https://doi.org/10.2352/ISSN.2470-1173.2017.10.IMAWM-172

Chromatic Domain Phase Features with Gradient and Texture for Efficient Human Detection

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

Abstract

In this paper, we propose a new human detection descriptor based on a combination of three major types of visual information: color, shape, and texture. Shape features are extracted based on both the gradient concept and the phase congruency in LUV color space. The Center-Symmetric Local Binary Pattern (CSLBP) approach is used to capture the texture information of the image. The fusing of these complementary information yields to capture a broad range of the human appearance details that improves the detection accuracy. The proposed features are formed by computing the phase congruency of the three color channels in addition to the gradient magnitude and CSLBP value for each pixel in the image with respect to its neighborhood. Only the maximum phase congruency values are selected from the corresponding color channels. The histogram of oriented phase and gradients, as well as the histogram of CSLBP values for the local regions of the image, are determined. These histograms are concatenated to construct the proposed descriptor, that fuses the shape and texture features, and it is named as Chromatic domain Phase features with Gradient and Texture (CPGT). Several experiments were conducted to evaluate the performance of the proposed CPGT descriptor. The experimental results show that the proposed descriptor has better detection performance and lower error rates when compared to several state of art feature extraction methodologies.

Introduction

Human detection is an active research topic in computer vision and pattern recognition systems. It can be stated simply as the localization of the regions in an image or video sequence containing humans. The fluctuating appearance of the human body combined with the occlusions, cluttered scenes and illumination changes, make the human detection task as one of the challenging categories in object detection. The human detection system is mainly consists of two major procedures: feature extraction and classification. The feature extraction algorithm is used to encode the image regions as low dimensional feature vectors that support high accuracy human/non-human decisions. The classifier unit uses these extracted features to determine whether the image region belongs to the object of interest or not. One of the earliest algorithms used for pedestrian detection system is proposed by Papageorgiou et al. [1]. This technique used sliding window detector and applied the Support Vector Machine (SVM) with the multi-scale Haar wavelet. Viola and Jones [VJ] [9] has upgraded this idea and introduced the integral images for fast features computation and a cascade structure for efficient detection [2]. In addition, they have developed a detector that integrates Haar-like features with Adaboost classifier [2]. These contributions continue today to serve as a foundation of the modern detection techniques. Important gains in the detection performance came with the adoption of the gradient based features. Inspired by Scale Invariant Features transform (SIFT), Dalal and Triggs [3] 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 [4]. Zhu et al [5] used the integral histogram to speed up the HOG features [6]. Shashua et al. [7] proposed Edge orientation histogram (EOH) which is a similar representation used with SVM for modeling of pedestrians [4]. Motion information is another cue for a human interception that used to improve the detection performance and the reliability for practical applications. Histograms of Flow descriptor is one approach that includes motion features and designed for situations with moving camera or background [8]. Haar wavelets and a cascade boosting scheme is another approach that includes the motion features and used to build a detector for static-camera [9], [10]. Although all of the previous low-level features showed considerable good results in the human detection system, however, they are still not enough to depict accurately all visual information especially at the poor image resolution, occlusion, and crowded scenes. Since the detection efficiency of the single feature based techniques is limited, it is envisaged that a representation based on multiple features would be more effective in capturing the humans in the various background environments. Each of the additional features would provide complementary information and the overall performance is reasonably improved. Mao et al [11] improved the contour detection by introducing a system based on edgelets, Haar-like features, and Viola's AdaBoost cascade framework. Wu and Nevatia [12] combined HOG, covariance and edgelets features in one descriptor. Wang et al [13] combined HOG features with texture features. Wojek and Schiele [20] combined Haar-like features, shapelets, shape context, and HOG features. Zhang and Ram [14] have combined the Edgelets with the HOG features and they improved the detection performance of IR images. Dollar et al [15] developed the integral channel features (ICF) that combined HOG, gradient magnitude, and LUV color. In this paper, we propose a new feature extraction algorithm that combines the texture information of the image employing The Center-Symmetric Local Binary Pattern (CSLBP) approach with the image shape features using both, the gradient concept and the phase congruency information in color space. This algorithm is named as Chromatic domain Phase features with Gradient and Texture ($CPGT$). This feature extraction algorithm leads to extract the complementary information of the image, hence significantly increases the detection rate of the system. The framework of the human detection system based the proposed descriptor algorithm is shown in Figure 1. The proposed features are formed by computing the phase congruency of the three color channels in addition to the gradient magnitude and CSLBP value for each pixel in the image with respect to its neighborhood. Only the maximum phase congruency values are selected from the corresponding color channels as shown in Figure 2. The histogram of oriented phase (HOP), and histogram of oriented gradients (HOG), as well as

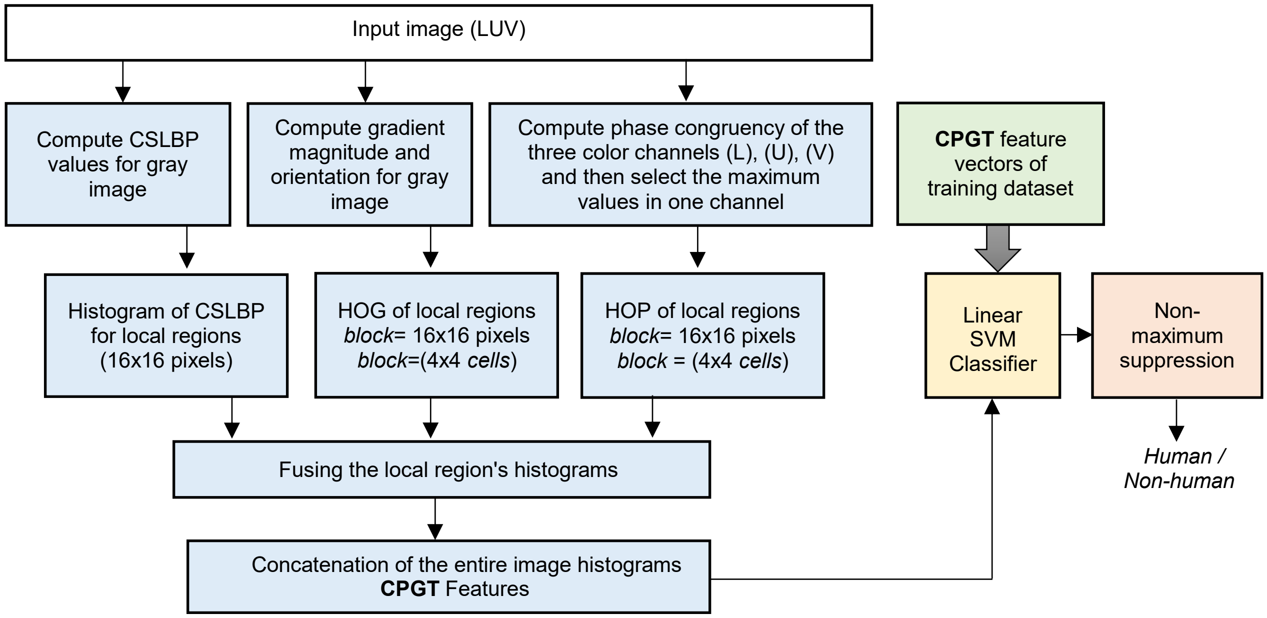

Figure 1. Framework of the human detection system based the proposed descriptor algorithm

the histogram of CSLBP values for the local regions (blocks) of the image are determined and combined together. These histograms are concatenated to construct the proposed descriptor. A support vector machine (SVM) classifier is used in this detection system to classify the CPGT features, where the scanning window approach is used to detect the presence of a human in an image. This approach returns a set of the detection window and the detection overlap that may occur due to sliding window is resolved using non-maximum suppression. Human detection system based on this algorithm is robust against illumination changes and able to detect the human objects in different scales, viewpoints, and postures as well as the detection of partial occlusion and realistic environments.

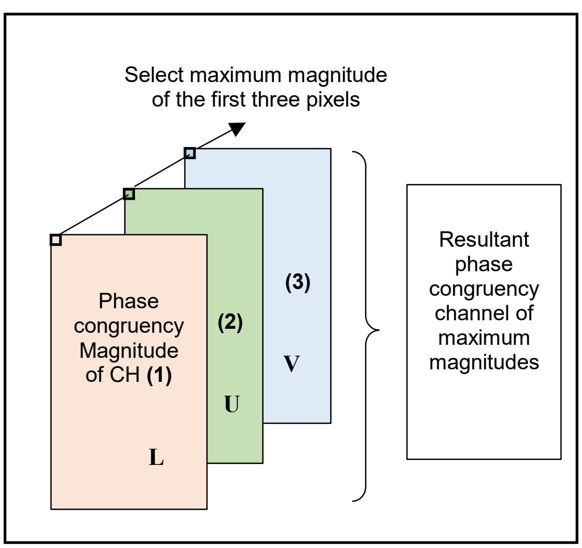

Figure 2. Phase congruency magnitude in chromatic domain

The remaining sections of this paper are organized in order as the following:

- Discussion of phase congruency principle and computation.
- Description of image gradients computation.
- Discussion of The texture features based on CSLBP technique.
- Implementation of $\boldsymbol{CPGT}$ descriptor.
- The experimental results.
- Conclusion.

Phase congruency principle and computation

Phase congruency PC is a technique developed by Kovesi [16] to localize the edges and corners of the image. It corresponds to features believed to be important in human vision. Some physiological evidence [17] has indicated that the human visual system responds strongly to points of high phase congruency in an image [18]. Phase is congruent when its values are equal or almost the same at a specific point in the signal [18]. For a one dimensional step signal shown in Figure 3, the Fourier components of this signal are meet together at the edges, where the phase congruency and the local energy $E(x)$ are maximum. Therefore, the phase congruency $PC(x)$ is directly proportional to the local energy . Venkatesh and Owens [23] show that local energy is equal to phase congruency PC scaled by the sum of the Fourier amplitudes A_n and given as [19], [20], [21], [22]:

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

where ε is a small number to avoid division by zero. The phase congruency becomes a dimensionless quantity by the normalization with A_n.

Assume $F(x)$ is equal the input signal $I(x)$ filtered from a DC component and $F_H(x)$ is 90° phase shift of $F(x)$ (Hilbert

Transform). The local energy function $E(x)$ can be defined as [21], [24], [25]:

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

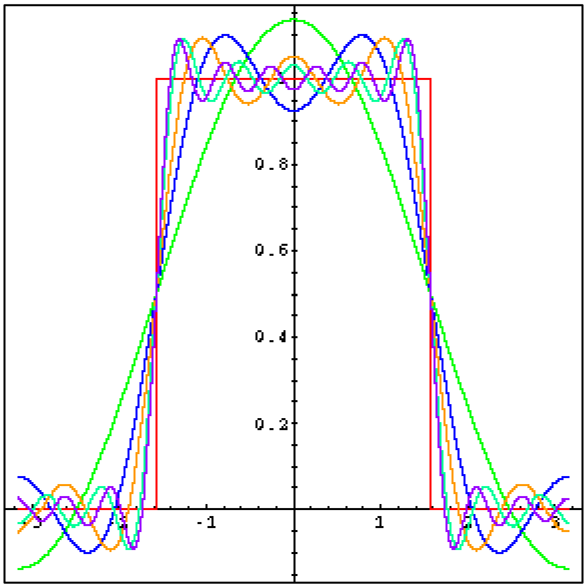

Figure 3. The Fourier components of a step signal

The first step for computing the phase congruency of the given image is to extract as much as possible of the local frequencies and the phase information by convolving this image by a pair of quadrature filters. Log-Gabor filter is an efficient band-pass filter used in this paper to extract the local phase information spread over a wide spectrum.

Consider M^o_{no} and M^e_{ne} are the odd symmetric and even symmetric components of the Log- Gabor filter at scale n and orientation o. 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 as [31]:

$$[e_{no}(x,y)\,,o_{no}(x,y)] = [I(x,y) * M^e_{no}\,,I(x,y) * M^o_{no}] \qquad (3)$$

$F(x,y)$ and $F_H(x,y)$ is defined as:

$$F(x,y) = \sum_o \sum_n e_{no}(x,y) \qquad (4)$$

$$F_H(x,y) = \sum_o \sum_n o_{no}(x,y) \qquad (5)$$

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

$$A_{no} = \sqrt{(e^2_{no}\,(x,y) + o^2_{no}\,(x,y))} \qquad (6)$$

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

Hence, the phase congruency $PC(x,y)$ of the 2D signal is computed as:

$$PC(x,y) = \frac{\sum_o \sqrt{(\sum_n e_{no}(x,y))^2 + (\sum_n o_{no}(x,y))^2}}{\varepsilon + \sum_o \sum_n A_{no}(x,y)} \qquad (8)$$

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

$$\varphi(x,y) = tan^{-1}\left(\frac{\sum_o \sum_n o_{no}(x,y)}{\sum_o \sum_n e_{no}(x,y)}\right) \qquad (9)$$

Phase congruency magnitude varying in the range (0 to 1). Figure 4 shows an example of the original image, the corresponding phase congruency image, and an extended cell illustrated the phase congruency magnitude and orientation in each pixel.

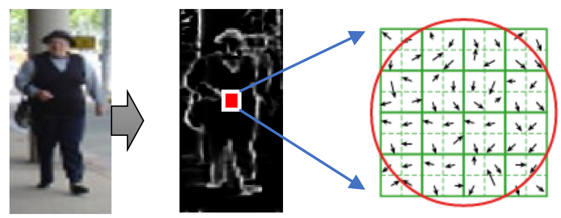

Figure 4. (a) Original image. (b) Corresponding phase congruency image. (c) Extended cell shows the phase congruency magnitude and orientation in each pixel

Image Gradient

Image gradient is defined as the directional change in the gray level or the color. Consider G_x and G_y is the center horizontal and vertical gradient of the input image $I(x,y)$ respectively. The horizontal gradient is computed by convolving the input image by the mask $(1 \quad 0 \quad -1)$. The horizontal gradient is computed by convolving the input image by the mask $(1 \quad 0 \quad -1)^T$. Therefore;

$$G_x(x,y) = I * (1 \quad 0 \quad -1) \qquad (10)$$

$$G_y(x,y) = I * (1 \quad 0 \quad -1)^T \qquad (11)$$

The gradient magnitude $G(x,y)$ and the orientation $\phi(x,y)$ for the image $I(x,y)$ can be computed as following [26];

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \qquad (12)$$

$$\phi(x,y) = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \qquad (13)$$

Texture features based on Center Symmetric Local Binary Pattern (CSLBP) operator

Center Symmetric Local Binary Pattern CSLBP is a powerful approach used to capture the texture features of the gray-level image as well as gradients [27]. CSLBP is a modified scheme of the original LBP algorithm [27]. It is computed by comparing the center-symmetric pairs of pixels with a central pixel (n_c), rather than comparing each pixel with the center [27], as shown in Figure 5. CSLBP approach reduces the computation and results a smaller dimension than LBP, while maintain certain characteristics such as robustness against monotonic gray level changes and tolerance of the illumination changes [27] [28] [29].

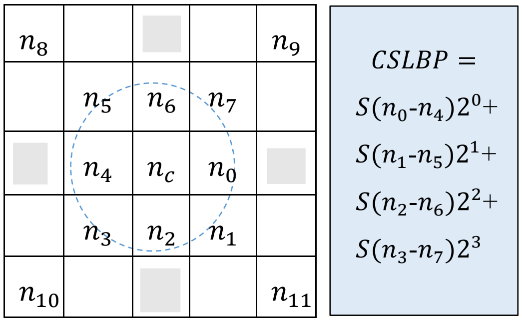

Figure 5. Center Symmetric Local Binary Pattern CSLBP features for neighborhood of 8-bit

Let $CSLBP_{N,R}$ denote the CSLBP feature of a pixel's circular neighborhood, where N represents the total number of sampling points and R represents the radius of the circle, then $CSLBP_{N,R}$ can be computed as:

$$CSLBP_{N,R}(x,y) = \sum_{i=0}^{N/2-1} S(|n_i - n_i + (N/2)|)\, 2^i \qquad (14)$$

$$S(z) = \begin{cases} 1 & if\ z \geq t \\ 0 & otherwise. \end{cases} \qquad (15)$$

where, n_i and $n_i + (N/2)$ is the gray level values of the center symmetric pairs of pixels, t is a small value for thresholding the gray level differences (t = 0.1 *is selected*). This thresholding maintains robustness in flat regions [27]. x and y are the coordinates of the center pixel. Notes that, CSLBP generates only 16 different binary patterns whereas LBP generates 256 different binary patterns.

CPGT descriptor

The proposed $CPGT$ descriptor is implemented by computing three channels from the input image: The first channel is the phase congruency magnitude and orientation in LUV color space; the second channel is the gradient magnitude and orientation in the gray level; the third channel is the CSLBP texture features. Each of these channels is divided into local regions called *blocks* of the size 16×16 pixel, and each *block* is divided into 4×4 *cells* as shown in Figure 6. Histogram of Oriented Phase features in LUV color space ($CHOP$) [22], and the Histogram of Oriented Gradient (HOG) are computed for each *cell* in the first *block* region ($CHOP_{c1}\ldots\ CHOP_{c16}$) and ($HOG_{c1}\ \ldots.\ HOG_{c16}$) Respectively. Histogram of the center symmetric local binary pattern for the first *block* region ($CSLBP_{b1}$) is also computed. These histograms are combined together to form the $CPGT_{b1}$ features for one *block* region as shown in Figure 7.

$$CHOP_{b1} = [CHOP_{c1}\ \ CHOP_{c2}\ \ldots CHOP_{c16}] \qquad (16)$$

$$HOG_{b1} = [HOG_{c1}\ \ HOG_{c2}\ \ldots HOG_{c16}] \qquad (17)$$

$$CPGT_{b1} = [CHOP_{b1}\ \ HOG_{b1}\ \ CSLBP_{b1}] \qquad (18)$$

The same is done for the rest local regions and the overall histograms are concatenated to construct the proposed $CPGT$ descriptor.

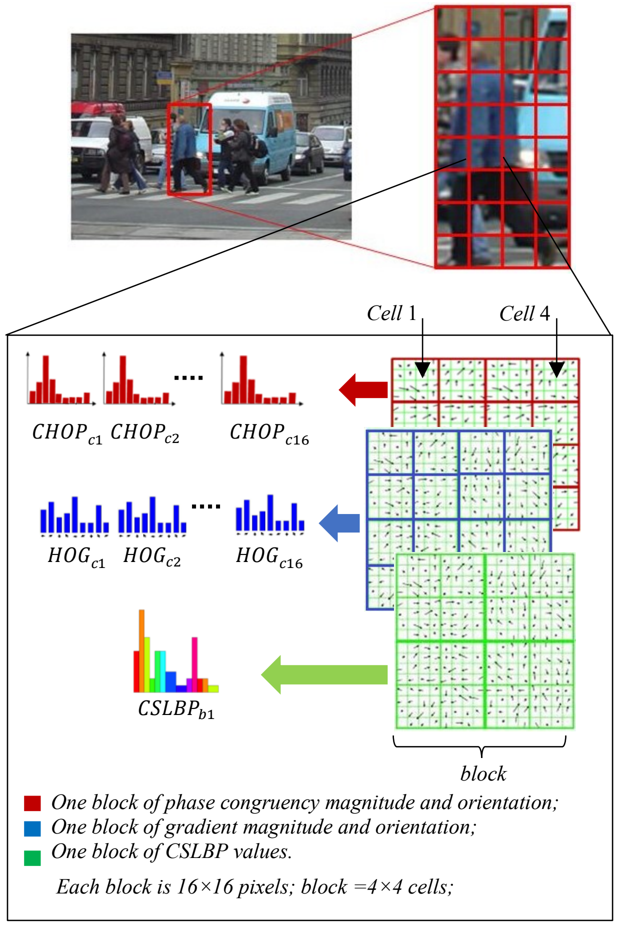

Figure 6. Blocks of one local region; phase congruency, gradients, and textures based on CSLBP features

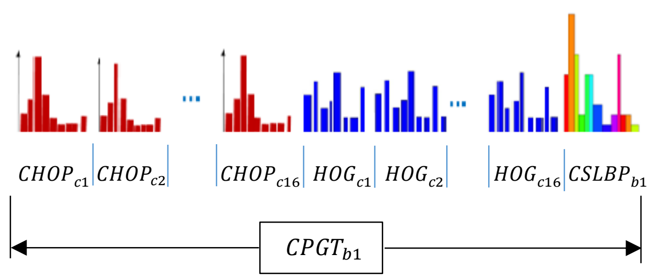

Figure 7. Construction of CPGT features for one local region

Experimental Results

In this section, a human detection system based on $CPGT$ descriptor is evaluated using INRIA and NICTA dataset. Linear Support Machine Vector (SVM) is used as classifier in this detection system. The detection performance of the proposed descriptor is compared with several algorithms including: HOG

(Dalal &Triggs), Pyramid-HOG (PHOG), Felzenszwalb-HOG (FHOG), CSLBP, Histogram of Oriented Phase (HOP) [31], HOP+PCA [31], (CHOP) [22], and the multi-features: HOGcslbp, Fused Structural and Texture (FST) features [24], Histogram of Oriented Phase and Gradient (HOPG) [25], FPGT [30], and (FPGT+PCA) [30].

In the first experiment, INRIA dataset is used for training and testing the human detection system based on $CPGT$ descriptor. 2416 positive (human) samples and 9750 negative (non-human) samples from INRIA dataset 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.

The performance of human detection system that based on $CPGT$ descriptor at different color spaces (RGB, LUV, HSV, YCrCB, and GRAY), is tested and analyzed. The results ilustrated in Figure. 8 shows that the optimal detection performance is achieved at the LUV and HSV color spaces. In the next experiments, the $CPGT$ descriptor at LUV color space is used for human detection and evaluated using INRIA and NICTA datasets.

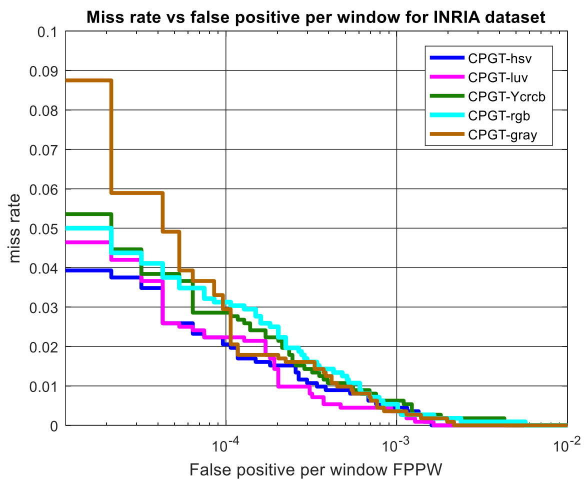

Figure.8. Human detection performance based on CPGT descriptor at different color spaces: RGB, LUV, HSV, and YCrCB

The detection performance of the system based on $CPGT$ descriptor and the performance of several state of the art methodologies is shown in Figure 9. At FPPW= 10^{-4} the miss rate of the proposed descriptor is 2.23%. However the miss rates of PHOG, CSLBP, HOG, FHOG, HOP, HOP+PCA, CHOP, HOGcslbp, FST, HOPG, FPGT, and FPGT+PCA based detectors are (66.16%), (37.59%), (33.48%), (26.87%), (24.2%), (20.8%), (15.18%), (11.07%), (6.071%), (5.089%), (3.034%), and (2.946%) respectively.

In the second experiment, NICTA dataset is used to evaluate the proposed descriptor using NICTA dataset. 1000 samples of human and 2000 non-human samples from this dataset are used in the training process. 500 positive samples and 500 negative samples from the NICTA dataset are used in the testing phase. The detection performance of the system based on $CPGT$ descriptor in comparison with other state of the art methodologies is shown in Figure 10. At FPPW=10^{-3}, the miss rate of the detection system based on $CPGT$ descriptor is 2.6%. On the other hand, the miss rates of the detectors based on uLBP, CSLBP, PHOG, HOG, HOP, FHOG, CHOP, FST, HOPG, and FPGT descriptors are (45.2%), (35.8%), (18.5%), (9.4%), (8.8%), (6.6%), (5.4%), (4.8%), (3.8%), and (3.78%) respectively.

It is obvious from the evaluation results applied on INRIA and NICTA dataset that the proposed descriptor CPGT has better detection performance over the mentioned algorithms.

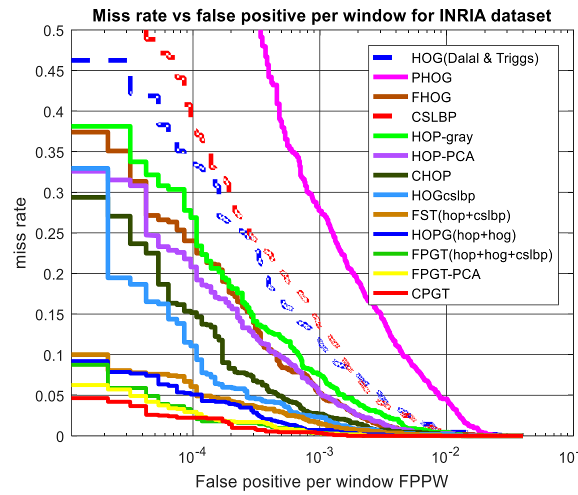

Figure 9. Detection performance of CPGT detector on INRIA dataset and its comparison with a several state of the art based detectors

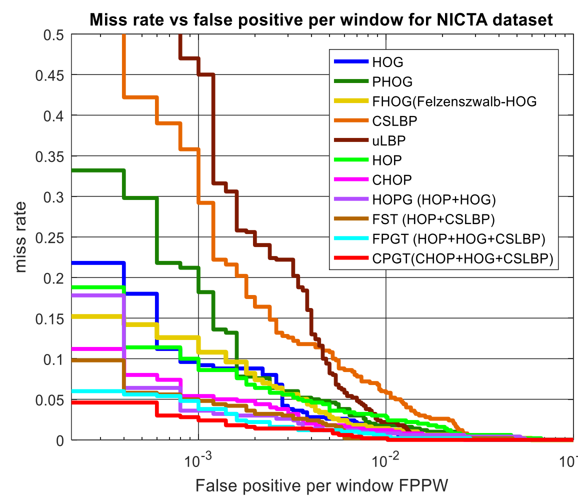

Figure 10. Detection performance of CPGT detector on NICTA dataset and its comparison with a several state of the art based detectors

Conclusion

Chromatic domain Phase features with Gradient and Texture ($CPGT$) is a new descriptor presented for improved human detection performance. The proposed feature extraction approach fuses multiple-features in one descriptor, including local phase information in LUV color space, image gradients, and texture features. This combination lead $CPGT$ descriptor to capture various information of the image complementary, hence increased the human detection performance significantly. Results of the experiments conducted on INRIA and NICTA datasets showed that $CPGT$ descriptor has better detection performance over several state of the art methodologies.

References

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

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

[3] 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.

[4] P. Dollar, C. Wojek, B. Schiele y P. Perona, "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.

[5] Q. Zhu, M.-C. Yeh, K. Cheng and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 1491-1498, 2006.

[6] F. M. Porikli, "Integral histogram: a fast way to extract histograms in cartesian spaces," IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 829-836, 2005.

[7] A. Shashua, Y. Gdalyahu y G. Hayun, "Pedestrian detection for driving assistance systems: single-frame classification and system level performance." IEEE International Conference on Intelligent Vehicles, pp. 1-6, 2004.

[8] N. Dalal, B. Triggs y C. Schmid, "Human detection using oriented histograms of flow and appearance," Proceedings of the 9th European Conference on Computer Vision, pp. 428-441, 2006.

[9] P. Viola, M. J. Jones y D. Snow, "Detecting pedestrians using patterns of motion and appearance," Proceedings of the 9th International Conference on Computer Vision, vol. 1, pp. 734-741, 2003.

[10] C. Sancho, "Pedestrian Detection using a boosted cascade of Histogram of Oriented Gradients" Linköping – Barcelona, Aug 2014.

[11] X. Mao, F. Qi, W. Zhu, "Multiple-part based pedestrian detection using interfering object detection," Proceedings of the 3rd International Conference on Natural Computation, vol. 2, pp. 165-169, 2007.

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

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

[14] 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.

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

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

[17] M. Morrone, J. Ross, D. Burr, R. Owens, "Mach Bands are Phase Dependent", Nature 324, 1986

[18] R. Dalvi and R. Abugharbieh, "Fast Feature-Based Multi-Slice to Volume Registration Using Phase Congruency", 30th Annual International IEEE EMBS Conference Vancouver, British Columbia, Canada, August 20-24, 2008.

[19] R. A. Owens, "Feature-free images," Pattern Recognition Letters, 15:35–44, 1994.

[20] P. Kovesi, "Image Features from Phase Congruency," Journal of Computer Vision Research, summer 1999, Volume 1, Number 3.

[21] S. Gundimada and 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.

[22] H. Ragb and V. Asari, "Color and local phase based descriptor for human detection," *National Aerospace & Electronics Conference & Ohio Innovation Summit (NAECON-OIS*), Dayton, Ohio, USA, 26 - 29 July 2016. (NAECON 2016).

[23] S. Venkatesh and R. Owens, "On the classification of image features,".Pattern Recognition Letters, 11:339–349, 1990.

[24] H. Ragb, V. Asari, "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.

[25] H. Ragb, V. Asari, "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, February 14, 2016.

[26] 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.

[27] B. Chul, D. Kim, J. Jung, and J. Nam, "Three-level cascade of random forests for rapid human detection", Optical Engineering 52(2), 027204 (February 2013).

[28] Y. Zheng, C. Shen, R. Hartley, X. Huang, "Effective Pedestrian Detection Using Center-symmetric Local Binary/Trinary Patterns", IEEE, SEPTEMBER 2010.

[29] M. Heikkilä, M.Pietikäinen, C. Schmid, "Description of Interest Regions with Center-Symmetric Local Binary Patterns", ICVGIP 2006: 58-69.

[30] H. Ragb and V. Asari, "Multi-feature fusion and PCA-based approach for efficient human detection," IEEE Computer Society Workshop on Applied Imagery and Pattern Recognition - AIPR 2016: Imaging and Artificial Intelligence: Intersection and Synergy, Washington DC, USA, 18 - 20 October 2016. (IEEE AIPR).

[31] H. Ragb and V. Asari, "Histogram of oriented phase (HOP): a new descriptor based on phase congruency," SPIE Conference on Commercial + Scientific Sensing and Imaging: Mobile Multimedia/Image Processing, Security, and Applications 2016, Baltimore, MD, USA, 17 - 21 April 2016.

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 received his MS in Electrical Engineering from the University of Belgrade (2000). Since then he worked as a lecturer at the University of Tripoli. Currently, he is a Ph.D. 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 Ph.D. in Electrical Engineering from the Indian Institute of Technology, Madras. He holds three patents and has published more than 500 papers, including 89 journal papers. Dr. Asari is a Member of IS&T and a Senior Member of IEEE and SPIE.