5-2016

Histogram of Oriented Phase (HOP): A New Descriptor Based on Phase Congruency

Hussin Ragb
University of Dayton

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

Histogram of oriented phase (HOP): A new descriptor based on phase congruency

Hussin K. Ragb*[a], Vijayan K. Asari[a]
[a]Vision Lab., 300 College Park, Dayton, OH, USA 45469-0232

ABSTRACT

In this paper we present a low level image descriptor called Histogram of Oriented Phase based on phase congruency concept and the Principal Component Analysis (PCA). Since the phase of the signal conveys more information regarding signal structure than the magnitude, the proposed descriptor can precisely identify and localize image features over the gradient based techniques, especially in the regions affected by illumination changes. The proposed features can be formed by extracting the phase congruency information for each pixel in the image with respect to its neighborhood. Histograms of the phase congruency values of the local regions in the image are computed with respect to its orientation. These histograms are concatenated to construct the Histogram of Oriented Phase (HOP) features. The dimensionality of HOP features is reduced using PCA algorithm to form HOP-PCA descriptor. The dimensionless quantity of the phase congruency leads the HOP-PCA descriptor to be more robust to the image scale variations as well as contrast and illumination changes. Several experiments were performed using INRIA and DaimlerChrysler datasets to evaluate the performance of the HOP-PCA descriptor. The experimental results show that the proposed descriptor has better detection performance and less error rates than a set of the state of the art feature extraction methodologies.

Keywords: Phase congruency, HOP, HOG, INRIA dataset, SVM classifier

1. INTRODUCTION

Human detection is considered one of the significant tasks in pattern recognition and computer vision due to its extended applications that include human computer interaction, autonomous navigation, visual surveillance, person identification, event detection, robotics, automotive safety systems, etc. The fluctuating appearance of the human body such as the pose, clothes and scale variations in addition to the occlusion, cluttered scenes, and illumination changes, make the human detection one of the challenging tasks in object detection. This task is tackled with different and diverse techniques. One of these techniques is the Holistic approach which is a training method proposed to classify and identify the full-human body as a single detection region. Part-based method is another approach proposed to identify each part of the body separately, and the human can be detected if some or all these parts are presented in a reasonable spatial configuration [1]. Various types of features are used to depict the human body, ranging from the low level features to complex visual descriptors. These features are used in combination with different types of classifiers [1], [2], [3]. One of the earliest algorithms used for the pedestrian detection system is proposed by Papageorgiou et al. [4]. This technique used sliding window detector and applied the Support Vector Machine (SVM) with the multi-scale Haar wavelet. Viola and Jones [VJ] [5] has up-graded this idea and introduced the integral images for fast features computations and a cascade structure for efficient detection. In addition they developed a detector that integrates Haar-like features with Adaboost classifier. These contributions continue today to serve as a foundation of the modern detection techniques. An important gain in the detection performance came with the adoption of the gradient based features. Inspired by Scale Invariant Features Transform (SIFT), Dalal and Triggs [6] introduced the Histogram of Oriented Gradient (HOG) algorithm. This algorithm is the most popular gradient based technique. It provides a high efficient and robustness features and showing substantial gains over the intensity based features [2]. HOG features is used as a base of the most modern complex descriptors of the human detection system. Wu and Nevatia [7] combined HOG, covariance and edgelets features into one descriptor to improve the detection perormance. Wang et al [8] combined HOG features with texture features. Wojek and Schiele [9] combined Haar-like features, shapelets, shape context and HOG features. Zhang and Ram [10] have combined the Edgelets with the HOG features and they improved the detection performance of IR images. Dollar [11] developed the integral channel features (ICF) that combined HOG, gradient magnitude and LUV color. Although HOG features played a vital role in human detection systems, it was lacking to deal effectively with images impacted by noise, contrast, and illumination variations [12], [13]. In this paper we present a descriptor based on

Mobile Multimedia/Image Processing, Security, and Applications 2016, edited by Sos S. Agaian,
Sabah A. Jassim, Proc. of SPIE Vol. 9869, 98690V · © 2016 SPIE
CCC code: 0277-786X/16/$18 · doi: 10.1117/12.2225159

the phase congruency concept, called Histogram of Oriented Phase (HOP). This descriptor can be used to depict and represent the human objects more efficiently than the gradient based approach especially those images exposed to the illumination and contrast variations. The significance of the phase information proved by the experiment of Oppenheim and Lim [12] played an important motivation factor for using the phase congruency as a base of the HOP descriptor. This experiment showed that a phase of the image can carry more structural information than the amplitude does. Hence, the phase congruency approach can be used to detect and localize the significant point features of the image better than the traditional gradient operators such as Prewitt, Laplace, Sobel, and Canny edge detector especially in the regions affected by illuminations. 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 [14],[15]. This property gives an advantage for phase congruency to detect and measure the significant features of the image over a constant threshold value that can be applied across a wide range of image classes. However the gradient based operators, which are sensitive to illumination and contrast variation, cannot localize the features accurately or consistently under a fixed threshold [14], [15]. The framework that illustrates the HOP descriptor is shown in Figure 1. The input image is divided into local regions. The phase congruency magnitude and orientation are computed for each pixel in the input image. The histogram of oriented phase are computed and normalized for each local region. These histograms are concatenated to construct the HOP descriptor. The dimensionality of HOP feature vector is reduced using the principal component analyses technique to form HOP-PCA descriptor.

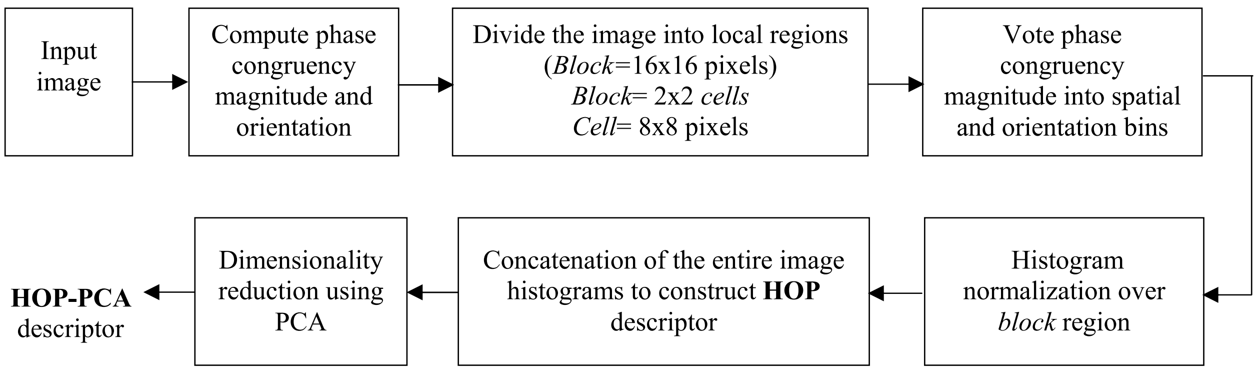


Figure 1. Framework of the construction of HOP-PCA descriptor.

The remaining sections of this paper are organized as following: In section 2, we discuss the concept and theoretical derivation of the phase congruency. In Section 3, we describe how to construct the HOP features. In section 4, the principal component analysis technique is discussed. Then, in Section 5, the experimental results are explained. Finally, the conclusion is presented in Section 6.

2. PHASE CONGRUENCY CONCEPT AND COMPUTATION

Phase congruency (PC) is an algorithm developed by Kovesi [16] to detect and localize edges and corners of digital images. This algorithm was based on a local energy model [17] and postulates that the features are perceived at the points of the strong phase congruency [16], [14], [25], [26].

Assume that an input periodic signal $I(x)$ is a one dimensional signal defined in the range [−π, π]. The Fourier components of the input signal meet in the edges where the phase congruency is maximum. Since the Fourier components meet at the edges, the local energy $E(x)$ would be also maximum at these points. The phase congruency becomes a dimensionless quantity by the normalization with the sum of all Fourier components amplitudes A_n [18], [19], [25], [26]. 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 [14], [16]. 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 [14], [25], [26].

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 [19], [25], [26].

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

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 [20], [25], [26]:

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

where ω_o is the center frequency of the filter. k/ω_o is kept constant for various ω_o [21], [22]. 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 [19], [21], [25], [26].

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 [25], [26]:

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

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:

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

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

Hence, phase congruency $P(x, y)$ of the 2D signal can be computed over various scales and orientation as [25], [26]:

$$PC(x, y) = \frac{\sum_o \sqrt{(\sum_n e_{no}(x, y))^2 + (\sum_n o_{no}(x, y))^2}}{\varepsilon + \sum_0 \sum_n A_{no}(x, y)} \tag{10}$$

The orientation angle $\varphi(x, y)$ is given by [25], [26]:

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

The level of phase congruency lies between the values 0 and 1. Figure 2 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 [25], [26].

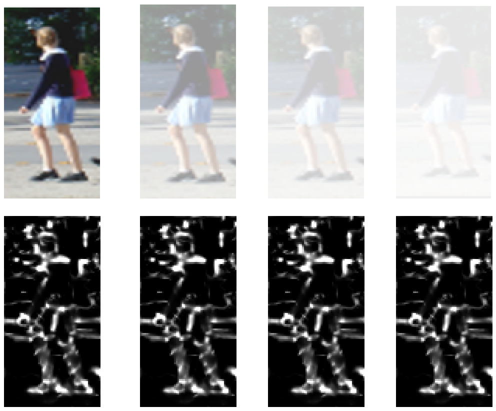


Figure 2. Phase congruency of the image at various illumination levels

3. EXTRACTION OF HOP FEATURES

After computing the phase congruency and its corresponding orientation angle for the input image $I(x, y)$, the Histogram of Oriented Phase (HOP) features can be constructed as described below.
Suppose (W) is a *cell* centred at (a, b) with a window size $(2d + 1) \times (2d + 1)$. N-bin histogram of oriented phase $B(a, b, z)$ within the *cell* window can be formed by accumulating the votes $T(a, b, z)$ of each pixel as in Eq. 12.

$$B(a, b, z) = \sum_{i=-d}^{d}\sum_{j=-d}^{d} T(a + i\,, b + j, z)\,, \qquad 0 \le z \le N - 1 \tag{12}$$

where N is the bin size of the histogram.

Since the bins of the orientation angle are in discrete values, a bilinear interpolation is used to vote into two neighbouring bins for the values that does not locate at the centre of the bin within the *cell* window [23]. Assume, $\varphi_1(a, b)$ and $\varphi_2(a, b)$ are the nearest bin centres to the orientation angle $\varphi(a, b)$. Suppose $z_1(a, b)$ and $z_2(a, b)$ are the indexes in the histogram of the oriented phase $B(a, b, z)$ corresponding to the bin centres $\varphi_1(a, b)$ and $\varphi_2(a, b)$ respectively. Therefore the vote $T(a\,, b, z)$ at the location (a, b) is consisting from the contribution of the bins $z_1(a, b)$ and $z_2(a, b)$ as given in Eq. 13.

$$T(a, b, z) = T\big(a, b, z_1(a, b)\big)\delta\big(z - z_1(a, b)\big) + T\big(a, b, z_2(a, b)\big)\delta\big(z - z_2(a, b)\big) \tag{13}$$

where δ is the discrete time impulse function.

$$T\big(a, b, z_1(a, b)\big) = \left(1 - \frac{\varphi(a, b) - \varphi_1(a, b)}{D}\right) PC(a, b) \tag{14}$$

where D is the bandwidth of the histogram such that $D = \pi/N$.

$$T\big(a, b, z_2(a, b)\big) = \left(\frac{\varphi(a, b) - \varphi_2(a, b)}{D}\right) PC(a, b) \tag{15}$$

$$z_1(a,b) = \left\lfloor \frac{\varphi(a,b) - D/2}{D} \right\rfloor + 1 \tag{16}$$

$$z_2(a,b) = \begin{cases} 0 & z_1(a,b) = N - 1 \\ z_1(a,b) + 1 & otherwise \end{cases} \tag{17}$$

$$\varphi(a,b) = (z_1(a,b) - 1/2)D \tag{18}$$

The operator $\lfloor . \rfloor$ denotes that the enclosed quantity is rounded towards minus infinity.

Now, the image $I(x, y)$ is divided into a local regions called *blocks* in the size 16×16 pixels. Each *block* consists of 2×2 *cells* with a *cell* size of 8×8 pixels. These *blocks* are 50% overlapped. The histogram of the oriented phase of *cell* in *block* region is computed. Finally, the overall Histogram of Oriented Phase for the entire image is constructed by concatenating the whole histograms extracted from each *block* in one feature vector representing the HOP features.

4. PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis is one of the well-known techniques used for dimensionality reduction to make the characteristic expression more compact. Assume there is a training sample set $U = \{u_1, u_2, \dots, u_R\}$, where R is the size of these training samples. The mean of the training sample can be computed as:

$$\bar{u} = \frac{1}{R}\sum_{m=0}^{R} u_m \tag{19}$$

The covariance matrix S of the training samples is defined as:

$$S = \frac{1}{R}\sum_{m=1}^{R} (u_m - \bar{u})\,(u_m - \bar{u})^T \tag{20}$$

We can note that the covariance matrix S is a real symmetric matrix. Therefore, its eigenvectors are also real values [27]. These eigenvectors are sorted in the descending order as $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p$ where p is the number of eigenvectors. So the q number of the selected feature vectors are corresponding to the q number of the largest eigenvectors.

5. THE EXPERIMENTAL RESULTS

In this section, the feature extraction algorithm HOP-PCA is applied for human detection system. The framework of this system is illustrated in Figure 3. The linear Support Vector Machine (SVM) from the VLfeat toolbox is used as classifier of this system.

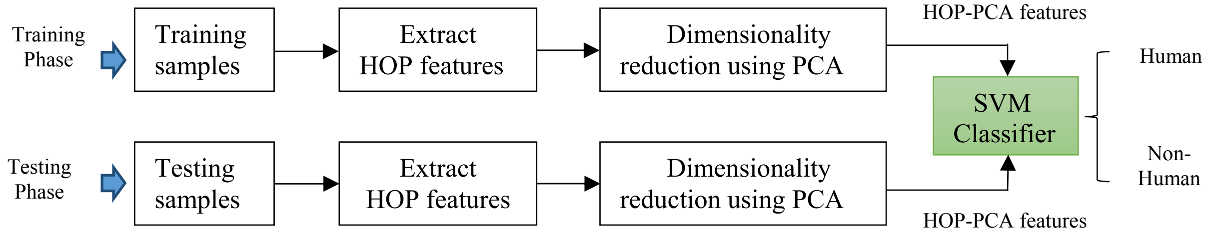


Figure 3. Framework of the human detection of HOP-PCA descriptor.

The detection system based on HOP-PCA descriptor is evaluated in comparison with the detection systems based on the feature algorithms: Histogram of Oriented Phase (HOP), Histogram of Oriented Gradient (HOG), Pyramid HOG (PHOG), Uniform Local Binary Pattern (uLBP), and Central symmetric Local Binary Pattern (CSLBP).

The performance of the detection system is analyzed and tested with various experiments using INRIA and DaimlerChrysler dataset. In INRIA dataset [6], 2416 positive training samples and 1126 positive testing samples of the size 128×64 pixels are used for training and testing the system respectively. On the other hand, 9000 negative training samples and 3750 negative testing samples of size 128×64 pixels are cropped from the negative images to train and test the system with the non-pedestrian samples. To evaluate the detector performance, we plot the curve of the miss rate ($miss\ rate\ = \frac{False\ Negative}{True\ Positive+False\ Negative}$) versus the false positive rate per window (FPPW) in the semi-log scale.

In the first experiment, the effect of the histogram binning size (6-bins, 9-bins, 18-bins) on the detection performance is analyzed. Figure 4 shows that the least miss rate of the detection system can be achieved at 9-bins orientation binning size.

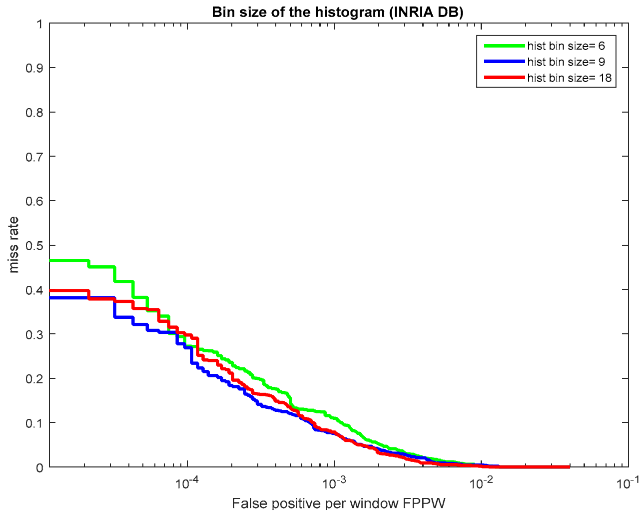


Figure 4. The detection performance at a different orientation binning size

Figure 5 illustrates the detection performance of the system at different forms of HOP descriptors. The first one uses a histogram with non-overlapped *blocks* {*block size=2×2 cells, cell size= 8×8 pixels*}; the second one uses *50% block-overlap* {*block size=2×2 cells, cell size= 8×8 pixels*}; the third one uses histogram of *50% block-overlap* {*block size=4×4 cells, cell size= 4×4 pixels*}. Based on these results, the optimal parameters that we selected to construct the HOP descriptor and used in the next analysis is 9-*bin* of orientation binning size; *50% block-overlap; block size=2×2 cells; and cell size= 8×8 pixels*.

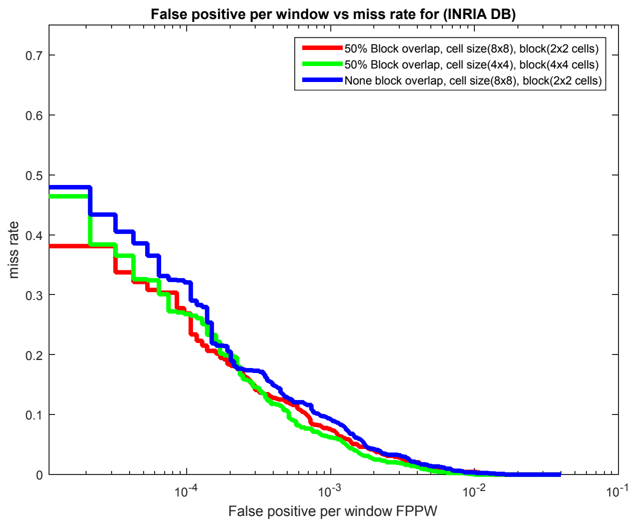
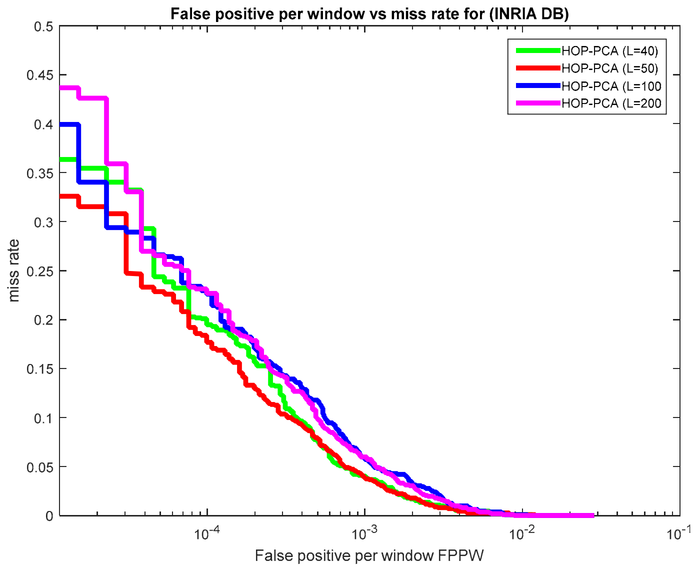


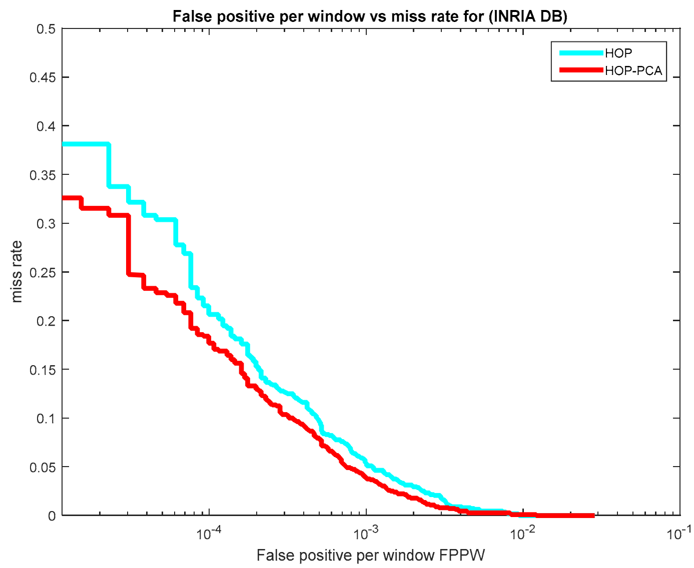
Figure 5. The detection performance at a different forms of the HOP descriptor histogram

In the next experiment, the principal component analysis PCA algorithm is applied to reduce the vector dimensionality of the HOP descriptor and to select just the strongest features. Figure 6a shows the miss rates of the detection system at various lengths of the HOP-PCA feature vector. Figure 6b illustrates the improvement in the system detection performance by applying the feature selection algorithm PCA and reducing the dimensionality of the HOP descriptor from 3780 to only 50 elements.



(a)

Figure 6. (a) The miss rates of the detection system at various dimensions of HOP feature vector.



(b)

Figure 6. (b) The improvement in the system detection performance by applying PCA algorithm

In the second experiment, the INRIA dataset used to evaluate the human detection system based on HOP-PCA descriptor in comparison with various detection systems based on low level feature algorithms as shown in Figure 7. At FPPW=10^{-4}, the miss rates of the detectors based on HOP, HOG, CSLBP, PHOG, and uLBP feature extraction algorithms are (26.87%), (33.48%), (37.02%),(66.16%), and (98.75%) respectively. The miss rate of HOP-PCA based detector is (20.8%).

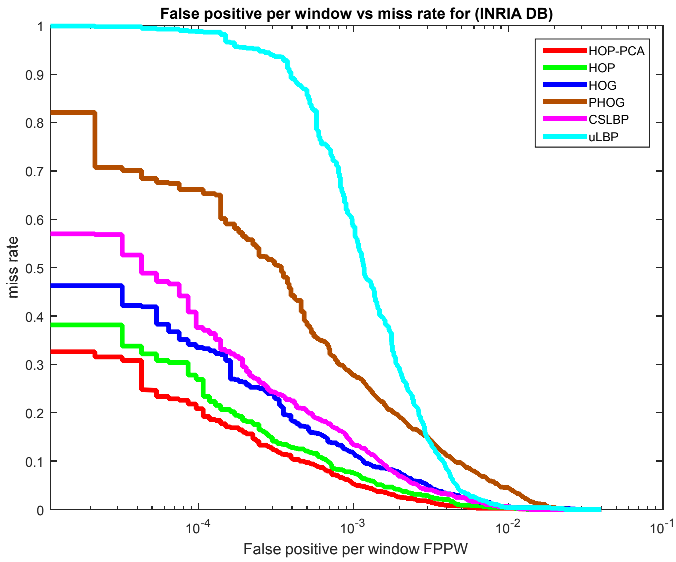


Figure 7. Detection performance of HOP_PCA based detector and its comparison with HOP, HOG, PHOG, uLBP, and CSLBP based detectors

In the third experiment, DaimlerChrysler dataset is used to evaluate the human detection system based on the proposed descriptor. DaimlerChrysler dataset contains a collection of low resolution grayscale pedestrian and non-pedestrian images in the size of 18×36 pixels [24]. 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). In this experiment the sets "1" & "T1" are used for training and testing the detection system respectively. At FPPW=10^{-3}, the miss rate of HOP-PCA based detector is (70.83%). However, the miss rates of the detectors based on HOP, HOG, PHOG, uLBP, and CSLBP feature extraction algorithms are (73.88%), (77.83%), (88.84%), (85.42%), and (93.85%) respectively as shown in Figure 8.

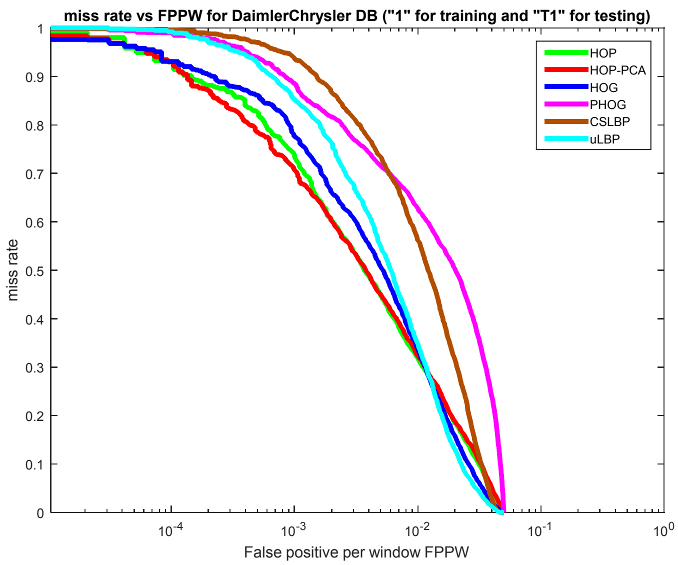


Figure 8. Detection performance of HOP_PCA based detector and its comparison with HOP, HOG, PHOG, uLBP, and CSLBP based detectors

6. CONCLUSION

Histogram of Oriented Phase (HOP) based on phase congruency is a feature extraction algorithm that has been presented for human detection. This descriptor can localize image details better than the gradient based approaches especially in the regions impacted by illumination and contrast changes. The principal component analysis PCA is applied to reduce the dimensionality of this descriptor and then improve the detection performance. The human detection performance based on HOP and HOP-PCA descriptors has been evaluated using the pedestrian datasets INRIA. The results of the experiments show better detection performance for the descriptor HOP-PCA over the HOP, HOG, PHOG, uLBP, and CSLBP feature extraction algorithms.

REFERENCES

[1] Cristina S., "Pedestrian Detection using a boosted cascade of Histogram of Oriented Gradients," Linköping – Barcelona, August 2014.

[2] Dollar P., Wojek C., Schiele B., Perona P., "Pedestrian detection: an evaluation of the state of the art.," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, No. 4, pp. 743-761, 2012.

[3] Gavrila D. M., "The visual analysis of human movement: a survey," Computer Vision and Image Understanding, vol. 73, nº 1, pp. 82-98, 1999.

[4] Papageorgiou C., Poggio T., "A trainable system for object detection," International Journal of Computer Vision, vol. 38, no. 1, pp. 15-33, 2000.

[5] Whitehill J., Omlin W., "Haar features for FACS AU recognition" Proc. IEEE International Conference ON Automatic Face and Gesture Recognition (AFGR06), pp. 217-222, 2006.

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

[7] Wu y B., Nevatia R.,"Optimizing discrimination-efficiency tradeoff in integrating heterogeneous local features for object detection," IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2008.

[8] Wang X., Han T., Yan, "An HOG-LBP human detector with partial occlusion handling," IEEE 12th International Conference on Computer Vision, pp. 32-39, 2009.

[9] Wojek C., Schiele B., "A performance evaluation of single and multi-feature people detection," Proceedings of the 30th DAGM symposium on Pattern Recognition, pp. 82-91, 2008.

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

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

[12] Skarbnik N., Yehoshua Y., Zeevi, Sagiv C., "The Importance of Phase in Image Processing," CCIT Report #773, August 2010.

[13] Yao Su C., Yang J., "Histogram of gradient phases: a new local descriptor for face recognition," Published in IET Computer Vision, 2014.

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

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

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

[17] Morrone M., Ross J., Burr, D., Owens R. "Mach bands are phase dependent," Nature 324 (1986) 250–253.

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

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

[20] Yuan X., Shi P.,"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.

[21] Basavaraj G., Reddy G., "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.

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

[23] Mathew A., Asari V., "Rotation-invariant histogram features for threat object detection on pipeline right-of-way", Proc. SPIE 9026, (March 5, 2014).

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

[25] Ragb H., Asari V., "Fused structure and texture (FST) features for improved pedestrian detection," International Journal of Computer and Information Engineering, (World Academy of Science, Engineering and Technology), vol. 3, no. 1, 2016.

[26] Ragb H., Asari V., "Histograms of oriented phase and gradient (HOPG) descriptor for improved pedestrian detection," IS&T International Conference on Electronic Imaging: Video Surveillance and Transportation Imaging Applications, San Francisco, California, USA, February 14, 2016.