https://doi.org/10.2352/ISSN.2470-1173.2018.10.IMAWM-375
© 2018, Society for Imaging Science and Technology

A feature fusion strategy for human detection in omnidirectional camera imagery

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

Abstract

Field of view of the traditional camera is limited such that usually more than three cameras is needed to cover the entire surveillance area. The use of multiple cameras usually requires more efforts regarding camera control and set up as well as they need additional algorithms to find the relationships among the images of different cameras. In this paper, we present a multi-feature algorithm that employs only one omnidirectional camera instead of using multiple cameras to cover the entire surveillance region. Here we use the image gradients, the local phase information based on phase congruency, the phase congruency magnitude, and the color features, and they are fused together to build one descriptor named as "Fused Phase, Gradients and Color features (FPGC). The image gradients, and local phase information based on phase congruency concept are used to extract the human body shape features. Either LUV or grayscale channel features are used according to the kind of camera used. The phase congruency magnitude and orientation of each pixel in the input image is computed with respect to its neighborhood. The resultant images are divided into local regions and the histogram of oriented phase, and the histogram of oriented gradient are determined for each local region and combined. A maximum pooling of the candidate features is generated for one channel of the phase congruency magnitude and the three LUV color channels. All these features are fed to a decision tree Adaboost classifier for training and classification between the classes. The proposed approach is evaluated on a challenging omnidirectional dataset and observed promising performance.

Introduction

Human 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 human detection application as one of the challenging categories in object detection [23]. Over the last decade, remarkable results in the area of human detection have been reported by numerous researchers. Many single feature extraction algorithms are proposed for depicting and describing the human appearance. One of the earliest algorithms used for human detection system is proposed by Papageorgiou et al [2]. This technique used sliding window detector and Harr-like features [3] for describing the person. Shape features such as Edgelet [4], Shapelet [5], and Histogram of Oriented Histogram (HOG) [6] are other features extraction algorithms proposed for human descriptor. In addition, texture features such as Local Binary Pattern (LBP) [7], as well as color features like color-self-similarity [1], and color histograms are also applied for human detection tasks. Among these features, the Histogram of Oriented Gradient (HOG) descriptor is the most popular gradient based technique. It provides high efficient and robustness features and showing substantial gains over the intensity based features [1], [8]. Although the mentioned single features played significant role in human detection, their ability of describing the objects in the challenging conditions was limited. That has raised up recently the need of multi-features descriptor which proposed for improving the human detection performance [30]. Wojek and Schiele [10] fused HOG, Haar-like features, shapelets and shape context in one descriptor to improve the detection performance. Wang [9] combined HOG and LBP features and used the linear Support Vector Machine (SVM) to train the human detector. Zhang and Ram [11] improved the detection of the IR images by the combination of the Edglets and HOG features. Dollar et al [12] developed the integral channel features (ICF) that combined HOG, gradient magnitude and LUV color, etc. All the mentioned algorithms are used and tested for traditional cameras. The field of view of the traditional camera is limited such that usually more than three cameras is needed to cover the entire surveillance area [31]. The use of multiple cameras is usually requiring more efforts regarding the control and set up as well as they need of additional algorithms to find the relationships among the images of the different cameras [31]. In this paper, we present a new multi-feature algorithm employs only one omnidirectional camera instead of using multiple cameras to cover all the surveillance view and to avoid these problems. A sample of an omnidirectional image is shown in Figure 1a. A simple transformation is used for unwrapping from omnidirectional into panoramic images as shown in Figure 1b.

(a) (b)

Figure 1. A sample of an omnidirectional imagery and its unwrapped image. (a) an omnidirectional image. (b) panoramic image.

In the proposed features extraction algorithm, the image gradients, the local phase information based on phase congruency, the phase congruency magnitude, and the color features of the panoramic image are extracted and fused together in one descriptor called "Fused Phase, Gradients and Color features (FPGC). The framework of the human detection system based on the proposed descriptor is shown in Figure 2. The phase congruency magnitude and the orientation of each pixel in the input gray image is computed with respect to its neighborhood. The resultant images are divided into local regions and

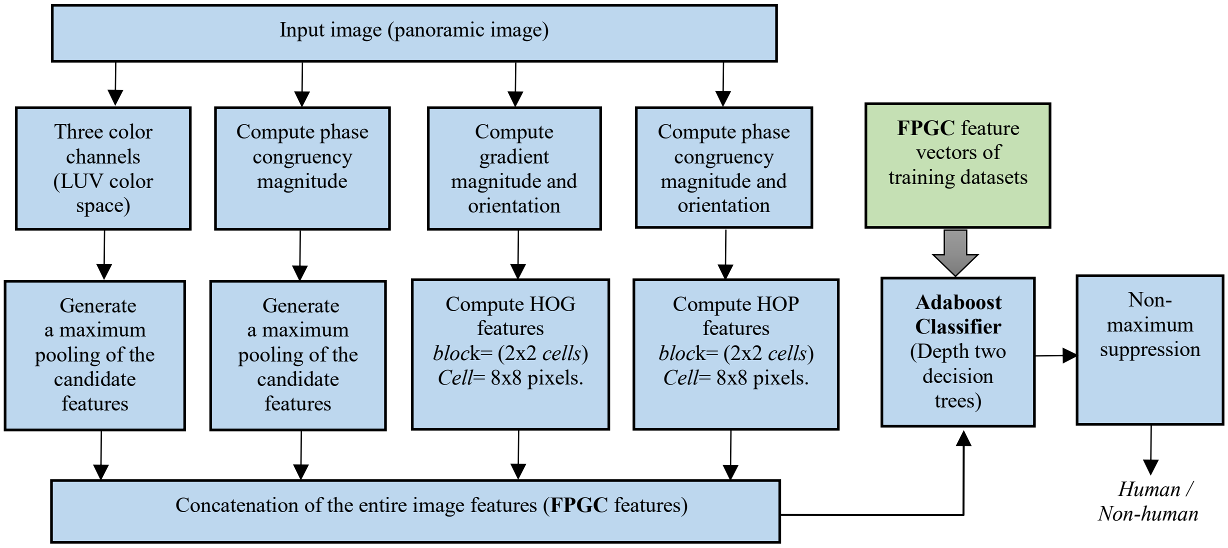

Figure 2. Framework of the human detection system based the proposed descriptor algorithm

the histogram of oriented phase, and the histogram of oriented gradient are determined for each local region and combined together. A maximum pooling of the candidate features is generated for a one channel of the congruency magnitude, and the three LUV color channels. All these features are concatenated and fed to a depth two decision tree Adaboost classifier for training and the classification between the classes. The proposed human detection system is implemented based on the scanning window approach that used to detect the presence of humans in an image. This method returns a set of detection window and the detection overlap that may occur due to sliding window is resolved using non-maximum suppression [30].

Phase Congruency computation

Phase congruency is an algorithm that was developed to localize the edges and corners of the image. Oppenheim and Lim [13] [14] have shown that the most significant information within an image is provided by the phase rather than amplitude [15]. Phase congruency provides a measure that is independent of the overall magnitude of the signal making it invariant to illumination and contrast variations [16]. The phase congruency function in terms of the Fourier series expansion of a signal at some location x is given as [24], [20], [30]:

$$PC(x) = max_{\bar{\phi}(x)\epsilon[0,2\pi]} \frac{\sum_n A_n \cos(\phi_n(x)-\bar{\phi}(x))}{\sum_n A_n} \quad (1)$$

where, A_n is the amplitude of the nth Fourier component [20]. $\phi_n(x)$ represents the local phase of the Fourier component. $\bar{\phi}(x)$ is the mean local phase angle of all the Fourier terms being considered at the point. An alternative and interpretation of phase congruency, Venkatesh and Owens [19] show that local energy is equal to phase congruency scaled by the sum of the Fourier amplitudes and given as [18], [20], [16], [28], [30]:

$$E(x) = PC(x) \sum_n A_n \quad (2)$$

For a one-dimensional input signal $I(x)$, the local energy function $E(x)$ can be defined as [16], [22], [23], [30]:

$$E(x) = \sqrt{F(x)^2 + F_H(x)^2} \quad (3)$$

where, $F(x)$ is the input signal filtered from a DC component and $F_H(x)$ is Hilbert Transform (90˚ phase shift of $F(x)$).

To compute the phase congruency, we should first extract the local frequencies and phase information by convolving the input image with a pair of quadrature filters. Log-Gabor filter is an efficient bandpass filter used in this paper to extract the local phase information spread over a wide spectrum [25]. The transfer function of the log-Gabor filter is given by [15], [22], [23]:

$$G(\omega,\theta) = exp\left(\frac{-\left(log(\omega/\omega_o)\right)^2}{2\left(log(k/\omega_o)\right)^2}\right) exp\left(\frac{-(\theta-\theta_o)^2}{2{\sigma_\theta}^2}\right) \quad (4)$$

where ω_o is the center frequency of the filter. k/ω_o is kept constant for various ω_o [34], [21]. θ_o is the center orientation of the filter, and σ_θ is the standard deviation of the Gaussian function in angular direction [16], [34], [22], [23]. The framework for computing the phase congruency is illustrated in Figure 3.

Consider M_{no}^o and M_{ne}^e are the odd symmetric and even symmetric components that represent the quadrature pair 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 by [22], [23], [28], [30]:

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

The convolution process is equivalent to the multiplication of Gabor filter $G(\omega,\theta)$ with the Discrete Fourier Transform (DFT) of the input image $I(x,y)$ followed by inverse DFT. The amplitude of

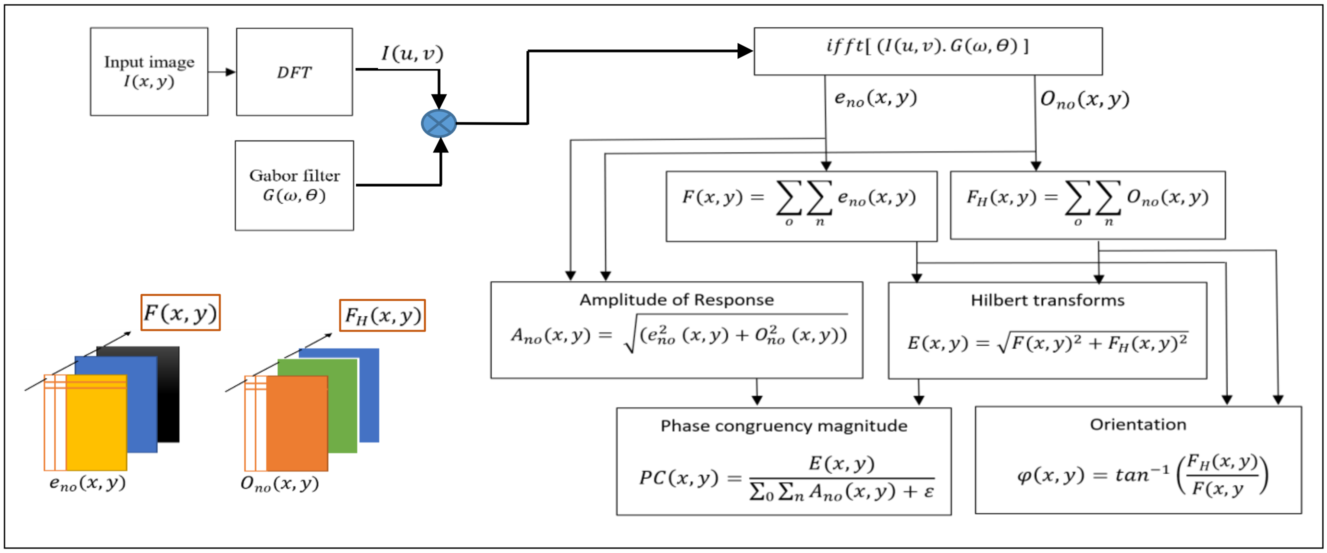

Figure 2. Block diagram shows the phase congruency computation.

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

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

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

$F(x,y)$ and $F_H(x,y)$ for a 2D signal are given by [25], [28]:

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

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

From Eq. (3), the energy of the two-dimensional signal is computed as;

$$E(x,y) = \sqrt{F(x,y)^2 + F_H(x,y)^2} \tag{10}$$

Therefore, phase congruency $PC(x,y)$ of the 2D signal is given as [22], [23], [25], [28], [30]:

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

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

$$\varphi(x,y) = tan^{-1}\left(\frac{\sum_o \sum_n o_{no}(x,y)}{\sum_o \sum_n e_{no}(x,y)}\right) \tag{12}$$

Image Gradient computation

Image Gradient is defined as the directional change in the color or an image intensity. The horizontal gradients $G_x(x,y)$ can be obtained simply by convolving the input image with the mask templet $(-1\ 0\ 1)$ and vertically $G_y(x,y)$ by convolving the image with $(-1\ 0\ 1)^T$[23]. The gradient magnitude $G(x,y)$ and the orientation $\phi(x,y)$ for the image $I(x,y)$ is computed as following [23], [24];

$$G_x(x,y) = I(x+1\,,y) - I(x-1\,,y) \tag{13}$$

$$G_y(x,y) = I(x\,,y+1) - I(x\,,y-1) \tag{14}$$

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \tag{15}$$

$$\phi(x,y) = \tan^{-1}\left(\frac{G_x(x,y)}{G_y(x,y)}\right) \tag{16}$$

Implementation of fused phase, gradient and color (FPGC) descriptor

Once the phase congruency magnitudes and orientations as well as the gradients of the input image are computed, the resultant images are divided into local overlapped regions called blocks of the size 16×16 pixels. The block region is formed from 2×2 sub-local regions called cells, where each cell is 8×8 pixels. The histogram of oriented phase [25] and the histogram of oriented gradient are determined for each cell region. These histograms are fused together to form the HOP and HOG features for each block region of the input image. The same is done for the rest of the blocks to form the HOP and HOG for the entire image. A maximum pooling of the candidate features is randomly generated for a one channel of the phase congruency magnitude and the same is done for the three LUV color channels (one-channel for grayscale). In proposed descriptor, maximum pooling is applied. These features are fused with HOP and HOG features to form the proposed descriptor FPGC. This descriptor is fed to Adaboost classifier (depth two decision tree) to select the strongest features to be used for training and the classification process.

Experimental results

A human detection system based on FPGC descriptor is tested and evaluated in comparison with the human detection systems based on the feature extraction algorithms: HOG (Dalal &Triggs) [6], ACF [32], and LDCF [33]. 2416 positive (person) samples and 12000 negative (non-person) samples from INRIA dataset are cropped in the size 128×64 pixels and used for training the detection system. Omnidirectional and panoramic image dataset is used for testing the human detection system. This dataset contains 30 omnidirectional images taken in different scenes including indoor and outdoor environment for detection standing human. In the evaluation of detection performance, we measured the accuracies of the human detection system in terms of the sensitivity (Recall) and the positive predictive value (PPV) (Precision), as given in Eq. (17) and Eq. (18). A higher value of sensitivity and PPV indicate a higher accuracy of human detection.

$$sensitivity = \frac{\#TP}{\#TP+\#FN} \tag{17}$$

$$PPV = \frac{\#TP}{\#TP+\#FP} \tag{18}$$

where, *TP* is "True Positive": mean that, the ground truth and the detected window are both belong to positive class "human". *FP* is "False Positive": mean that, the ground truth is belonging to negative class (non-human), but the detected window belongs to positive class. *FN* is "False Negative": mean that, the ground truth belongs to positive class, but the detected window belongs to negative class. Figure 3 and Figure 4 show the sensitivity and the PPV value of the human detection based on the proposed descriptor FPGC in comparison with HOG, ACF, and LDCF. These results show that FPGC, ACF, and LDCF have equal PPV value (100%). However, FPGC has higher sensitivity (100%) over HOG, LDCF, and ACF that equal (71.8 %), (81.7%), and (84.5%) respectively. Samples of human detection based on FPGC descriptor for omnidirectional images is shown in Figure 5.

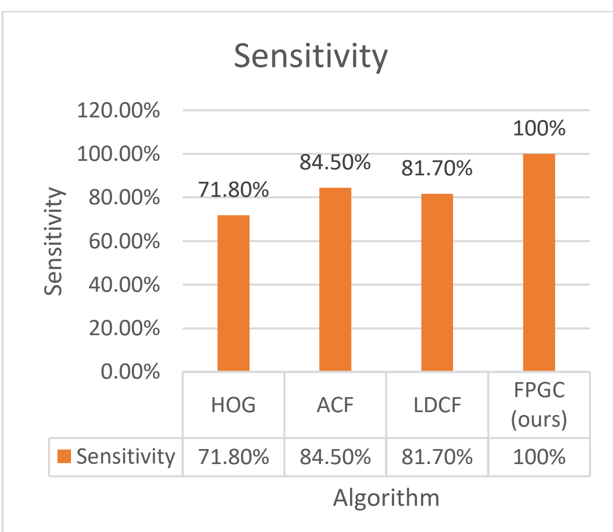

Figure 3 Sensitivity of human detection based on FPGC descriptor and it is comparison with HOG, ACF, and LDCF based detection technique.

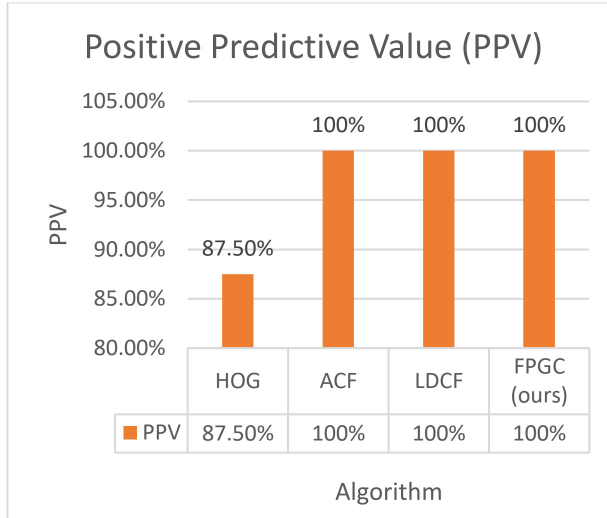

Figure 4 PPV of human detection based on FPGC descriptor and it is comparison HOG, ACF, and LDCF based detection technique.

Figure 5 Samples of human detection based on FPGC descriptor for omnidirectional images.

Conclusion

Fused Phase, Gradients and Color features (FPGC) is a new descriptor presented for human detection in omnidirectional camera imaginary. Only one omnidirectional camera is used to cover the entire surveillance region instead of using multiple cameras. The proposed feature extraction approach fuses image gradients, phase congruency magnitude, histogram of oriented phase, and color features in one descriptor. This combination lead FPGC descriptor to capture various information of the image complementary, hence increased the human detection performance significantly. Results of the experiments showed that FPGC descriptor has better detection performance over several state of the art methodologies.

References

[1] H. Liu , T. Xu , X. Wang, Y. QianA "Novel Multi-Feature Descriptor for Human Detection Using Cascaded Classifiers in Static Images", 31 December 2014# Springer Science+Business Media New York 2014.

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

[3] 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.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

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

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

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

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

[13] A.V. Oppenheim, J.S. Lim, "The importance of phase in signals," Proc. IEEE 69 (5), 529–541. 1981.

[14] A. Oppenheim, V., Lim, J.S., Kopec, G.E., Pohlig, S.C., 1979. Phase in speech and pictures. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, April, pp. 632–637.

[15] 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.

[16] 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, pp. 2074-2080, 2007.

[17] M. C. Morrone and R. A. Owens, "Feature detection from local energy. Pattern Recognition Letters," 6:303–313, 1987.

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

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

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

[21] 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.

[22] 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.

[23] 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, San Francisco, Feb. 14, 2016.

[24] P. Kovesi, "A dimensionless measure of edge significance." In the Australian Pattern Recognition Society, Conference on Digital Image Computing: Techniques and Applications, pages 281–288, Melbourne, 4-6 December 1991.

[25] 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.

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

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

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

[29] A. Savakis, R. Sharma1 and M. Kumar, "Efficient eye detection using HOG-PCA descriptor", Proceedings of SPIE - The International Society for Optical Engineering · February 2014.

[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: Washington DC, Oct. 2016. (IEEE AIPR).

[31] Yi Huang et al, “A Human Fall Detection System Using an Omni-Directional Camera in Practical Environments for Health Care Applications” MVA2009 IAPR Conference on Machine Vision Applications, May 20-22, 2009, Yokohama, JAPAN

[32] P. Dollar, R. Appel, S. Belongie, and P. Perona. Fast feature pyramids for object detection. ´ PAMI, 2014.

[33] W. Nam, P. Dollár, and J. H. Han. Local decorrelation for improved detection. In NIPS, 2014.

[34] 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.

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 96 journal papers. Dr. Asari is a Member of IS&T and a Senior Member of IEEE and SPIE.