ARPN Journal of Engineering and Applied Sciences

THE IMPLEMENTATION OF PERSONAL MONITORING SYSTEM USING HISTOGRAM OF GRADIENT, HAAR CLASSIFIER AND EIGENFACES FOR HUMAN DETECTION AND RECOGNITION

Z. Abal Abas, M. R. Ramli, A. S. Shibghatullah, H. Rahmalan, A. F. N. Abdul Rahman, Z. Zainal Abidin and B. Hussin

Optimization, Modelling, Analysis, Simulation and Scheduling Research Group, Centre for Advanced Computing Technology, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Durian Tunggal, Melaka, Malaysia

E-Mail: zuraidaa@utem.edu.my

ABSTRACT

The development of personal monitoring system aims as an initial study for home surveillance system that covers only frontal view of human face. This study focuses on a combination of computer vision analysis and video surveillance which has a wide range of application in monitoring and surveillance system. In general, the monitoring and surveillance system is vital in reducing crime cases as well as to protect public safety. Moreover, the development of personal monitoring focuses on two objectives; detecting the existence of human and performing face recognition for the detected human, which covers the pre-processing, processing and post processing stages. In the processing stage, background subtraction and Histogram of Gradient texture analysis techniques are used for human detection process. On the other hand, the Haar cascade and Eigenface recognizer techniques are used in the face recognition process. As the results, the combinations of these techniques from two phases produce a better human identification and faster detection. The impact of the personal monitoring development system shows a successful achievement on human recognition as well as an accurate identification of personal monitoring.

Keywords: personal monitoring, human detection, face recognition, image processing.

INTRODUCTION

Video surveillance has been an emerging research topic in image processing, which the technology is important to safeguard various environments and organization application system such as people, equipment, premises and vehicles. According to [1], there are three main categories for video surveillance: (1) Operator controlled video surveillance system. (2) Basic automated video surveillance system. (3) Intelligent video surveillance system.

Along with the third category that is the intelligent video surveillance system, machine learning in image processing is a significant concept that needs to be adopted, since this technique is important to transform an image into digital form and execute some operation in such a way of obtaining useful information regarding the image. In general, the framework of the general system model for video surveillance system starts with image acquisition followed by pre-processing, motion segmentation, object classification, human body tracking, human pose modeling, pose recognition and finally activity analysis [2].

Among the applications of video surveillance are home monitoring for smart home, elderly monitoring, security monitoring in public places and etc. In monitoring elderly at home environment, the object are detected based on various posture-based event such as sitting, standing, bending/squatting, side lying as well as lying towards the camera [3-5].

Home monitoring is an essential part of security in Smart Home now days since it is able to detect unusual event based on simultaneous tracking and event detection [6]. Personal monitoring application is a part of smart home monitoring, which is the main focus in this work. It must be noted that this work serve as an initial state for home surveillance. In order to analyse the video in this personal monitoring application, the image processing techniques are used. The human existence is firstly detected. Once human is detected, the individual is determined whether he or she is a familiar or not familiar. We propose a framework of personal monitoring and develop the system based on the framework. The development is based on the combination of histogram of gradient (HOG), HAAR and Eigenface recognizer. The outcome of the development is explained in the section of result and discussion. Future work section explains the extension of this current work. Finally, the conclusion section summarized the research at the end of this paper.

RELATED WORK

The advancement of current technology in computer vision has lead to a more tracking moving object been discovered. This includes the advancement in hardware and software. At the early state of hardware solution, there are some kind of sensor been used in order to get best human recognition such as multilayer scanner [7] or a video camera [8], infrared video [9]. New technology now days present system that merge information coming from different sensor types, like

LIDAR [10] and vision-based sensors[11]. Moreover, in order to obtain more comprehensive detection information for better result, multi sensor systems are required like multi-cue vision system [12]. However, it must be noted that this leads to higher material cost and maintenance efforts.

A general learning-based approach has been used for almost all of current vision based pedestrian detection system. In this approach, the appearance of human is represented as a simple low-level feature from region of interest [9]. A simple segmentation procedure is employed to get the foreground region of interest. However this simple segmentation such as background subtraction or a temporal differencing do have its limitation as it cannot deal with small or gradual changes in the background [9].

To overcome the problems, several techniques have been implemented such as Pfinder [13], [14]. However, this technique cannot deal with a large sudden change in the background. Another technique which is Independent Motion Detection will solve the large sudden change problem but it is quite difficult to develop and is not feasible for non-rigid object extraction [15], [16]. In pedestrian detection, there are two process involved which is the foreground detection and the recognition. For the recognition process, there are two approaches that can be applied; motion-based and shape-based. Motion based approaches is the most popular approaches used even though it requires a sequence of frames which make the processing time longer [17]. Shape based approaches in the other hand recognizes the pedestrian in moving and stationary state but still need more room to be improve as there are a lot of shape, pose, size and appearance of pedestrian that need to be model. There are also other techniques being proposed in the literature to solve the pedestrian tracking as well as face detection such as Gibson's approach [18], shape-based pedestrian detection [19], Probabilistic Template Based [9], View-Based and Modular Eigenspaces [20].

PROPOSED METHODOLOGY

There are three main stages in our framework of analysing video for personal monitoring as illustrated in Figure-1. The first stage is pre-processing where training is done for familiar faces being kept in the database. The second stage is the processing where it covers the process of detecting the existence of human followed by face recognition process. The post-processing which is the final stage will display the name of the person being detected in the database if the face is recognized as familiar. On the other hand, if the face is not recognized, hence the person being detected will be label as unknown. This section will explained more on the processes of human existence detection and face recognition.

Process 1: Human existence detection

The initial detection is performed by tracking the video in a frame-by-frame manner. The functionality of human existence detection aims to sense and localising the object with human shape. As a matter of facts, human detection process can be broken down into two processes which are object detection and object classification. In object detection, any foreground image will be detected as object, using background subtraction technique.

Background subtraction is one of the technique that has been used widely for generating a foreground mask (namely, a binary image containing the pixels belonging to moving objects in the scene) [9]. It computes the foreground mask by performing subtraction between the current frame and the background frame. It must be noted that the background frame is the initial frame and it is treated as the static part of the scene [9]. Among the approaches for background subtraction methods are mixture of Gaussian model, non-parametric background model, temporal differencing, warping background and hierarchical background model. The one being used in this work is a mixture of Gaussian model. The values of each pixel in the frame are models as a mixture of gradient. The mixture of Gradient is updated using new pixel value using K-means approximation [11].

Once object is detected, the object will then be classified as human or non-human whereby a rectangle region will be constructed as a hypothesis on detected human using texture based technique; Histogram of Gradient (HoG). The HOG is used as feature descriptor and it was firstly proposed by Dalal and Triggs [21]. It generalize the object in such a way that a feature of an object detected is close as possible to the similar human feature no matter what different condition it is being viewed. Another feature of HoG is, it is sensitive to the direction of gradients. Within one cell, the histogram of a given orientation is a simple sum [22] and Support Vector Machines (SVM) have been used effectively for classifying between two or more classes, or estimating a continuous variable using regression [23]. Once human is detected in the frame, the personal monitoring systems proceed to the second process which is the face recognition.

Process 2: Face recognition

In the face recognition process, there are two stages that need to be done. The first one is the face detection and the second stage is the face recognition. In this work, frontal face detection is done using the approach of Haar cascade or also known as Viola-Jones. Once face is detected in an image, a rectangle will be constructed at the face area in the video.

When face is detected, the second stage which is the face recognition will be done towards the face image. In this personal monitoring work, Principal Component Analysis (PCA) or known as eigenface is used for training the images. In order to carry on with the training image, preprocessing is firstly done in order to standardize the image being supplied to the face recognition system. The preprocessing is done by firstly converting the image into greyscale image. Histogram equalization is then performed

VOL. 10, NO 19, OCTOBER, 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

towards the greyscale image in order to have consistent brightness and contrast.

Having the pre-processing image ready, Eigenface or PCA can now be performed. In order to recognize someone face, a set of training image is needed. In this work, 30 photo for each person will be preprocessed at first. Supposed that there is $I_1, I_2, .. I_M$ training images represented by vector $\Gamma_1, \Gamma_2, ... \Gamma_M$ respectively. An average face vector will be computed as $\Psi = \frac{1}{M}\sum_{i=1}^{M}\Gamma_i$. The difference between the training images Γ_i and the average face vector Ψ is then computed as $\Phi_i = \Gamma_i - \Psi$. The covariance matrix C is then computed as $C = \frac{1}{M}\sum_{n=1}^{M}\Phi_n\Phi_n^T = AA^T$ where $A = [\Phi_1\ \Phi_2 ... \Phi_M]$. The eigenvector, u_i of AA^T is then computed in which only K eigenvector with regards to K largest eigenvalue will be kept. The normalized training face is then represented in the basis by each vector as in equation (1):

$$\Omega_i = \begin{bmatrix} w_1^i \\ w_2^i \\ \vdots \\ w_K^i \end{bmatrix},\ i = 1,2,..., M \qquad (1)$$

In other words, PCA will convert all the training images into a set of Eigenfaces which represents the differences between training images. An average face image will be obtained by computing the mean value of each pixel and then the first eigenface will be generated in which this is the most dominate face differences. The second eigenfaces will be generated as well as all the consecutive eigenfaces.

In order to recognize the person in the personal monitoring system, PCA calculation will be done, where it will try to match with the closes value of eigenface. This is achieved by computing Euclidean distance as in equation (2):

$$\left\|\Omega - \Omega^k\right\| = \sum_{i=1}^{K}\frac{1}{\lambda_i}\left(w_i - w_i^k\right)^2 \qquad (2)$$

If the value is out of the range of eigenface face value, the personal monitoring system will label the face as unknown.

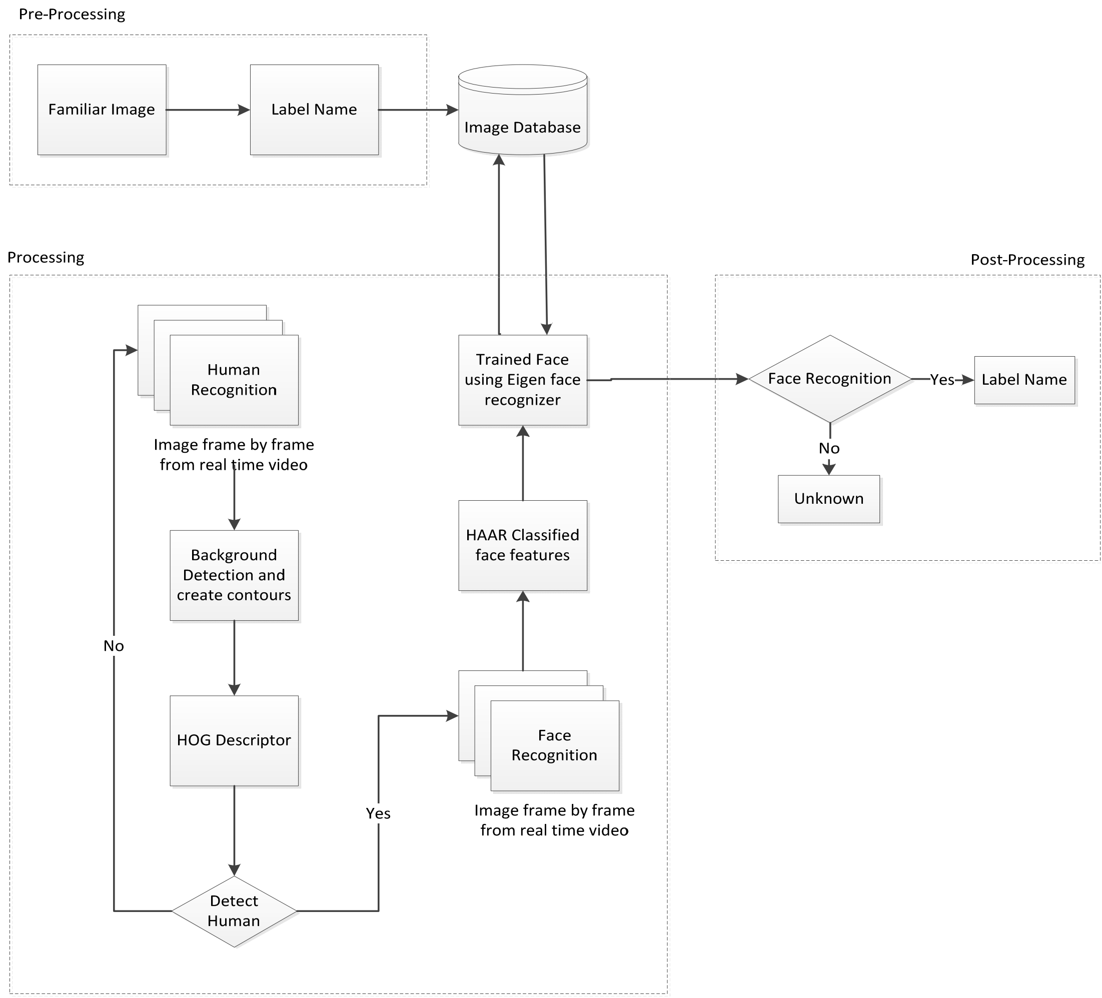


Figure-1. The overview of personal monitoring framework.

RESULTS AND DISCUSSIONS

A personal monitoring application is developed based on the proposed personal monitoring framework using C# programming language. Figure-2 illustrates the main window developed for personal monitoring application. There is a frame to display the real time image being captured. In this figure, there is only a background with no human detected. If the application detects human, the red box will be contracture at the detected human. The program then continues to detect faces constantly for each segmentation of frame. The face will be marked in blue box. If the person is known, the name will be displayed at the frame in yellow font color. At the same time, the familiar name will be displayed at the right side of the whole window, under the result of person present in the scene. Moreover, the system is able to count the number of people, where it will display the total number under the result of number of faces detected

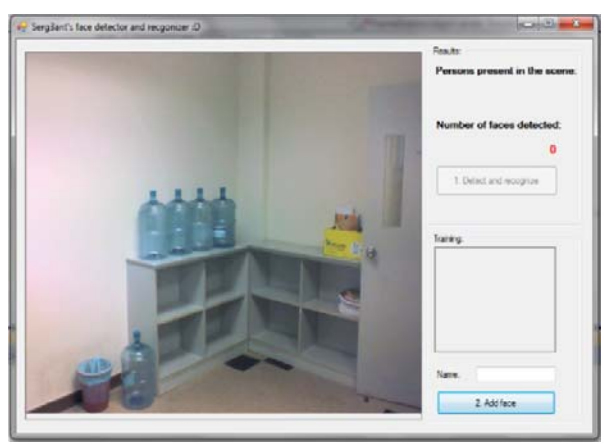


Figure-2. A window of personal monitoring system.

Figure-3 presents the image of face detection can be added in the database by clicking on the textbox labelled as Add Face. The faces that are added in the

VOL. 10, NO 19, OCTOBER, 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

database will be used for training purposes. While adding the faces in the database, the face will be shown simultaneously in the training box.

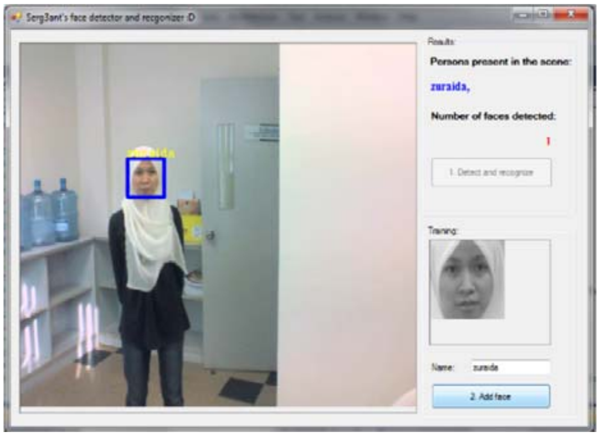


Figure-3. Added face with label into database.

Figure-4 illustrates human is detected where he is marked with red box. The face is marked with blue box. Since he is familiar, the name is appeared in the frame as well as at the right side of the window. The total number of person detected is also shown. Figure-5 illustrates two human are detected. Again, there are two red boxes constructed for each human detected as well as the names are appeared since both of them are familiar.

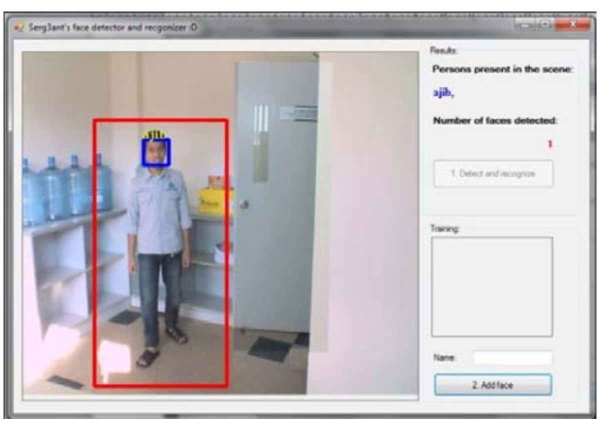


Figure-4. An example of human detection and face recognition for one person.

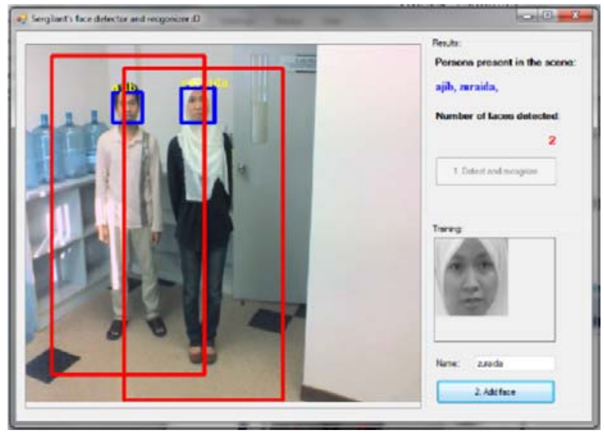


Figure-5. An example of human detection and face recognition for multiple person.

A standard Haar Cascase classifier perform well in human face detection, however, how far it accurately detect based on the training is unknown. Mostly the hypothesis state that the more the images of an individual are collected, the more accurate the human face recognition will be. Therefore, an experiment of testing this hypothesis is done as shown in Table-1. The experiment serves as the performance measure for the personal monitoring application developed using the proposed framework. The face database contains facial images of individuals with various facial expression such as sad, happy, surprised and etc. Figure 6 illustrates some of individual images in the database while Figure-7 illustrated different features images for an individual with each image resolution is 100x100 pixels.

Figure-6. Face database images.

VOL. 10, NO 19, OCTOBER, 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

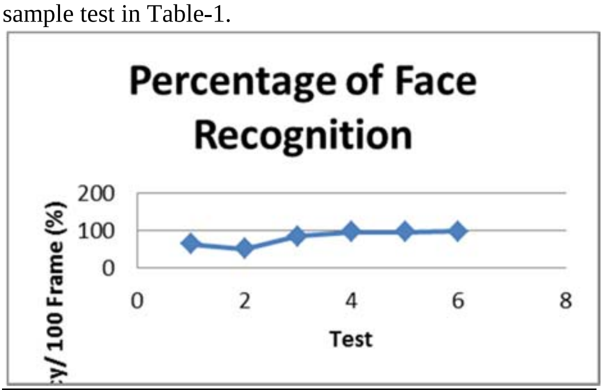
Figure-7. Different features images for an individual.

In this work, to perform the optimum face features detection, the experiment had been conducted by performing set of training face images. Table-1 represents number of individual tested with number of training images per person.

Table-1. Number of individual tested with number of training per person.

Test	**1**	**2**	**3**	**4**	**5**	**6**
No. of Individual	1	2	3	4	5	6
No. of training images/person	30	30	60	90	90	120

The results are collected base on either true or false recognition for each individual of human faces. The video camera will be recorded for 100 frames. Figure-8 shows the percentage of true faces that have been recognise by the video surveillance. Notice that the true recognition result for 30 training images per person for an individual only 64% then decrease to 52% for two individual tested. After that, the result percentage of true recognition increase along the increasing of number of individual tested and number of training images per person which is 85%, 97%, 897% and 98%.The result accept the hypothesis that have been stated before. Figure-8 represents the percentage of true faces recognition base on sample test in Table-1.



Test	**1**	**2**	**3**	**4**	**5**	**6**
Accuracy/100frame (%)	64	52	85	97	97	98

Figure-8. Percentage of face recognition for each test.

Based on the output from the personal monitoring system, the human detection integrated with face recognition in this system proves to work well. It must be noted that the work reported here is serves as the initial work for home monitoring. Hence, the system still has some limitation in recognizing human face because it can only work for frontal images. Therefore, if the image capture is not completely frontal image, the system may failed to perform face recognition properly.

FUTURE WORK

At the moment, the system is able to detect human in a pedestrian pose only. Therefore, the work will be extended to cover human with highly articulated poses such as bending, crouching and etc.

CONCLUSIONS

In the area of video surveillance, human detection is one of the major topics of computer vision research because it has very wide range of applications. This paper work on the personal monitoring where the existence of human is firstly detected. Once it is detected as human, then it will be match with the one keep in the database in order to determine either the human is recognized or not. This work serves as an initial study for home surveillance system.

The work consists of two major processes which are human existence detection followed by face recognition. For human existence detection, the human is detected using mixture of gradient which is the approach of background subtraction. The classification of human or non-human is done using the approach of texture-based method which is Histogram of Gradient. For face recognition, HAAR techniques are used to detect face while Eigenface recognizer is used to determine the human is recognized or not. The whole integrated processes are described in the overview of personal monitoring framework.

ACKNOWLEDGEMENT

A highly appreciation to Universiti Teknikal Malaysia Melaka (UTeM) for sponsor this research through FRGS and RAGS research grant (no. FRGS/2/2013/ICT07/FTMK/02/7/F00190 and RAGS/2013/FTMK/ICT02/01/B00039). Thank you to all friends and staff from Faculty of Information and Communication Technology (FTMK) for giving support to complete this project.

REFERENCES

[1] M. Sivarathinabala and S. Abirami. 2013. “An Intelligent Video Surveillance Framework for Remote Monitoring,” Int. J. Eng. Sci. Innov. Technol., vol. 2, no. 2, pp. 297-301.

[2] C. L. Devasena, R. Revathí, and M. Hemalatha, 2011. “Video Surveillance Systems – A Survey,” Int. J. Comput. Sci. Issues, vol. 8, no. 4, pp. 635–642.

[3] A. H. Nasution and S. Emmanuel, 2007. “Intelligent Video Surveillance for Monitoring Elderly in Home Environments,” Multimedia Signal Processing. MMSP. IEEE 9th Workshop on. pp. 203-206.

[4] Z. Zhou, W. Dai, J. Eggert, J. T. Giger, J. Keller, M. Rantz, and Z. He, 2009. “A real-time system for in-home activity monitoring of elders,” Engineering in Medicine and Biology Society. EMBC. Annual International Conference of the IEEE. pp. 6115–6118.

[5] A. M. Tabar, A. Keshavarz, and H. Aghajan, 2006. “Smart home care network using sensor fusion and distributed vision-based reasoning,” Proc. 4th ACM Int. Work. Video Surveill. Sens. networks - VSSN ’06, p. 145.

[6] M. Hofmann and M. Kaiser, 2011.“Event Detection in a Smart Home Environment using Viterbi Filtering and Graph Cuts in a 3D Voxel Occupancy Grid.,” in International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications.

[7] K. C. Fuerstenberg, D. T. Linzmeier, and K. C. J. Dietmayer, 2010. “Pedestrian Recognition and Tracking of Vehicles using a vehicle based Multilayer Laserscanner,” IEEE Trans. Intell. Transp. Syst., vol. 11, pp. 579–588.

[8] A. H. Nasution and S. Emmanuel, 2007. “Intelligent video surveillance for monitoring elderly in home environments,” IEEE 9th Int. Work. Multimed. Signal Process. MMSP - Proc., pp. 203-206.

[9] H. Nanda and L. Davis, 2002. “Probabilistic template based pedestrian detection in infrared videos,” Intell. Veh. Symp. IEEE, vol. 1, pp. 1-11.

[10] C. Premebida and U. Nunes, 2013. “Fusing LIDAR, camera and semantic information: A context-based approach for pedestrian detection,” Int. J. Rob. Res., vol. 32, no. Iv, pp. 371-384.

[11] C. Premebida, G. Monteiro, U. Nunes, and P. Peixoto, 2007. “A Lidar and vision-based approach for pedestrian and vehicle detection and tracking,” IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC, pp. 1044-1049.

[12] D. M. Gavrila and S. Munder, 2007. “Multi-cue pedestrian detection and tracking from a moving vehicle,” Int. J. Comput. Vis., vol. 73, no. 1, pp. 41–59.

[13] C. Wren and A. Azarbayejani, 1997.“Pfinder: Real-time tracking of the human body,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 780–785.

[14] I. Haritaoglu, D. Harwood, and L. S. Davis, 2000.“W4: real-time surveillance of people and their activities,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 22, no. 8. pp. 809–830.

[15] R. Polana, R. Nelson, and N. York, 1994. “Low Level Recognition of Human Motion,” Proc. 1994 IEEE Work. Motion Non-Rigid Articul. Objects. pp. 77-82.

[16] P. J. Burt, J. R. Bergen, R. Hingorani, R. Kolczynski, W. A. Lee, A. Leung, J. Lubin, and H. Shvayster, 1989. “Object tracking with a moving camera,” Visual Motion, 1989, Proceedings. Workshop on. pp. 2–12.

[17] R. Cutler and L. Davis, 1999. “Real-Time Periodic Motion Detection, Analysis, and Applications otherwise,” Comput. Vis. Pattern Recognit., vol. 218.

[18] A. Turner and A. Penn, 2002. “Encoding natural movement as an agent-based system: An investigation into human pedestrian behaviour in the built environment,” Environ. Plan. B Plan. Des, vol. 29, pp. 473-490.

[19] A. Broggi and I. Informazione, “Shape-based Pedestrian Detection.”

[20] A. Pentland, B. Moghaddam, and T. Starner, 1994. “2 A large face database 1 Introduction 3 General viewing geometries,” no. 245, pp. 2–8.

[21] N. Dalal, B. Triggs, and D. Europe, 2005. “Histograms of Oriented Gradients for Human Detection”.

ARPN Journal of Engineering and Applied Sciences

[22] C. Doersch and A. Efros, 2010. "Improving the HoG descriptor".

[23] T. Gandhi and M. M. Trivedi, 2008. "Image based estimation of pedestrian orientation for improving path prediction," IEEE Intell. Veh. Symp., pp. 506–511.