2-2016

Histogram of Oriented Phase and Gradient (HOPG) Descriptor for Improved Pedestrian Detection

Hussin Ragb
University of Dayton

Vijayan K. Asari
University of Dayton, vasari1@udayton.edu

Histogram of Oriented Phase and Gradient (HOPG) Descriptor for Improved Pedestrian Detection

Hussin K. Ragb and Vijayan K. Asari; Vision lab, University of Dayton; Dayton, Ohio/USA

Abstract

This paper presents a new pedestrian detection descriptor named Histogram of Oriented Phase and Gradient (HOPG) based on a combination of the Histogram of Oriented Phase (HOP) features and the Histogram of Oriented Gradient features (HOG). The proposed descriptor extracts the image information using both the gradient and phase congruency concepts. Although the HOG based method has been widely used in the human detection systems, it lacks to deal effectively with the images impacted by the illumination variations and cluttered background. By fusing HOP and HOG features, more structural information can be identified and localized in order to obtain more robust and less sensitive descriptors to lighting variations. The phase congruency information and the gradient of each pixel in the image are extracted with respect to its neighborhood. Histograms of the phase congruency and the gradients of the local segments in the image are computed with respect to its orientations. These histograms are concatenated to construct the HOPG descriptor. The performance evaluation of the proposed descriptor was performed using INRIA and DaimlerChrysler datasets. A linear support vector machine (SVM) classifier is used to train the pedestrians. The experimental results show that the human detection system based on the proposed features has less error rates and better detection performance over a set of state of the art feature extraction methodologies.

Introduction

Pedestrian detection is one of the widely used applications in the pattern recognition and computer vision systems. Over the last decade, detection of human beings in a visual surveillance system is a significant task due to its extended applications including human computer interaction, person identification, event detection, counting people in crowded regions, gender classification, automatic navigation, safety systems, etc. The fluctuating appearance of the human body combined with the occlusions, cluttered scenes and illumination changes, make the pedestrian detection application as one of the challenging categories in object detection. This task has been tackled by different techniques. Histogram of Oriented Gradient (HOG) introduced in 2005 by Dalal and Triggs [1], is one of the most popular techniques used for the pedestrian detection features. In this approach, the distribution of the gradient information is used to depict the shape and the appearance of the objects. This can be achieved by scanning the local portions of the image and counting the occurrences of the gradient orientation. In the last decade, this method and other single features algorithms such as Scale Invariant Feature Transform (SIFT) [14], Local Binary Pattern (LBP) [19], Edgelet [20], Haar [21], and Shapelet [22] played vital role in the pedestrian detection systems. However they were lacking to deal efficiently with images impacted by the noise and the illumination variations. Histogram of Oriented Phase (HOP) descriptor based on phase congruency has been developed to eliminate such problems and improve the detection performance. This descriptor can be used to detect and localize the significant point features of the image better than the traditional gradient operators especially in the regions affected by illuminations and where the perceived edge does not have a sharp step or a large derivative. However HOP descriptor is sensitive to noise and it is not efficient to detect objects in such case. Although, the above mentioned single feature algorithms could be used to depict the appearance of the objects reasonably, however they are still not enough to deal efficiently with the complex scenes such as poor image resolution, occlusion, and crowded scenes. The efficiency limitation of the single features based techniques raised up the need of combined multi-features to improve the performance of human detection. In 2007, Zhang and Ram [2] improved the detection of the IR images by the combination of the Edglets and HOG features. Wojek [3] combined HOG, Haar and shapelet features. Wang et al [4] connected HOG and LBP features to improve pedestrian detection performance. Dollar [5] also developed the integral channel features (ICF) that combined HOG, gradient magnitude and LUV color, etc. In this paper we propose a new feature extraction algorithm that combines the image shape features based on the local phase information with the histogram of oriented gradient features denoted as $HOPG$ descriptor. This approach leads an extraction of the complementary information of the image, hence significantly increases the detection rate of the system. The pedestrian detection framework based on $HOPG$ features is shown in Figure 1.

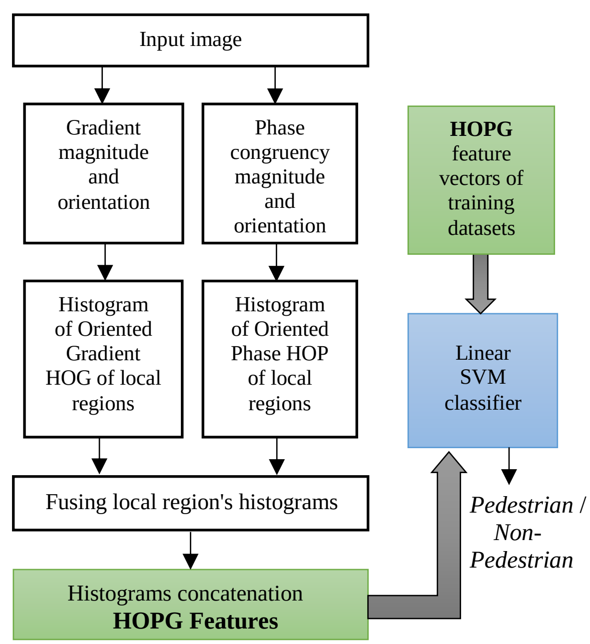

Figure 1. Framework of the pedestrian detection system based on HOPG features

The input image is divided into local regions. The gradients and phase congruency magnitudes and their orientations are computed in each pixel of the input image with respect to its neighborhood. The histogram of oriented gradient and the histogram of the oriented phase are computed in each local region and combined to each other. These histograms are concatenated to form the $HOPG$ descriptor that representing the input sample. Thses features are fed to a support vector machine (SVM) classifier that should be trained first by the $HOPG$ feature vectors of the pedestrian and non-pedestrian training dataset. Based on these input data, the classifier will be able to decide whether the tested image belongs to the positive (pedestrian) or a negative (non-pedestrian) category. The remaining sections of this paper are organized as following:

- Phase congruency concept and computation.
- Image gradient computation.
- Implementation of the Histogram of Oriented Phase and Gradient ($HOPG$) descriptor.
- Experimental results.
- Conclusion

Phase congruency concept and computation

Phase congruency PC is a technique developed by Kovesi [6] to detect features such as corners and edges of 2D digital images. The phase congruency algorithm postulates that features are perceived at points where the Fourier components are maximal in phase [6], [7].

First, consider a 1D signal $I(x)$ defined in the range $[0,2\pi]$. This signal is represented by a sinusoidal in the frequency domain by applying the Fourier transform. Figure 2 shows that all the Fourier components meet in the edges where the PC is maximum. Since the Fourier components meet at the edges, the local energy $E(x)$ would be also maximum at these points. Therefor the phase congruency is directly proportional to the energy [8], [9], [10], [15]. The phase congruency becomes a dimensionless quantity by the normalization with the sum of all Fourier components amplitudes A_n [15], [18].

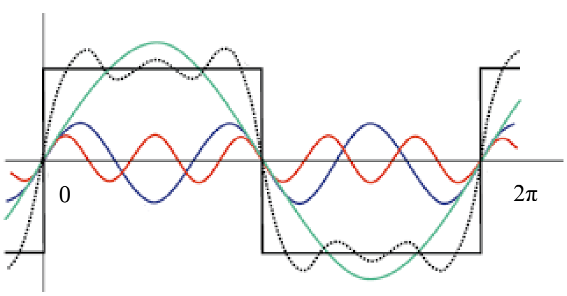

Figure 2. The Fourier components of a step signal.

Hence, the phase congruency PC is given by:

$$PC(x) = \frac{E(x)}{\varepsilon + \sum_n A_n} \tag{1}$$

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

Equation 1 shows that the peaks in the local energy correspond to the peaks in phase congruency [7], [11]. It shows also that the phase congruency does not depend on the signal overall magnitude, making the features invariant to scale, illumination and contrast variations [7].

Assume $F(x)$ is the same as the input signal $I(x)$ filtered from a DC component and $F_H(x)$ is 90° phase shift of $F(x)$ (Hilbert Transform). Using $F(x)$ and Hilbert Transform $F_H(x)$, The local energy function $E(x)$ can be defined as given in Eq. 2 [18].

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

Calculation of phase congruency using Gabor filter

Phase congruency can be computed by convolving the two dimensional signal with a pair of quadrature filters to extract the local frequencies and phase information. Log Gabor filter is an efficient band pass filter used in this paper to extract the local phase information spread over a wide spectrum. The transfer function of the Log-Gabor filter is given by [16]:

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

where ω_o is the center frequency of the filter. k/ω_o is kept constant for various ω_o [8], [12]. The transfer function of 2-D log Gabor filter constructed by using Gaussian function in the angular direction is given by:

$$G(\theta) = exp\left(\frac{-(\theta-\theta_o)^2}{2{\sigma_\theta}^2}\right) \tag{4}$$

where θ_o is the center orientation of the filter, and σ_θ is the standard deviation of the Gaussian function in angular direction [8], [12].

consider M_{no}^o and M_{ne}^e are the odd symmetric and even symmetric components of the Log- Gabor filter at scale n and orientation o so that they form a quadrature pair. The response vector at scale n and orientation o is obtained by the convolution of each quadrature pair with the input signal $I(x,y)$ and is given by:

$$[e_{no}(x,y)\,, o_{no}(x,y)] = [I(x,y) * M_{no}^e\,, I(x,y) * M_{no}^o] \tag{5}$$

The amplitude of the response A_{no} and the phase angle ψ_{no} at scale n and orientation o are given by:

$$\boldsymbol{A_{no}} = \sqrt{(\boldsymbol{e_{no}^2}\,(\boldsymbol{x},\boldsymbol{y}) + \boldsymbol{o_{no}^2}\,(\boldsymbol{x},\boldsymbol{y}))} \tag{6}$$

$$\psi_{no}(x,y) = tan^{-1}\left(\frac{o_n(x,y)}{e_n(x,y)}\right) \tag{7}$$

Referring to equation (2), $F(x,y)$ and $F_H(x,y)$ for a 2D signal are given by:

$$F(x,y) = \sum_n e_{no}(x,y) \tag{8}$$

$$F_H(x,y) = \sum_n o_{no}(x,y) \tag{9}$$

Hence, phase congruency $P(x,y)$ of the 2D signal can be computed over various scales and orientation as:

$$PC(x,y) = \frac{\sum_o \sqrt{(\sum_n e_{no}(x,y))^2 + (\sum_n o_{no}(x,y))^2}}{\varepsilon + \sum_o \sum_n A_{no}(x,y)} \quad (10)$$

The orientation angle $\varphi(x,y)$ is given by

$$\varphi(x,y) = tan^{-1}\left(\frac{F_H(x,y)}{F(x,y)}\right) \quad (11)$$

The level of phase congruency lies between the values 0 and 1. Figure 3 illustrates an example of phase congruency for an image at various illumination and contrast levels. We can note that the phase congruency is the same and invariant with illumination changes.

Figure 3. Phase congruency of the image at various illumination levels

Image Gradient computation

Gradient of the image is defined as the directional change of the image intensity or color. Simply it can be obtained horizontally $G_x(x,y)$ by convolving the image with the mask templet $(-1\ 0\ 1)$ and vertically $G_y(x,y)$ by convolving the image with $(-1\ 0\ 1)^T$. The gradient magnitude $G(x,y)$ and the orientation $\phi(x,y)$ for the image $I(x,y)$ can be calculated as following [17];

$$G_x(x,y) = I(x+1\,,y) - I(x-1\,,y) \quad (12)$$

$$G_y(x,y) = I(x\,,y+1) - I(x\,,y-1) \quad (13)$$

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \quad (14)$$

$$\phi(x,y) = \tan^{-1}\left(\frac{G_x(x,y)}{G_y(x,y)}\right) \quad (15)$$

Implementation of Histogram of Oriented Phase and Gradient (HOPG) descriptor

$HOPG$ descriptor proposed in this paper can be implemented by dividing the image into connected local regions called *blocks* of size 16×16 pixels with no *block* overlap. Each *block* is built from 4×4 *cells* with a cell size of 4×4 pixels as shown in Figure 4. The phase congruency $PC(x,y)$ value, the gradient magnitude $G(x,y)$ and their orientation angles $\varphi(x,y)$ and $\phi(x,y)$for each pixel in the *cells* regions are computed (as explained above). Each *cell* is discretized into angular bins according to the orientation interval. The histogram of oriented phase $(HOP_{c1},\dots,HOP_{c16})$ and histogram of oriented gradient $(HOG_{c1},\dots,HOG_{c16})$ of the 16-*cells* in *block*1 are computed for the interval [0˚, 180˚] with 9-orientation binning size. The corresponding orientation bin is found and the orientation's phase congruency value and gradient magnitude are voted to this bin. Histograms of the oriented phase HOP_{b1} and histogram of oriented gradient $HOG_{b1}1$ for *block*1 are given as:

$$HOP_{b1} = [HOP_{c1}\quad HOP_{c2}\ \dots HOP_{c16}] \quad (16)$$
$$HOG_{b1} = [HOG_{c1}\quad HOG_{c2}\ \dots HOG_{c16}] \quad (17)$$

HOP_{b1} and HOG_{b1} are combined to construct the histogram of oriented phase and gradient for *block*1 and denoted by $HOPG_{b1}$ as given in Eq. 18.

$$HOPG_{b1} = [HOP_{b1}\quad HOG_{b1}] \quad (18)$$

Similarly, $HOPG_{b2}$ for *block*2 is given by:

$$HOPG_{b2} = [HOP_{b2}\quad HOG_{b2}] \quad (19)$$

The same are done for the rest *blocks* of the image. The overall histogram of oriented phase and gradient for the entire image are constructed by the concatenation of the whole histograms extracted from the *block* regions in one feature vector representing the $HOPG$ descriptor where,

$$HOPG = [HOPG_{b1}\quad HOPG_{b2}\quad HOPG_{b3}\ \dots\dots\dots\ HOPG_{bd}] \quad (20)$$

where d is the number of *block*s in the image (d=32 for the image of size 64×128 pixels).

The feature vectors of the *block* regions should be normalized for better illumination and shadowing invariances. The resultant descriptor $HOPG$ should also be normalized with a convenient normalization method such as (L1-norm, L2-norm, L1-sqrt, and L2-Hys) as shown later in the experimental results.

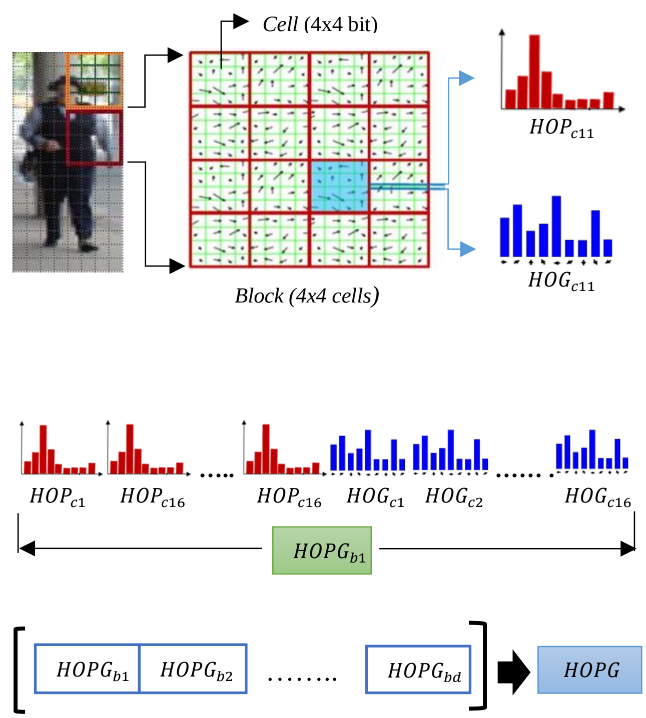

Figure 4 Construction of Histogram of Oriented Phase and Gradient (HOPG) descriptor

Experimental Results

A pedestrian detection system based on the proposed $HOPG$ descriptor are tested and evaluated using two well-known pedestrian datasets; INRIA and DaimlerChrysler datasets. The linear Support Vector Machine (SVM) from VLFeat toolbox is used as classifier of this detection system. $HOPG$ descriptor based detection system is evaluated in comparison with the feature extraction algorithms; HOG (Dalal &Triggs) , Histogram of Oriented Phase (HOP), Felzenszwalb-HOG (FHOG), Center Symmetric Local Binary Pattern (CSLBP), and the multi-features HOG+CSLBP. In the detection performance evaluation, we plot the curve of the miss rate versus the false rate per window (FPPW) in semi-log scale. Better detection performance requires lower miss-rate and higher detection rate at the same FPPW. The miss rate and the FPPW are given by;

$$miss\ rate\ = \frac{False\ Negative}{True\ Positive + False\ Negative} \tag{21}$$

$$FPPW = \frac{False\ Positive}{Total\ Negative\ windows} \tag{22}$$

Experiment-1: Evaluation using INRIA dataset

For the INRIA dataset [1], 2416 positive (person) samples and 9750 negative (no-person) samples are cropped in the size 128×64 pixels and used for training the detection system.1126 positive samples and 3750 negative samples are cropped in the size 128×64 pixels and are used for testing. Samples of the pedestrians and the non-pedestrians from INRIA data set are shown in Figure 5.

Figure 5 Samples of the pedestrian and non-pedestrian images from INRIA dataset

In the first part of this experiment, we have analyzed the effects of the descriptor normalization methods (L1-Norm, L1-sqrt, L2-Norm and L2-Hys) and the orientation binning size (6-bin, 9-bin, 18-bin) on the detection performance. Figure 6 shows that the normalization method has a big influence and gives the best results when L2-norm has applied. Figure 7 illustrates that the optimal performance can be achieved at the orientation binning size of 9-bin.

In the next part of this experiment, the INRIA pedestrian dataset is used. The detection performance of the system based on the proposed $HOPG$ descriptor is shown in Figure 8. At FPPW=10^{-4}, the detection rates for the HOG, CSLBP, HOP, FHOG (Felzenszwalb-HOG), and HOG+CSLBP based detectors are (70%), (70%), (79.6%), (84.2%), and (91.07%) respectively. However the detection rate of the HOPG based detector is (94.92%). The miss rates of the HOG, CSLBP, HOP, FHOG (Felzenszwalb-HOG), and HOG+CSLBP detectors at FPPW=10^{-4} are (30%), (30%), (20.4%), (15.8%), and (8.93%) respectively. The miss rate for the proposed descriptor is (5.089%). From the results illustrated in Table 1, we can observe that a significant improvement has been achieved by using the proposed $HOPG$ descriptor.

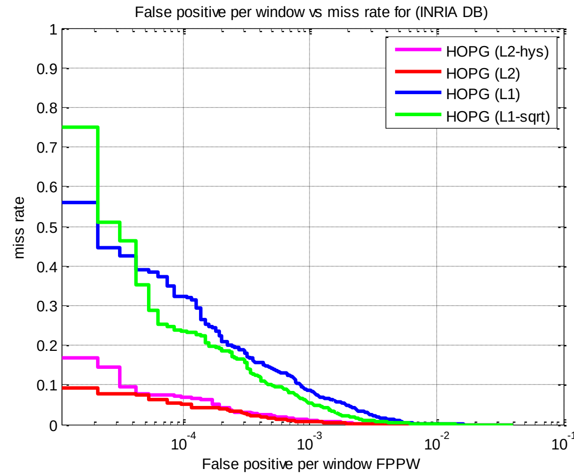

Figure 6 The detection performance at a different normalization method

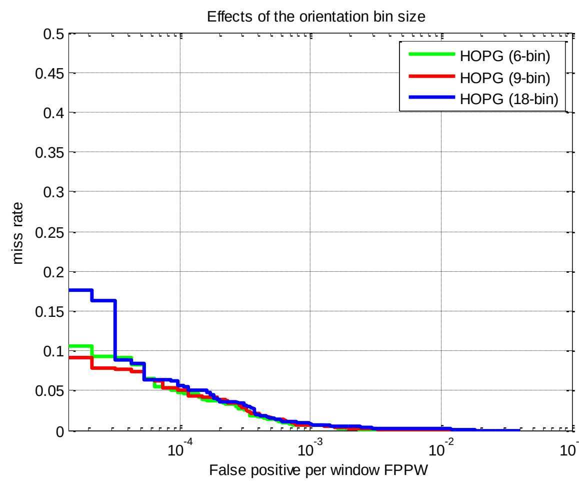

Figure 7 The detection performance at a different orientation binning size

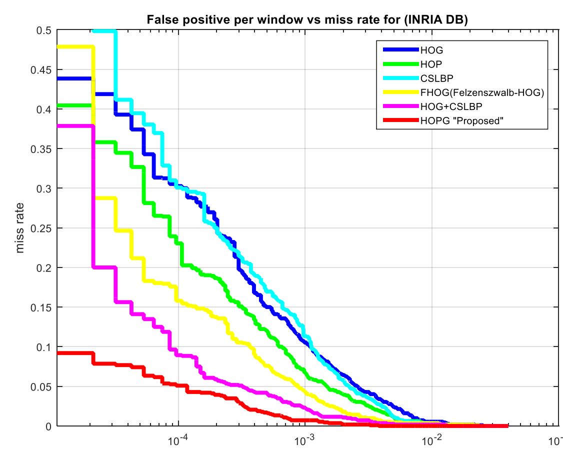

Figure 8 Detection performance of HOPG detector on INRIA dataset and its comparison with HOG, HOP, CSLBP, FHOG, and HOG+CSLBP based detectors

Table 1: Detection performance at $FPPW$ = 10^{-4}, INRIA Dataset

Algorithm	Miss rate	Detection Rate
HOG	30%	70%
CSLBP	30%	70%
HOP	20.45%	79.6%
FHOG	15.8%	84.3%
HOG+CSLBP	8.93%	91.07%
HOPG "proposed"	**5.089%**	**94.92%**

Experiment-2: Evaluation using DaimlerChrysler dataset

In this experiment, the pedestrian detection system based on $HOPG$ descriptor is evaluated using DaimlerChrysler dataset [13]. This dataset is composed of five low resolution gray level disconnected sets. Three of them ("1", "2", and "3") comprise a collection of pedestrian and no-pedestrian samples of size 18×36 pixels. These sets are resized to scale 20×40 pixels and are used for training the classifier. DaimlerChrysler dataset include two sets ("T1", "T2") of the pedestrian and no-pedestrian samples used in the testing phase of the pedestrian detection system. Each of these sets consists of 4800 pedestrian and 5000 non-pedestrian. Samples of the pedestrian and non-pedestrian images of DaimlerChrysler dataset are shown in Figure 9.

Figure 9 Pedestrian & non-pedestrian samples from DaimlerChrysler dataset

This experiment is performed two times with different training and testing datasets. Figure 10 shows the detection performance when set "1" is used for training and "T2" for testing. Figure 11 shows the performance when the set "3" is used for training and "T2" for testing. At $FPPW$ $=10^{-3}$ the miss rates of the $HOGP$ based detection system is 37.8% for the datasets ("1" &"T2") and 37.4% for the datasets ("3" &"T2"). These results are the lowest values in comparison with the miss rates of the HOG, CSLBP, HOP, FHOG (Felzenszwalb-HOG), and HOG+CSLBP algorithms as illustrated in Table 2. It is obvious from these experiments that the proposed HOPG descriptor has better detection performance over the other descriptors.

Table 1: Detection performance at $FPPW$ = 10^{-3} for DaimlerChrysler Dataset

Algorithm	Miss rate sets "1" & "T2"	Miss rate sets "3" & "T2"
HOG	85%	65.5%
CSLBP	93.8%	84.6%
HOP	77.9%	58.7%
FHOG	70.1%	41.2%
HOG+CSLBP	60.5%	44.1%
HOPG "proposed	**37. 8%**	**37.4%**

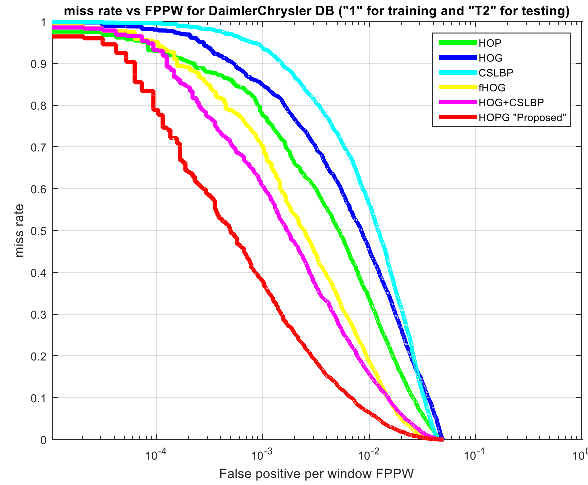

Figure 10 Detection performance of HOPG detector on Daimler dataset and its comparison with HOG, HOP, CSLBP, FHOG, and HOG+CSLBP based detectors (Dataset "1" for training & dataset "T2" for testing)

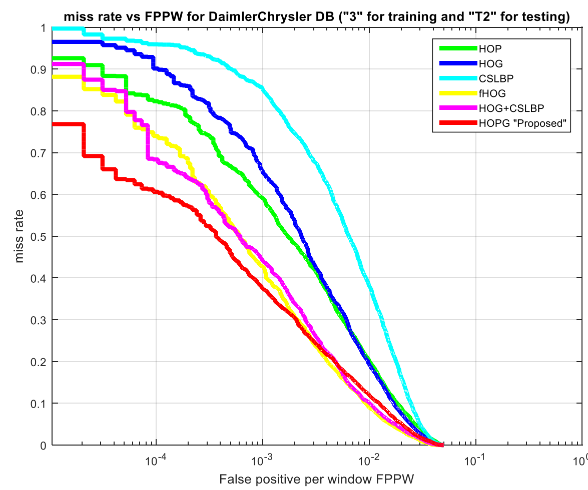

Figure 11 Detection performance of HOPG detector on Daimler dataset and its comparison with HOG, HOP, CSLBP, FHOG, and HOG+CSLBP based detectors (Dataset "3" for training & dataset "T2" for testing)

Conclusion

Histogram of Oriented Phase and Gradient ($HOPG$) is a new descriptor introduced in this paper for improved pedestrian detection system. The proposed algorithm combines the local phase based features with the image gradients in one descriptor. This approach lead $HOPG$ descriptor to capture complementary information of the image, hence significantly increase the detection rate of the system. The experimental evaluations performed on INRIA and DaimlerChrysler datasets showed that $HOPG$ descriptor has better detection performance than the single feature methods such as HOG, HOP, CSLBP, Felzenszwalb-HOG (FHOG) and the multi- feature method HOG+CSLBP.

References

[1] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," In IEEE Conf. Computer Vision and Pattern Recognition (CVPR), vol. 1, pp 886–893, 2005.

[2] Li Zhang，Bo Wu and Ram Nevatia, "Pedestrian Detection in Infrared Images Based on Local Shape Features," In CVPR, June 2007.

[3] C. Wojek, , B. Schiele, "A performance evaluation of single and multi-feature people detection," Proceedings of DAGM Symposium on Pattern Recognition (2008) 82-91.

[4] X. Wang, Pietikain, "partial occlusion handling," proceedings of IEEE International Conference on Computer Vision (2009) 2-29

[5] P. Dollar, Z. Tu, P. Perona, S. Belongie, "Integral channel features," Proceedingsof British Machine Vision Conference (2009) 1-11.

[6] P. Kovesi, "Phase Congruency Detects Corners and Edges," Proceedings DICTA 2003, Sydney Dec 10-12.

[7] P. Kovesi, "Image Feature from Phase Congruency," Robotics and Vision Research Group, Technical Report 95/4, March 1995.

[8] G. Basavaraj, G. Reddy, "An Improved Face Recognition Using Neighborhood Defined Modular Phase Congruency Based Kernel PCA," (IJERA) ISSN: 2248-9622 Vol. 2, Issue 2, Mar-Apr 2012, pp.528-535.

[9] M. Morrone, D. Burr, "Feature detection in human vision: A phase-dependent energy model," in Proc. Roy. Soc., London B, 1988, vol. 235, pp. 221–245.

[10] M. Morrone and R.Owens, "Feature detection from local energy," Pattern Recognit. Lett. , vol. 6, pp. 303–313, 1987.

[11] V. Santhaseelan and V. Asari, "Utilizing Local Phase Information to Remove Rain from Video," Int J Comput Vis, DOI 10.1007/s11263-014-0759-8, August 2014.

[12] T. Nguyen, S. Kim, I. Seop, "Fast Pedestrian Detection Using Histogram of Oriented Gradients and Principal Components Analysis," International Journal of Contents, Vol.9, No.3, Sep 2013.

[13] S. Munder, D. Gavrila, "An Experimental Study on Pedestrian Classification," IEEE Trans. on Pattern Analysis and Machine Intelligence, 2006.

[14] D. Lowe, “Distinctive image features from scaleinvariant keypoints,” International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.

[15] S. Venkatesh and Owens, " An energy feature detection scheme," In IEEE International Conference on Image Processing: Conference Proceedings ICIP’89, Sep 5–8 1989, Singapore: IEEE.

[16] X. Yuan, P. Shi," Iris Feature Extraction Using 2D Phase Congruency Proceedings of the Third International Conference on Information Technology and Applications (ICITA’05), 0-7695-2316-1/05.

[17] S. Yang, X. Liao, " A Pedestrian Detection Method Based on the HOG-LBP Feature and Gentle Ada-Boost", International Journal of Advancements in Computing Technology(IJACT) Volume4, Number19,October. 2012.

[18] S. Gundimada,V. Asari," A Novel Neighborhood Defined Feature Selection on Phase Congruency Images for Recognition of Faces with Extreme Variations", International Journal of Information Technology Volume 3 Number 1.

[19] 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, Jul. 2002.

[20] B. Wu, R. Nevatia, "Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors," Proceedings of IEEE International Conference on Computer Vision 1 (2005) 90-97.

[21] P. Viola, M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition 1 (2001) 511-518.

[22] P. Sabzmeydani, G. Mori, "Detecting pedestrians by learning shapelet features," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2007) 1-8.

Author Biography

Hussin K. Ragb received his BS in Electrical and Electronic Engineering from the University of Tripoli (1991). Since then he has worked in the optical research center in, Tripoli-Libya. His work has focused on the development of the laser based systems. He has received his MS in Electrical Engineering from the University of Belgrade (2000). Since then he worked as lecturer at University of Tripoli. Currently he is PhD student in Vision Lab in the University of Dayton-USA.

Vijayan K. Asari is a Professor and Endowed Chair in electrical and computer engineering at the University of Dayton, USA. He is the Director of the Vision Lab at UD. Dr. Asari received his PhD in Electrical Engineering from the Indian Institute of Technology, Madras. He holds three patents and has published more than 500 papers, including 83 journal papers. Dr. Asari is a Member of IS&T and a Senior Member of IEEE and SPIE.