Empty Seat Revelation Using Face Detection And Correlation Matching

Sree Sharmila T
Department of IT
SSN College of Engineering
Chennai,India.
sreesharmilat@ssn.edu.in

Pavithra L K
Department of IT
SSN College of Engineering
Chennai, India.
pavithralk@ssn.edu.in

Abstract— **The main initiative of this paper is to explain the concept of empty seat revelation system and hence find the number of empty seats left unoccupied in a hall. The efficient empty seat revelation system is achieved using the combination of Viola-Jones algorithm and template based correlation matching. The proposed empty system proves to be highly efficient. This system supports crowd management in the form of updating the number of vacant seats periodically.**

Keywords— ***Crowd management, Empty seat, Face detection, Template-based correlation matching, video surveillance, Viola-Jones algorithm***

I. Introduction

Empty seat revelation system aims to detect the number of empty seats left unoccupied in a crowded hall. High resolution, high definition (HD) cameras are set at various corners of the hall and the video is captured continuously. The obtained video is split into frames and each of these frames is analyzed for human presence. Initially, the obtained frames are subjected to template based correlation matching[1],[2],[3] to identify the initial count of the seats present in the hall. Once this initial count of the seats is obtained, this count serves as the input to the further process. The frames are then further processed using Viola Jones face detection algorithm [4],[6],[12],[14] to identify the human presence[11],[36].Thus, from the obtained information the system can conclude the number of empty seats present in the hall. The empty seat count is displayed in a monitor held outside the hall and thus enabling people outside to easily know the presence of any empty seats available. The system helps in quickly organizing the crowd [5] and reduces unwanted waiting time. Thus the system plays a vital role in crowd management. This proposed work is compared with Adaboost [6], a boosting algorithm which is used to detect the human faces automatically [7],[8] and then the extracted human face is subjected to the Camshift algorithm [9],[10].

Fig. 1 depicts a brief idea of what the system actually does when it is implemented. The video is captured continuously with the help of various cameras fixed at different positions in the hall. Initial frames of the empty hall are subjected to template based correlation matching to obtain the initial count of the empty seats. Then, the sequence of frames obtained from the video are analyzed and checked for human presence. If humans are detected then the count of the same is obtained for all the frames. The empty seat count is updated at a periodic interval of 5 seconds. The count thus obtained is a cumulative value obtained from the processing of the frames during the span of those 5 seconds. If the human presence is detected then the empty seat count is updated otherwise the initial seat count remains unchanged.

II. Detailed Methodology

This paper proposes face detection using Viola-Jones algorithm, a well-known algorithm for detecting human presence. The Viola-Jones algorithm is a real-time face detection algorithm [4]. This algorithm is efficient in feature selection [12], and is a fast feature computation technique, faster than the Schneiderman-Kanade detector [13]. It scales only the features and not the image itself [14]. It is a generic algorithm that is used not only to detect faces but can also be trained to detect other objects of our choice [6]. This algorithm helps to identify the human presence and the count for the same[15]. This work also features yet another successful algorithm called template matching which uses the template to find the correlation coefficient , that is used to obtain the initial seat count in the hall. The time complexity of the correlation coefficient calculation is reduced by the Fast Fourier Transform (FFT) [22].

A. Template Based Correlation Matching

This paper proposes to calculate the initial number of seats present in the hall using the technique of template based correlation matching[16],[17],[18],[19],[20],[21]. Individual frames from the input video are processed using this technique to identify the number of empty seats. Images of empty seats are collected and used as template images for this process. Multiple windows of the frames are correlated towards the template image and the window with the most

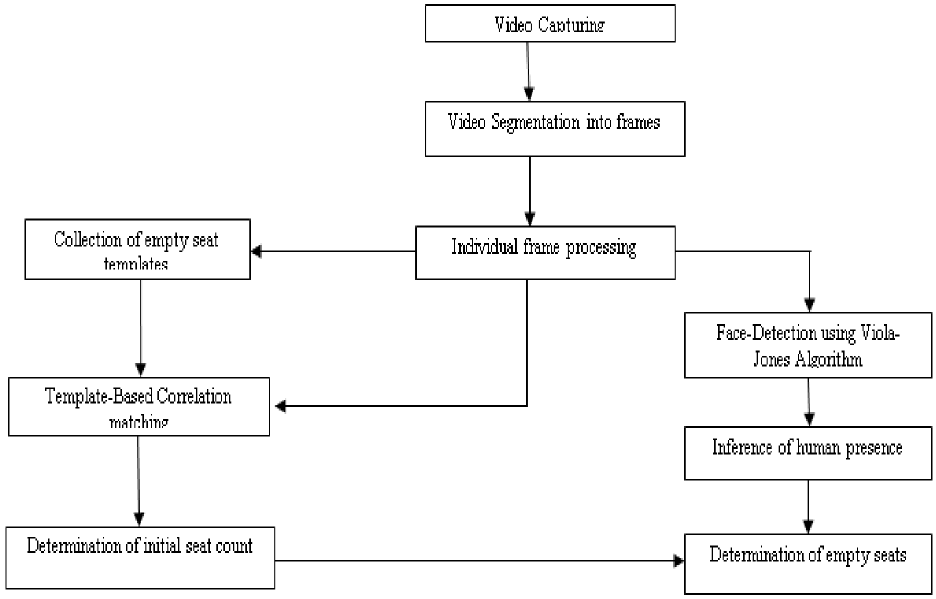

Fig. 1. Empty Seat Revelation System

similarity score is identified as the empty seat. Template matching is used to identify the presence of an object in a target image using the computation of the correlation coefficient, as this is invariant to linear intensity distortions. The correlation coefficient is calculated in the spatial domain, to make use of the redundancy in the computation of the correlation coefficient. Partial elimination techniques [23], use the property that the measures of the correlation coefficient increase monotonically when processed closer to the search location[24]. Figure 2 shows a template that is used to match the required target in the original input Fig 3. The number of highly matched template image is identified in the target image using the correlation coefficient, the identified template is marked using a bounding box.

Fig. 2. Template image

Fig. 3. Original input image

1) Computation Of Cross Correlation

Considering the reference image r size $m \times n$ pixels and a template image t of size $p \times q$ pixels, the correlation coefficient between the template image t and the reference image block $r_{i,j}$ where (i, j) the first block is defined as

$$P_{t,i,j} = \sum_{x=1}^{m} \sum_{y=1}^{n} \frac{S_t(x, y)}{\alpha_t} \frac{S_{i,j}(x, y)}{\alpha_{i,j}} \quad (1)$$

where, $S_t(x, y) = t(x, y) - \mu_t$,

$S_{i,j}(x, y) = r_{i,j}(x, y) - \mu_{i,j}$

and

$$\alpha_t = \sqrt{\sum_{x=1}^{m} \sum_{y=1}^{n} (t(x, y) - \mu_t)^2}$$

$$\alpha_{i,j} = \sqrt{\sum_{x=1}^{m} \sum_{y=1}^{n} (r_{i,j}(x, y) - \mu_{i,j})^2}$$

The summation represents the monotonic behavior of the partial symmetry value. The integral images of the target image can be pre-computed [25], [38]. This increases the speed of computation of the correlation coefficient value, to identify the matching of the template.

B. Viola Jones Algorithm For Face Detection

The Viola Jones object detection algorithm has high competitive detection rates [6]. Though the algorithm can be trained to detect multiple types of objects, it is mainly used in the field of face detection [26]. This algorithm, proposed in 2001, is highly robust and produces very low false detection rate and it is very efficient in real time face detection [27]. The algorithm basically involves four steps such as haar features, Integral Image and Cascade architecture. One of the steps involves Adaboost, a Meta learning algorithm that is used to automatically detect faces. But Adaboost as a single algorithm is not efficient; thereby it is always used in combination with the other algorithms to produce very high detection rates.

1) Haar Feature Detection

The feature detection is the first step of the face detection process [28], [29], [30], [31]. The input image is initially broken up into several 24×24 sub windows depending on the size and the resolution of the image. Each of the obtained sub windows are subjected to the haar feature test. There is an initial set of 1, 62,336+ features to be checked for in each sub window. Figure 4 shows the different types of haar features. These features can be further classified into edge, line and centre surround features[6]. The subtraction operation between the sum of intensities in the black area and the sum of intensity information in the white area give single value.

$$Value = \sum(\text{Pixels in the black area}) - \sum(\text{Pixels in the white area}) \quad (2)$$

As shown in the Fig. 4, the nose is detected using the haar feature as the nose bridge region is brighter than the eye. Similarly, Fig. 5 shows the other case, where to detect an eye, the haar features are evaluated as the eye region will be darker than the cheek.

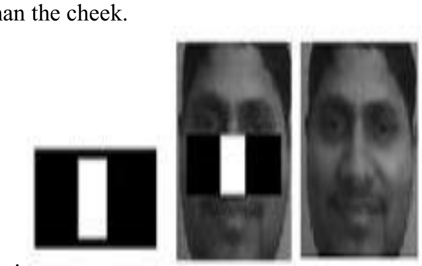

Fig. 4. Detection of nose using Haar features

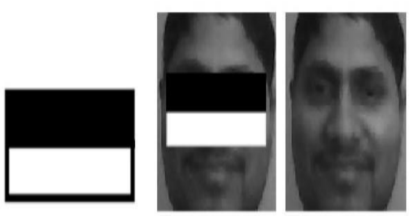

Fig. 5. Detection of eyes using haar features

In Fig. 6, A and B shows two rectangular haar features, C depicts three rectangular haar features and D shows four rectangular haar features, wherein the two rectangular haar features are widely used in the Viola Jones algorithm for the face detection. The output value obtained for each feature is evaluated in the next step using the integral image.

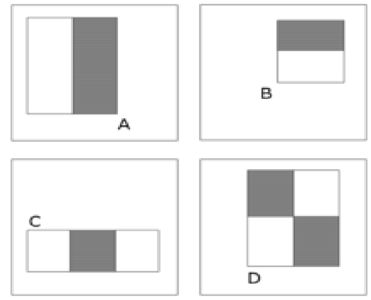

Fig. 6. Different types of rectangular Haar features

2) Integral Image (Or) Summed Area Table

Integral image is a fast feature evaluation technique [28], [29], [30], [31]. This is used to compute the value obtained from the haar feature detection step. In an integral image the value at pixel (x, y) is the sum of pixels above and to the left of (x, y). The Fig. 7 illustrates the integral image computation for the given input image.

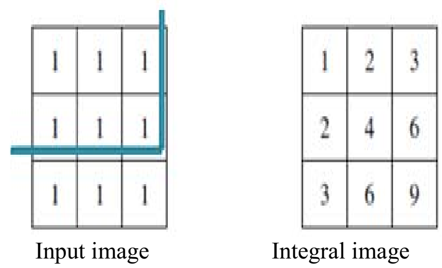

Fig. 7. Integral image computation

3) Adaboost Training

Adaboost is a meta learning face detection algorithm [28], [29], [30], [31], [33]. This is used as the third step in the Viola Jones face detection algorithm. Since there are more than 1,60,000 features to evaluate, it becomes a very tedious process while evaluating all the features for every single sub-window. Therefore, Adaboost is used here for performing the job of feature selection. Feature selection in this context basically means the process of choosing the appropriate features to be evaluated for each chosen 24×24 sub window. Adaboost is used to train the classifiers that use these features. It selects the best features from the evaluation results of the integral image. It builds the strong classifiers for each stage as a linear combination of the weak classifiers. Adaboost algorithm selects the best features from the set of 1,60,000 features which has the capability of separating the face and non face images efficiently.

4) Cascade Architecture

It is the final step in the face detection process, where the stages of classifiers are built. The each stage of this algorithm decides whether the sub window has face or not. At any stage of process, sub windows are discarded immediately when they have non face information in it [34].

Progress in the form of stages of classifiers proves to be fast, efficient and less cumbersome. Consider a case where 100 rectangular sub-windows pass through stage 1 of the classifiers. Suppose if only 50 sub-windows are considered to be positive i.e .there exists the presence of face they are selected for evaluation by stage 2. If the other 50 sub-windows are considered to be negative i.e., absence of face, then those sub windows are dropped and not considered for the evaluation by future stages. This methodology basically reduces the number of frames to be processed at each stage. As the stage progresses, the test for the presence of features will be more rigorous and at the same time higher, hence in this stage the features evaluated at that stage will also grow higher.

Fig. 8. Gray scale Image 1

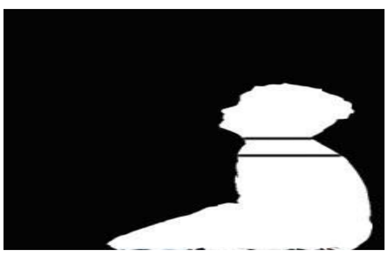

Fig. 9. Gray scale Image 2

C. Detection Of Human Head

The proposed techniques also use head shoulder detection [39] in addition to face detection. This head shoulder detection technique assumes the area of human head as H_H, which separates the skin region from the other regions [34] ,[35]. The human head with two different poses are shown in Fig. 8 and 9. The middle line separates the area of the head into two different regions. The front and back poses of the face symmetrically splits the two regions where as side poses of the face leads to non-symmetric splits. The left half, the right half and the whole head region are denoted by H_L, H_R, H_H respectively. The head poses are given below:

$$\begin{cases} 1 & \begin{cases} \left|\mathrm{H_L} - H_R\right| < T \\ \dfrac{\mathrm{H_L} + H_R}{H_H} < T \end{cases} \\ 0 & \begin{cases} \mathrm{H_H} = \left|\mathrm{H_L} - H_R\right| < T \\ \dfrac{\mathrm{H_L} + H_R}{H_H} \geq 0.5 \\ \dfrac{\mathrm{H_L} + H_R}{H_H} \left|H_L - H_R\right| \geq T \end{cases} \end{cases} \quad (3)$$

where, T is a threshold. H_H equal to one indicates that the camera faces the back side of the head. H_H equal to zero indicates that the camera focuses the front face. H_H greater than zero denotes that the left side of the head is focused by the camera. H_H less than zero denotes that the camera focuses right side pose of the head. The minimum distance between the H_L and H_R indicates that the camera is focusing the full head area either front or back. The $H_L + H_R H_H$ is greater than the 0.5 indicates that the front portion of the head is focused by the camera. On the other hand, the camera focuses the back portion of the head. The large distance between the H_L and H_R illustrates that the side poses of the head is focused by the camera. The value of $H_L - H_R H_H$ is greater than the zero for the left side head poses and less than zero for the right side poses. In order to improve the accuracy, this system introduced the angle of rotation θ which is used to measure the head with respect to the position of the shoulder.

The trace window for the face detection can be manually selected or can be done with the help of the Camshift algorithm [37]. With the help of the above mentioned techniques we have been able to effectively detect the presence of a human being seated in the hall. These techniques help to differentiate between the human being occupied seats and an empty seat that is left unoccupied or occluded with an object.

III. Experimental Results And Discussions

For every acquired image size of 1920×1080, the seat is about 116×100 pixels on an average. First, the video of an empty hall is captured and it is split into frames. One of the obtained frames is subjected to template based correlation matching to identify the initial count of the seats in the hall. The obtained value is used to set initial value of the empty seat count, after which the video samples of the people seated are obtained. From the video, the individual frames obtained are subjected to face detection algorithms for detecting the human presence. The Fig. 10 shows the output of the template based correlation matching. Then the initial count of the empty seats obtained as a result of correlation match. An exact matching of the pattern of interest cannot be obtained as it will be corrupted by noise, geometric distortion, and occlusion. Matching can also be used with an infinite number of transformations; hence a

TABLE 1 DETECTION RATE OF HUMAN FACES

Face detector Algorithm	Number of human faces given as input in the video (%)						
	7	*16*	*32*	*37*	*41*	*49*	*68*
Viola Jones	100	100	100	99.2	97.3	96.6	95.8
Adaboost and camshift	100	100	100	99.6	98.5	97.3	96.4
Proposed work	100	100	100	100	99.6	98.4	97.1

number of separated similar patterns can be joint using rubber links. If the image is rotated, scaled or distorted, an exact match is difficult to find, hence a number of transformations of the template image are used to identify different types of seats. The variations in appearance of the target object can also be handled by using various templates. The Fig. 11 shows the face detection from the video in which the people are seated on the chairs. The Viola Jones face detection algorithm is used to detect the presence of people. The count is thus obtained for each processed frame. The sequences of frames are processed for every 5 seconds and the count thus obtained is a cumulative count obtained from the processing of the frames during that span of 5 seconds. The proposed technique, a combination of template based correlation matching with Viola Jones algorithm has achieved 97.1% detection accuracy which is increased by 1.3% for performing only the Viola Jones algorithm. Table 1 shows detection rate of various input videos of the proposed technique.

Fig. 10. Detection of seats – template matching Input Videos

Fig. 11. Detection of human presence

IV. CONCLUSION

In this research work, it has been proposed to combine the template image based correlation matching algorithm and Viola Jones face detection algorithm to count the number of human present in a hall. An empty seat of the hall is identified by the correlation coefficient. The proposed method can handle detection of the empty seats using the Viola Jones face detection and automatic tracking. Thereby, this proposed work effectively supports its application in the field of video surveillance and seat allotment for a large crowded place. The future work lies in optimal techniques for devising; the same work can be implemented for low resolution cameras as well. This system can further be enhanced to spot out where the vacant seats lie exactly and how to reach the same.

ACKNOWLEDGEMENT

The authors would like to thank the Management of SSN Institutions for supporting the internal funding project to complete this research work successfully.

REFERENCES

[1] Scheiderman Henry, Takeo Kanade, "A statistical method for 3D object detection applied on faces and cars,"*in Proceedings of IEEE Computer Vision and Pattern Recognition Conference,*vol.1,2000.

[2] Brunelli Roberto, Tomaso Poggio, "Face recognition features versus template," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol.10, pp: 1042-1052, 1993.

[3] Lipton J Alan, Hironobu Fujiyoshi, Raju S. Patil, "Moving target classification and tracking from real-time video," in *Proceedings of IEEE Applications of Computer Vision,* 1998.

[4] Yi Qing Wang, "An Analysis of the Viola-Jones Face Detection Algorithm," Image Processing on Line, vol.4, pp: 128-148, 2014.

[5] A Davies, J Yin, S Velastin, " Crowd monitoring using image processing," *Journal of Electronics and Communication Engineering,* vol.7,pp:37-47,1995.

[6] Viola Paul, Michael Jones, "Rapid object detection using a boosted cascade of simple features," *Computer Vision and Pattern Recognition (CVPR)*, vol.1, 2001.

[7] Jesorsky Oliver, Klaus J Kirchberg, Robert W Frischholz, "Robust face detection using the Hausdorff distance," *International Conference on audio and video based biometric person authentication,* pp:90-95, 2001.

[8] Ahonen Timo, Abdenour Hadid, Matti Pietikainen, "Face recognition with local binary patterns," *in European conference on Computer Vision,* pp: 469-481, 2004.

[9] G.R Bradski, "Computer vision face tracking for use in a perceptual user interface,"*Journal on Intel Technology,* pp:214-219,1998.

[10] S. Kai and L. Shirong, "Combined algorithm with modified camshaft and Kalman filter for multi-object tracking," *Information and Control*, vol.8, no.1, pp: 178-180, 2009.

[11] Yahiaoui T,Meurie C,Khoudour L, Cabestaing F, "A people counting system based on dense and close stereovision," *Image and Signal Processing,* pp:59-66,2008.

[12] Viola Paul, Michael Jones, "Robust real-time face detection," *International Journal of Computer Vision,* vol.57, no.2, pp: 137-154, 2004.

[13] Michael Jones, Viola Paul, "Fast multi-view face detection," *Mitsubishi Electric Research Lab TR*, vol.3, no.14, 2003.

[14] Beymer David, "Person counting using stereo human motion," *in Proceedings Workshop on IEEE,* pp: 127-133, 2000.

[15] Wei Shou-Der,Shang Hong Lai, "Fast template matching based on normalized cross correlation with adaptive multi-level winner update," *in IEEE Transactions on Image Processing*, vol.17,no.11,pp:2227-2235,2008.

[16] Luigi Di, Stefano Mattoccia, "Fast template matching using bounded partial correlation," *Machine Vision and Applications,* vol.13, pp: 213-221, 2003.

[17] Milos Malinsky,et al. "Registration of FA and T1-weighted MRI data of healthy human brain based on template matching and normalized cross-correlation," *Journal of digital imaging,*vol.26,no.4,pp:774-785,2013.

[18] Lei Ming,Zhang Guang-jun, "A novel algorithm for a rotation invariant template matching," *Optoelectronics Letters,*vol.4,no.5,pp.379-483, 2008.

[19] A Manzar,T.A Cheema, I.M Qureshi, "New scale invariant template matching technique using hyperspace image representation, " *Pattern Analysis and Applications,*vol.12,no.3,pp:201-214,2009.

[20] Palleja, Tomas, et. al, "Implemenation of a roubust absolute virtual head mouse combining face detection, template matching and optival flow algorithms", *Telecommunications Systems,* vol-52(3), 2013, pp: 1479-1489.

[21] Dou, Yunguang, et al., "A fast template matching-based algorithm for railway bolts detection", *International Journal of Machine Learning and Cybernetics* 5(6), 2014,pp:835-844.

[22] Frigo, Matteo and Steven G. Johnson, "The design and implmentation of FFTW3", *in Proceddings of IEEE* 93(2)*,*2005, pp: 216-231.

[23] Mahmood, Arif and Sohaib Khan, "Exploiting transitivity of correlation for fast template matching", *IEEE Transaction on Image Processing* 19(8), 2010, pp: 2190-2200.

[24] Hel-Or .Y and Hel-Or .H, "Real time pattern matching using projection kernels", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol-27(9), 2005, pp:1430-1445.

[25] Schweitzer .H, Bell. J.W., Wc .F, "Very fast template macthing", 7th European Conferene on Computer Vision, In: Heyden A., Sparr G., Nielsen M., Johansen P (eds) Computer Vision –ECCV 2002. ECCV 2002. Lecture Notes in Computer Science, vol-2353, pp: 358-372, Springer, Berlin, Heidelberg.

[26] Lai K.C., Dennis Wong M. L., Islam S.Z., "A HW/SW Co-Design Implementation of Viola-Jones Algorithm for Driver Drowsiness Dectection In: Jung, HK., Kim J., Sahama T., Yang CH. (eds) Future Information Communication Technology and Applications. Lecture Notes in Electrical Engineering, vol-235,2013, pp:427-435. Springer, Dordrecht.

[27] Mena, Alexandra Paz, Margarita Bachiler Mayoral and Estela Diaz-Lope, "Comparative Study of the features used by algorithms based on viola and jones face detection algorithm" , International work – conference on the Interplay between Natural and Artifcial Computation. pp: 175-183, Springer, Cham 2015.

[28] Sri-Kaushik Pavani, David Delgado-Gamez, Alejandro F. Frangi, "Fast training procedure for Viola- Jones type object detectors using Laplacian clutter models", *Pattern Analysis and Applications*, vol-17 (2), 2014, pp: 441-449.

[29] Castrillon, Modesto, et al, "A comaparison of face and facial feature dectors based on the Viola-Jones general object detection framework", *Machine Vision and Applications,* vol-22(3), 2011, pp:481-494.

[30] Xu, Zhenchao, et. al, "Improving dector of Viola and Jones through SVM", *Computer Vision- ACCV 2010 Workshops,* pp: 64-73, Springer Berlin Heidelberg, 2011.

[31] Anisetti, Marco., "Fast and robust face detection", Multimedia Techniques for Device and Ambient Intelligence, pp: 43-72. Springer US, 2009.

[32] Crow. F, "Summed area tables for texture mapping", *In proceedings of SIGGRAPH*, vol -18(3), 1984, pp:207-212.

[33] E. Osuna, R. Freund , F. Girosi, "Training support vector machines: an application to face detection", *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1997, pp: 130-136.

[34] F. Fleuret, D. Geman, "Coarse-to-fine face detection", International Journal of Computer Vision, vol-41(1-2), 2001, pp:85-107.

[35] H.Wuand, X. Zheng, "Improved and efficient object tracking algorithm based on Camshift", *Computer Engineering and Applications*, vol-45(7), 2009, pp:178-180.

[36] S. Karungaru , M. Fukumi, N. Akamatsu , "Detection of human faces in visual scene", *In Proceedings of the Seventh Australian and New Zealand Intelligent Information System Conference,* 2001, pp:165-170.

[37] B. Liu, H.Yao, W. Gao. "Real –time human face tracking in colour images", *Chinese Journal of Computers*, vol-6, 1998, pp:527-532.

[38] Mahmood, Arif and Sohaib Khan, "Correlation –coefficient based fast template matching through partial elimination", *IEEE Transaction on Image Processing* , vol-21(4), 2012, pp:2099-2108.

[39] Li Chunyan, Zhu Yulian, Xue Zhimei, "Detecting human head and shoulder trajectory in a smart classroom", *International Journal of Smart Home*, vol-8(4),2014, pp:293-302.