based on the combination of the local phase information with the texture features. Since the phase of the signal conveys more structural information than the magnitude, the phase congruency concept is used to capture the structural features. On the other hand, the Center-Symmetric Local Binary Pattern (CSLBP) approach is used to capture the texture information of the image. The dimension less quantity of the phase congruency and the robustness of the CSLBP operator on the flat images, as well as the blur and illumination changes, lead the proposed descriptor to be more robust and less sensitive to the light variations. The proposed descriptor can be formed by extracting the phase congruency and the CSLBP values of each pixel of the image with respect to its neighborhood. The histogram of the oriented phase and the histogram of the CSLBP values for the local regions in the image are computed and concatenated to construct the FST descriptor. Several experiments were conducted on INRIA and the low resolution DaimlerChrysler datasets to evaluate the detection performance of the pedestrian detection system that is based on the FST descriptor. A linear Support Vector Machine (SVM) is used to train the pedestrian classifier. These experiments showed that the proposed FST descriptor has better detection performance over a set of state of the art feature extraction methodologies.

***Keywords*—**Pedestrian detection, phase congruency, local phase, LBP features, CSLBP features, FST descriptor.

I. Introduction

PEDESTRIAN detection is considered as one of the motivated tasks in computer vision systems due to its importance in applications such as human computer interaction, visual surveillance, autonomous navigation, robotics, and automotive safety systems. This task however, is rather challenging because of the fluctuations in appearance of the human body and the cluttered scenes, pose, occlusion, and the illumination variations. In the last decade, several single feature algorithms were developed for the pedestrian detection systems. Some of the most popular features of these algorithms are the Histogram of Oriented Gradient (HOG), the Scale Invariant Feature Transform (SIFT), Edgelet, Haar, shapelet, the Local Binary Pattern (LBP), and the Histogram of Oriented Phase (HOP). Since the detection efficiency of the single feature based techniques are limited, it is envisaged that a representation based on multiple features would be more effective in capturing the pedestrians in the various background environments. Zhang and Ram [1] improved the detection performance of IR images by combining the Edgelets and HOG features. Wang et al. [2] combined the feature set HOG and LBP to improve the pedestrian detection performance. Wojek [3] also combined HOG, Haar and Shapelet features, etc. All of these combined algorithms improved the pedestrian detection performance, however they lack to deal effectively with the poor resolution images and cluttered background scenes. In this paper we propose new pedestrian detection features that fuse the structural and textual information of the image into one descriptor. The phase congruency concept is used to capture the structural information. The significance of the phase information proved by the experiment of [4] played an important role for using the phase congruency as a base of the structural features. This experiment has shown that, a phase of the image can carry more structural information than the magnitude does. Furthermore, phase congruency is a dimension-less quantity that makes it more robust to image scale changes as well as illumination and image contrast variations [5], [6], [10]. The center symmetric local binary pattern (CSLBP) algorithm [7] is used to capture the texture features and the gradients of the image. This algorithm is computationally simple. It has robustness on the flat image areas, blurring, and the illumination changes [7]. The traits of these algorithms lead the FST features proposed in this paper to be more precise, more robust, and less sensitive to light variations. The pedestrian detection framework based on the FST descriptor is shown in Fig. 1. The phase congruency (magnitude & orientation) and the CSLBP values of the local regions in the input image are computed. The histogram of oriented phase and the histogram of the CSLBP values are computed and fused for each local region. These histograms are normalized and concatenated to each other to construct the proposed FST descriptor. Once the FST features of the input image is computed, it is fed to the Support Vector Machine (SVM) classifier which is trained by the FST feature vectors of the positive and negative samples of the datasets. Finally, the SVM classifier will decide whether the objects in the input image belong to the pedestrian or non-pedestrian.

The remaining sections of this paper are organized as follows. In Section II, we discuss the structural features computation based on the phase congruency. In Section III, we describe how the texture information is captured using CSLBP operator. Then, in Section IV, we explain the construction of the FST descriptor. The performance evaluation and the experimental results are illustrated in Section V. Finally, the conclusion is presented in Section VI.

Hussin Ragb and Vijayan Asari are with the Department of Electrical and Computer Engineering at the University of Dayton, Dayton, Ohio, USA (phone: 937-229-3611; fax: 937-229-4529, 937-229-4504; e-mail: ragbh1@udayton.edu, vasari1@udayton.edu).

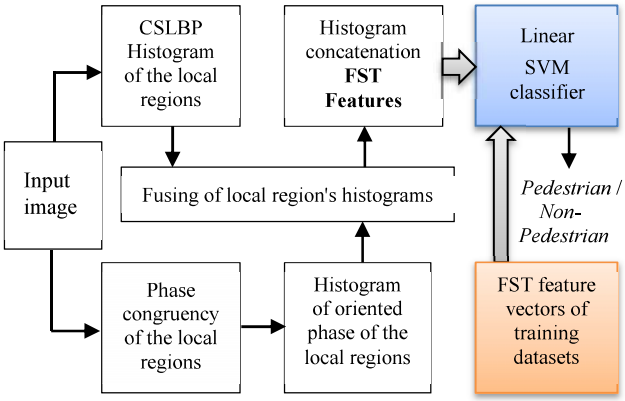


Fig. 1 The pedestrian detection framework based on the proposed FST descriptor

II. Structural Features Based on Phase Congruency

Phase congruency (PC) is an algorithm that was developed by Kovesi [5], [6] to detect and localize the structural information (edges and corners) of the digital images. The local energy model of feature detection [18] postulates that the features are perceived at the points of the strong phase congruency [8], [9].

Assume that an input periodic signal $I(x)$ is a one dimensional signal defined in the range $[-\pi, \pi]$ and $F(x)$ is the same as $I(x)$ filtered from a DC component. The local energy function $E(x)$ can be defined from $F(x)$ and the Hilbert Transform $F_H(x)$ as in (1).

$$E(x) = \sqrt{F(x)^2 + F_H(x)^2} \tag{1}$$

where, $F_H(x)$ is 90° phase shift of $F(x)$, (Hilbert Transform). Venkatesh and Owens [17] has shown that the energy is equal to the product of phase congruency PC and the sum of Fourier amplitudes A_n as in (2).

$$E(x) = PC(x) \sum_n A_n \tag{2}$$

Equation (2) shows that the peaks in the energy function correspond to the peaks in PC [8]. It shows also that the PC is independent of the overall magnitude of the signal. This makes the features invariant to scale, illumination, and contrast changes.

A. Computing the Phase Congruency

In order to compute the PC, we have to extract the local frequencies and phase information of the signal. This is done by convolving the two dimensional signal with a pair of quadrature filters. In this paper we are using Log-Gabor filter which is an efficient band pass filter to extract and localize the phase information spread over a broad spectrum. Log-Gabor filter does not have DC component and its frequency response can be represented by the following transfer function [8]:

$$G(\omega, \theta) = exp\left(\frac{-(log(\omega/\omega_o))^2}{2(log(k/\omega_o))^2}\right) exp\left(\frac{-(\theta-\theta_o)^2}{2\sigma_\theta{}^2}\right) \tag{3}$$

where ω_o is the center frequency of the filter, k is the frequency width parameter. The ratio k/ω_o should be kept constant for various ω_o. It is used for controlling the bandwidth of the filter. σ_θ is the standard deviation of the Gaussian function in angular direction, and θ_o represents the centre orientation of the filter [8], [11].

Starting with the one dimensional input signal $I(x)$, consider M_n^o and M_n^e are the odd symmetric and even symmetric components of the Log- Gabor filter at scale n and they form a quadrature pair. The convolution of each quadrature pair with the input signal $I(x)$ will form a response vector at position x on scale n as given in (4).

$$[e_n(x), o_n(x)] = [I(x) * M_n^e, I(x) * M_n^o] \tag{4}$$

Therefore; the amplitude $A_n(x)$ and the phase angle $\phi_n(x)$ of the response at scale n are given by:

$$A_n(x) = \sqrt{(e_n^2(x) + o_n^2(x))} \tag{5}$$

$$\phi_n(x) = tan^{-1}\left(\frac{o_n(x)}{e_n(x)}\right) \tag{6}$$

$F(x)$ and $F_H(x)$ in (1) can be computed as:

$$F(x) = \sum_n e_n(x) \tag{7}$$

$$F_H(x) = \sum_n o_n(x) \tag{8}$$

Therefore, the one dimensional phase congruency PC is given as:

$$PC(x) = \frac{E(x)}{\varepsilon + \sum_n A_n} \tag{9}$$

where ε is a small number to avoid a division by zero.

The phase congruency $PC(x, y)$ of the two dimensional signal $I(x, y)$ can be computed in the same manner. The response vector at scale n and orientation o is given by:

$$[e_{no}(x, y), o_{no}(x, y)] = [I(x, y) * M_{no}^e, I(x, y) * M_{no}^o] \tag{10}$$

where M_{no}^e, M_{no}^o is the even symmetric and odd symmetric components of the log-Gabor filter at scale n and orientation o. The amplitude at scale n and orientation o is given as:

$$A_{no} = \sqrt{(e_{no}^2(x, y) + o_{no}^2(x, y))} \tag{11}$$

As in (7) and (8), $F(x, y)$ and $F_H(x, y)$ are given by:

$$F(x, y) = \sum_o \sum_n e_{no}(x, y) \tag{12}$$

$$F_H(x, y) = \sum_o \sum_n o_{no}(x, y) \tag{13}$$

Therefore; PC of the image can be computed as:

$$PC(x, y) = \frac{\sum_o \sqrt{(\sum_n e_{no}(x,y))^2 + (\sum_n o_{no}(x,y))^2}}{\varepsilon + \sum_o \sum_n A_{no}(x,y)} \tag{14}$$

The orientation angle $\varphi(x,y)$ where the phase congruency occurs can be computed as:

$$\varphi(x,y) = tan^{-1}\left(\frac{F_H(x,y)}{F(x,y)}\right) \quad (15)$$

III. Texture Features Computation

In this paper, the center symmetric local binary pattern approach (CSLBP) is used to capture the texture as well as the gradient features of the image. The CSLBP algorithm is a modified version of the local binary pattern method (LBP). As illustrated in Fig. 2, CSLBP operator compares the gray level of the center symmetric pairs of the pixel, which represent the gradients of the image. The texture information is obtained by replacing each pixel in the image by the pixel's CSLBP value [12], [13]. The CSLBP features can be computed by:

$$CSLBP = \sum_{i=1}^{N/2} S(P_i - P_{i+N/2})\, 2^i \quad (16)$$

$$S(z) = \begin{cases} 1 & if\ z \geq t \\ 0 & otherwise. \end{cases} \quad (17)$$

where, P_i is the gray values of the neighbor pixel, N is the number of the neighbors (Fig. 2 shows neighbors for $N = 8$), t is threshold value ($t = 0.1\ is\ selected$).

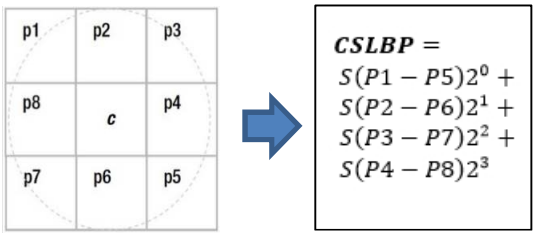


Fig. 2 The *CSLBP* descriptor

IV. Construction of the Fused Structural and Texture (FST) Descriptor

The input image $I(x,y)$ is divided into local regions, called *blocks,* in the size of 16×16 pixels. Each *block* is built from 2×2 *cells* with *cell* size 8×8 pixels. These *block*s are 50% overlapped. The phase congruency value and its orientation angle for each pixel in the *cell* regions are computed as explained in Section II. The histogram of the oriented phase of the four *cells* in *block*1 $(HOP_{c1}, \ldots, HOP_{c4})$ are computed for the interval $[0^\circ, 180^\circ]$ with 9-orientation binning size as illustrated in Fig. 3 (a). For each pixel in the *cell* region, the phase congruency magnitude is contributing to the *cell* region histogram by adding its value to the corresponding orientation bin. The histograms of the oriented phase for the four *cells* in *block*1 are combined and denoted by HOP_{b1} as given in (18).

$$HOP_{b1} = [\, HOP_{c1} \quad HOP_{c2} \quad HOP_{c3} \quad HOP_{c4}] \quad (18)$$

The histogram of the center symmetric local binary pattern values of *block*1 $(CSLBP_{b1})$ is computed. The fused structural and texture features of the *block*1 (FST_{b1}) can be obtained by the combination of HOP_{b1} with $CSLBP_{b1}$ and given by:

$$FST_{b1} = [\, HOP_{b1} \quad CSLBP_{b1}] \quad (19)$$

The same has done for *block*2 to compute FST_{b2} as in (20).

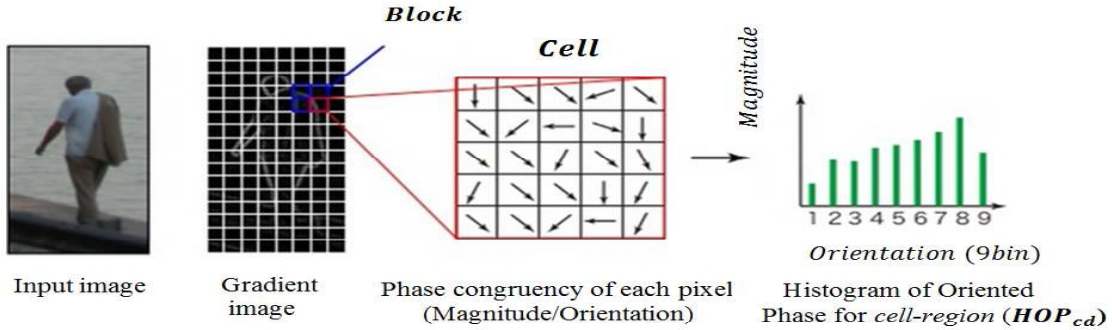
$$FST_{b2} = [\, HOP_{b2} \quad CSLBP_{b2}] \quad (20)$$

Therefore, the overall FST features of the input image can be computed by the concatenation of all *block's* FST histograms as shown in Fig. 3 (b) and the FST feature vector is given by:

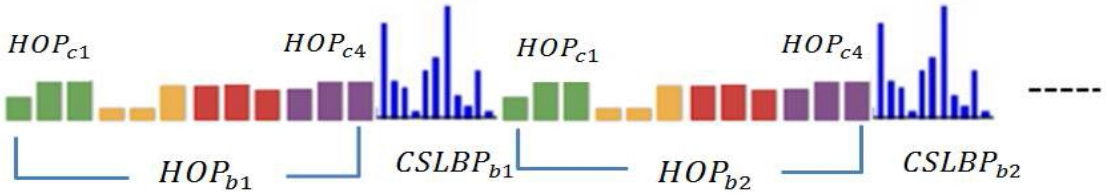
$$FST = [\, FST_{b1} \quad FST_{b2} \quad FST_{b3} \ldots\ldots \quad FST_{bd}] \quad (21)$$

where, d is the number of *block* regions.

The fused structural and texture features of each *block* region FST_{bd} and the overall FST feature vector should be normalized for better tolerance against the shadowing and illumination effects. Several methods can be used for normalization such as (L1-norm, L2-norm, L1-sqrt, and L2-Hys). The optimal choice of these methods will be explained later in the experimental results.



(a)



(b)

Fig. 3 (a) Extraction of the Histogram of Oriented Phase for the *cell* local region (HOP_{cd}), (b) Concatenation of *FST* features

V. Experimental Results and Analysis

Pedestrian detection systems based on the proposed FST descriptor are tested and evaluated using INRIA and DaimlerChrysler datasets. The linear Support Vector Machine (SVM) from the VLFeat toolbox is used to train the pedestrian classifier. The FST based system is evaluated in comparison with the feature algorithms, HOG [16], the histogram of oriented phase (HOP), CSLBP, FHOG (Felzenszwalb-HOG), and the multi feature HOG+CSLBP. To evaluate the detector performance, we plot the curve of the miss-rate versus the false positive rate per window ($FPPW$) in the semi-log scale, where the miss rate and FPPW is given as:

$$\text{miss rate} = \frac{\text{False Negative}}{\text{True Positive+False Negative}} \quad (22)$$

$$\text{FPPW} = \frac{\text{False Positive}}{\text{Total Negative windows}} \quad (23)$$

Better detection performance requires a lower miss-rate and higher detection rate at the same FPPW.

A. Experiment-1: Using INRIA Dataset

INRIA dataset [16] consists of 2416 positive (person) samples and 12180 negative (no-person) windows randomly sampled in the size 128×64 pixels from 1218 negative images and used for training the classifier of the detection system. 1126 positive samples in the size 128×64 pixels and 453 person free images are used in the testing phase. This dataset is the most commonly used dataset for the evaluation of the human detection systems. It is comprised of challenging samples with different occlusion, pose, clothing, and illumination. Some of the pedestrian and non-pedestrian samples from INRIA dataset are shown in Fig. 4.

Fig. 4 Samples of the pedestrian and non-pedestrian images from INRIA dataset

In the first part of *experiment-1*, the effects of various orientation binning sizes (6-bin, 9-bin, and 18-bin) of the FST descriptor on the detection performance are analyzed. In addition, various normalization methods (L1-Norm, L2-Norm, L1-sqrt and L2-Hys) have applied in order for the FST descriptor to analyze their effects on the performance of the detection system. The results illustrated in Fig. 5 show that the best performance is obtained at the binning size of 18-bin. The dimension size of the FST feature vector at the binning sizes (6-bin, 9-bin, and 18-bin) is 4200, 5460, and 9240 elements respectively. At FPPW=10^{-4} the miss rates of the detection system based on FST descriptor at the orientation binning size (6-bin, 9-bin, and 18-bin) is 8.1%, 5.8%, and 5.4% respectively. Since the dimensionality of the FST descriptor at the binning size 9-bin is shorter than that in 18-bin, and their miss rates are close to each other, the descriptor with the orientation binning size 9-bin is selected in the next evaluations of the pedestrian detection system. The influence of the normalization process on the detection performance is illustrated in Fig. 6. The results show that the normalization method has a significant effect on the detection performance, and the best result can be achieved when L2-hys is applied.

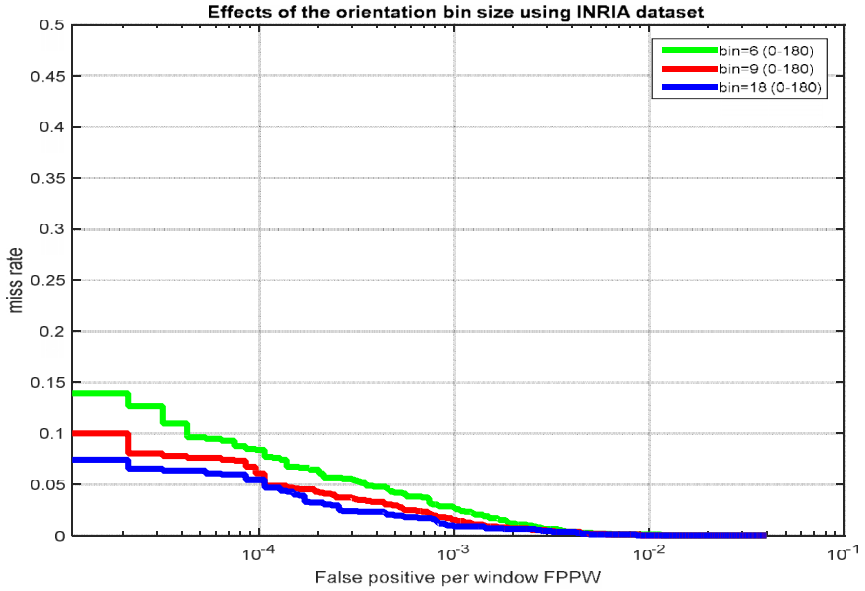


Fig. 5 The detection performance at a different orientation binning size (INRIA dataset are used)

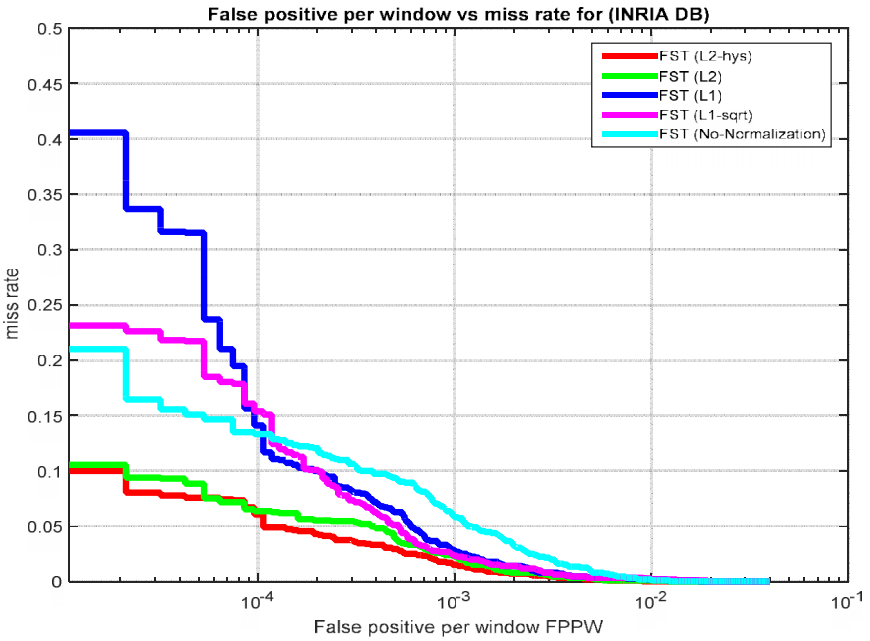


Fig. 6 The detection performance at a various normalization method

In the second part of this experiment the evaluation of the pedestrian detection system, which is based on the proposed descriptor using the INRIA dataset, is performed. Its detection performance results are shown in Fig. 7. The miss rates and the detection rates of the detection system that is based on the FST descriptor, along with and its comparison with several other systems based on various algorithms at $FPPW = 10^{-4}$, are illustrated in Table I. These results show that the proposed FST descriptor has the lowest miss rate (5.8%) in comparison with HOG, CSLBP, HOP, FHOG (Felzenszwalb-HOG), HOG+CSLBP algorithms and has the highest detection rate as well (94.2%).

TABLE I
DETECTION PERFORMANCE AT $FPPW = 10^{-4}$ FOR INRIA DATASET

Algorithm	Miss Rate	Detection Rate
HOG (Dalal &Triggs)	33.4 %	66.52 %
CSLBP	30.8 %	69.2 %
FHOG	24.02%	77.41%
HOP	20.54%	79.64%
HOG+CSLBP	11.1%	88.9%
FST (proposed)	**5.8 %**	**94.2%**

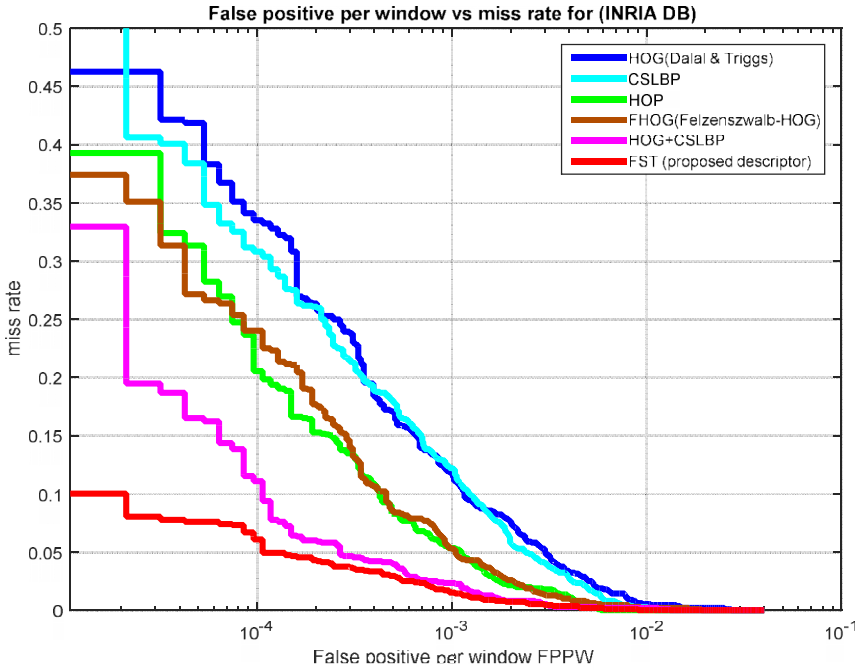


Fig. 7 Detection performance of the FST detector based on INRIA dataset and its comparison with HOG, CSLBP, HOP, FHOG, and HOG+CSLBP detectors

B. Experiment-2: Using DaimlerChrysler Dataset

In this experiment, the DaimlerChrysler dataset [14] is used to evaluate the human detection system based on FST descriptor. This dataset includes a collection of low resolution gray level pedestrian and non-pedestrian images in the size 18×36 pixels. It is composed of five disconnected sets, three of them ("1", "2", and "3") are assigned for training phase and two ("T1", "T2") are used for testing. Each of these sets consists of (4800 pedestrian, 5000 non-pedestrian). The images of this dataset is difficult to be classified due to the small size of the samples and the challenges of the negative sets [15]. Some of the pedestrian and non-pedestrian samples of the DaimlerChrysler dataset are shown in Fig. 8.

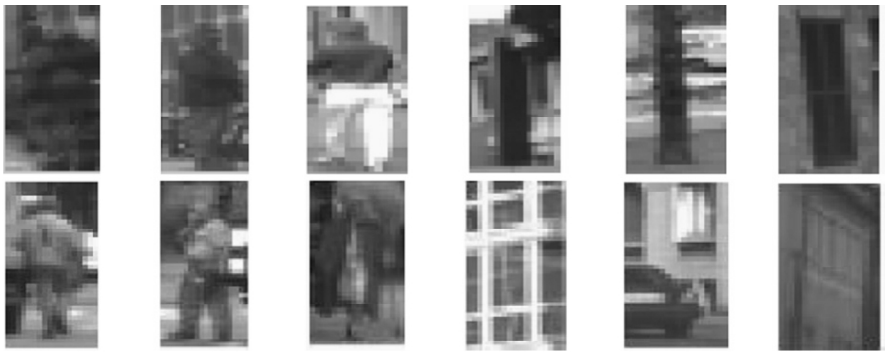


Fig. 8 Samples of the pedestrian and non-pedestrian images from DaimlerChrysler dataset

Experiment-2 is performed two times using different composition of the DaimlerChrysler dataset. Fig. 9 shows the performance of the detection system based on the FST descriptor when a set "1" is used for training the classifier and "T2" is used in the testing phase. Fig. 10 shows the performance when set "3" is used for training the classifier and "T2" is used for the testing. At $FPPW = 10^{-3}$, the miss rates of the detection system based on the FST descriptor for both cases in comparison with the various algorithms are illustrated in Table II. The miss rates of the FST based system is 32.7% for the used datasets ("1" &"T2") and 26.83% for the used datasets ("3" &"T2"). These results are the lowest rates in comparison with the miss rates of the HOG, CSLBP, HOP, FHOG (Felzenszwalb-HOG), and HOG+CSLBP algorithms. These results prove that the proposed descriptor has better detection performance over the mentioned algorithms.

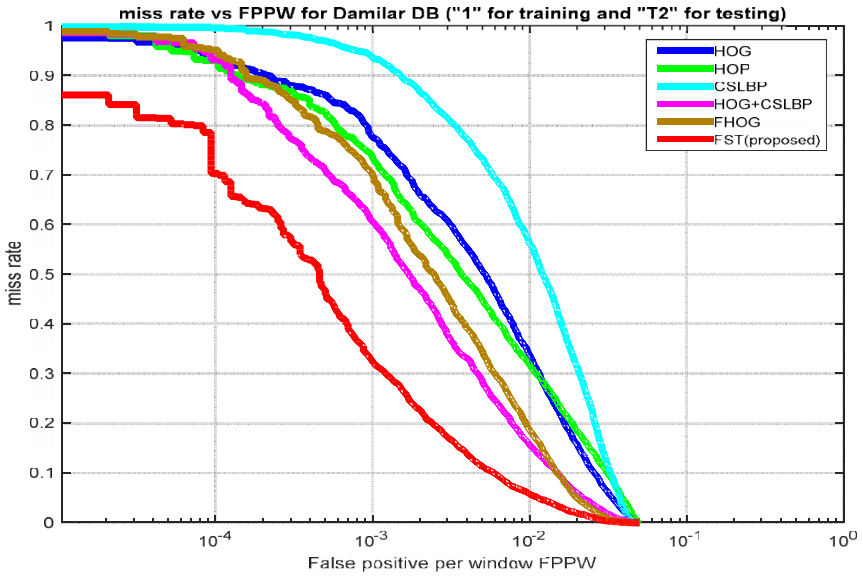


Fig. 9 Detection performance of FST detector on the DaimlerChrysler dataset and its comparison with various algorithms when (Dataset "1" used for training & dataset "T1" for testing)

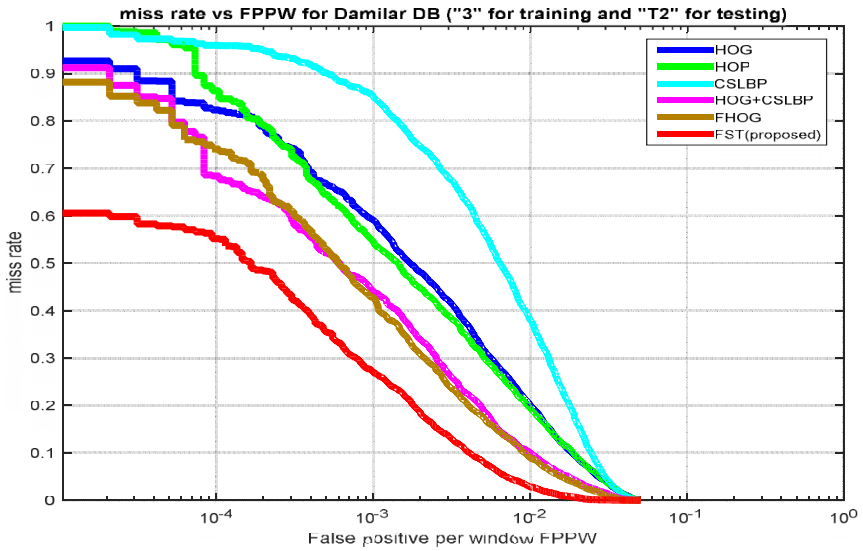


Fig. 10 Detection performance of FST detector on DaimlerChrysler dataset and its comparison with various algorithms when (Dataset "3" used for training & dataset "T2" for testing)

TABLE II
DETECTION PERFORMANCE AT $FPPW = 10^{-3}$ FOR DAIMLER CHRYSLER DATASET

Algorithm	Using set "1" & T2 Miss Rate	Using set "3" & T2 Miss Rate
HOG (Dalal &Triggs)	77.4 %	58.5 %
CSLBP	93.6 %	84.6 %
HOP	72.5%	54.4 %
FHOG	69.3%	40.9 %
HOG+CSLBP	60.5%	43.7 %
FST (proposed)	**32.7 %**	**26.8 %**

V. CONCLUSION

The Fused Structure and Texture features (FST) is a descriptor developed based on the phase congruency concept and the Center Symmetric Local Binary Pattern (CSLBP) operator, and it is intended to improve the pedestrian detection system. The novelty of the proposed descriptor is the fusing of the local phase information with textures and the image gradient. By combining these operators more significant information of the upright human body could be extracted in order to permit the detection of pedestrians more efficiently with less sensitivity to light variations. The experimental evaluation results performed on the INRIA and DaimlerChrysler datasets showed that the detection performance of the pedestrian detection system based on the proposed descriptor is better than that based on HOG, CSLBP, HOP, FHOG, HOG+CSLBP features.

REFERENCES

[1] Li Zhang Bo Wu and Ram Nevatia, Pedestrian Detection in Infrared Images Based on Local Shape Features, In CVPR, June 2007.
[2] X. Wang, T. X. Han and S. Yan, An HOG-LBP Human Detector with Partial Occlusion Handling, In ICCV, pp. 32-39, Kyoto, 2009.
[3] Wojek, C., Schiele, B.: A performance evaluation of single and multi-feature people detection. Proceedings of DAGM Symposium on Pattern Recognition" (2008) 82–91
[4] Oppenheim and J. S. Lim, The importance of phase in signals, Proceedings of the IEEE, vol. 69, no. 5, pp. 529-541, 1981.
[5] P. Kovesi, Image Feature from Phase Congruency, Robotics and Vision Research Group. Technical Report 95/4, March 1995.
[6] P. Kovesi, Phase Congruency Detects Corners and Edges, Proceedings DICTA 2003, Sydney Dec 10-12.
[7] M. Heikkila and C. Schmid, Description of interest regions with local binary patterns, Pattern Recognition., vol. 42, no. 3, pp. 425–436, 2009.
[8] V. Santhaseelan · V. Asari, Utilizing Local Phase Information to Remove Rain from Video, Int J Comput Vis, DOI 10.1007/s11263-014-0759-8, August 2014.
[9] P. Kovesi, Image Feature from Phase Congruency Robotics and Vision Research Group. Technical Report 95/4, March 1995.
[10] C.Yao Su, J. Yang, Histogram of gradient phases: a new local descriptor for face recognition, Published in IET Computer Vision, 2014.
[11] X. Yuan, P. Shi, Iris Feature Extraction Using 2D Phase Congruency, Institute of Image Processing and Pattern Recognition, China, 200030.
[12] Y. Zheng, C. Shen, R. Hartley, X. Huang, Effective Pedestrian Detection Using Center-symmetric Local Binary/Trinary Patterns, IEEE, September 2010
[13] M. Heikkilä, M. Pietikäinen, C. Schmid, Description of Interest Regions with Center-Symmetric Local Binary Patterns, ICVGIP 2006: 58-69.
[14] S. Munder, D. M. Gavrila, An Experimental Study on Pedestrian Classification, IEEE Trans. on Pattern Analysis and Machine Intelligence, 2006.
[15] L. Nanni, S. Brahnam, A. Lumini, A simple method for improving local binary patterns by considering non-uniform patterns, Pattern Recognition 45 (2012) 3844–3852.
[16] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In IEEE Conf. Computer Vision and Pattern Recognition (CVPR), volume 1, pages 886–893, 2005.
[17] S. Venkatesh, and R. Owens, (1989). An energy feature detectionscheme. In IEEE International Conference on Image Processing: Conference Proceedings ICIP'89, Sep 5–8 1989, Singapore: IEEE.
[18] M. Morrone and R. Owens. Feature detection from local energy. Pattern Recognition Letters, 6:303–313, 1987.