Detection of Unauthorized Human Entity in Surveillance Video

D Radha, J Amudha, P Ramyasree, Ranju Ravindran, Shalini Snehansh
Department of Computer Science and Engineering,
Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Bangalore, India.
{d_radha, j_amudha}@blr.amrita.edu
{ramyasree.510, ranjuchakkudu, shalini.snehansh}@gmail.com

Abstract--**With the ever growing need for video surveillance in various fields, it has become very important to automate the entire process in order to save time, cost and achieve accuracy. In this paper we propose a novel and rapid approach to detect unauthorized human entity for the video surveillance system. The approach is based on bottom-up visual attention model using extended Itti Koch saliency model. Our approach includes three modules- Key frame extraction module, Visual attention model module, Human detection module. This approach permits detection and separation of the unauthorized human entity with higher accuracy than the existing Itti Koch saliency model.**

Keywords—Video surveillance, Histogram, Key frame extraction, Visual Attention Model, Saliency map, Connected component, Aspect ratio.

I. INTRODUCTION

The importance of video surveillance has increased manifolds in the present scenario where there is a constant threat to security. Video surveillance is nowadays used in many fields for monitoring and security purposes. The various fields where surveillance videos are being used extensively are military, shops and banks for detection and prevention of robbery, detection of unauthorized entry in any institute, at airports and railway stations to monitor any suspicious activity, tracking of terrorist activities by intelligence agencies. Millions of cameras are deployed all over, producing huge amount of information. This information has to be put under extensive analysis and scrutiny. Doing so manually is very tedious, expensive, time consuming and requires huge manpower. Therefore in recent years automation of video surveillance has become an extremely active research area.

The need for advanced video surveillance systems has inspired progress in many important areas of science and technology including traffic monitoring, transport networks, traffic flow analysis, understanding of human activity, home nursing, monitoring of endangered species, observation of people and vehicles in crowded area, detection of unauthorized entity, abandoned object detection and many more[1].

Here we propose a model which automatically detects any unauthorized entry in an organization. We have restricted our work to detection of unauthorized entity in school grounds, but the same concept can be applied for any organization. For our project the authorized entity and unauthorized entity are distinguished by the dress code.

The techniques like histogram difference for key frame extraction, visual attention model for salient region detection, 4-connected component method for redundant salient region detection and removal, aspect ratio estimation for human identification etc. are used for automatic detection of unauthorized human entity in a surveillance video.

Gaussian Mixture Model makes use of Decision rule to classify pixels into background or moving objects categories and then Markov regularization of detection is done to smooth object detection. Though this procedure provides high accuracy, it is highly complex and is not able to eliminate shadow [2]. The Approximate Median Method takes the median value of a set number of previous frames to construct a back plate model and then compares the pixels with other frames. This procedure takes long processing time and is unable to eliminate shadow [3]. Kehuang Li and Yuhong Yang proposed a model called sigma delta model for object detection where global background variance is used as threshold for detecting background points and local intensity variance is used to detect foreground pixels. Though this model works better for illumination changes, it is quite sensitive to noise [4]. There are many other similar models which have been proposed earlier like the Frame Difference method [5] and Single Gaussian method [6]. Another model uses edge detection to segment moving object where background is reconstructed from image sequences and edge detection eliminate the disturbance of shadow and uses Gaussian filter to eliminate noise[7].Most of the models which have been proposed so far rely on background subtraction as the first step. Background subtraction techniques are generally quite sensitive to global illumination changes and environmental noise and often tend to include complex computations [8]. Another approach called salient motion detection which is a complementary approach to background subtraction assumes that a scene will have many different types of motion, of which some types are of interest from a surveillance perspective.This method is sensitive to noise and is computationally complex [9].

In this paper, we propose a model of unauthorized human entity detection using extended Itti Koch model [10], which works on the basis of features like colour, intensity and orientation of an image. This method is not only robust to noise but also computationally simple. A video is given as input to the system. The first step in our algorithm is extraction of key frame which is done using histogram key frame extraction technique [11]. Next, the saliency map is generated for the key frame using Itti Koch model based on features like color, intensity and orientation [10]. Later human detection is done from the salient objects found in the previous step using aspect ratio analysis. The rest of the paper is organized as follows: Section II describes the system model and algorithm in detail; Section III gives the performance evaluation and experimental results; Section IV gives the conclusion and future scope in the work.

II. DETECTION OF UNAUTHORIZED HUMAN ENTITY SYSTEM DESIGN AND IMPLEMENTATION

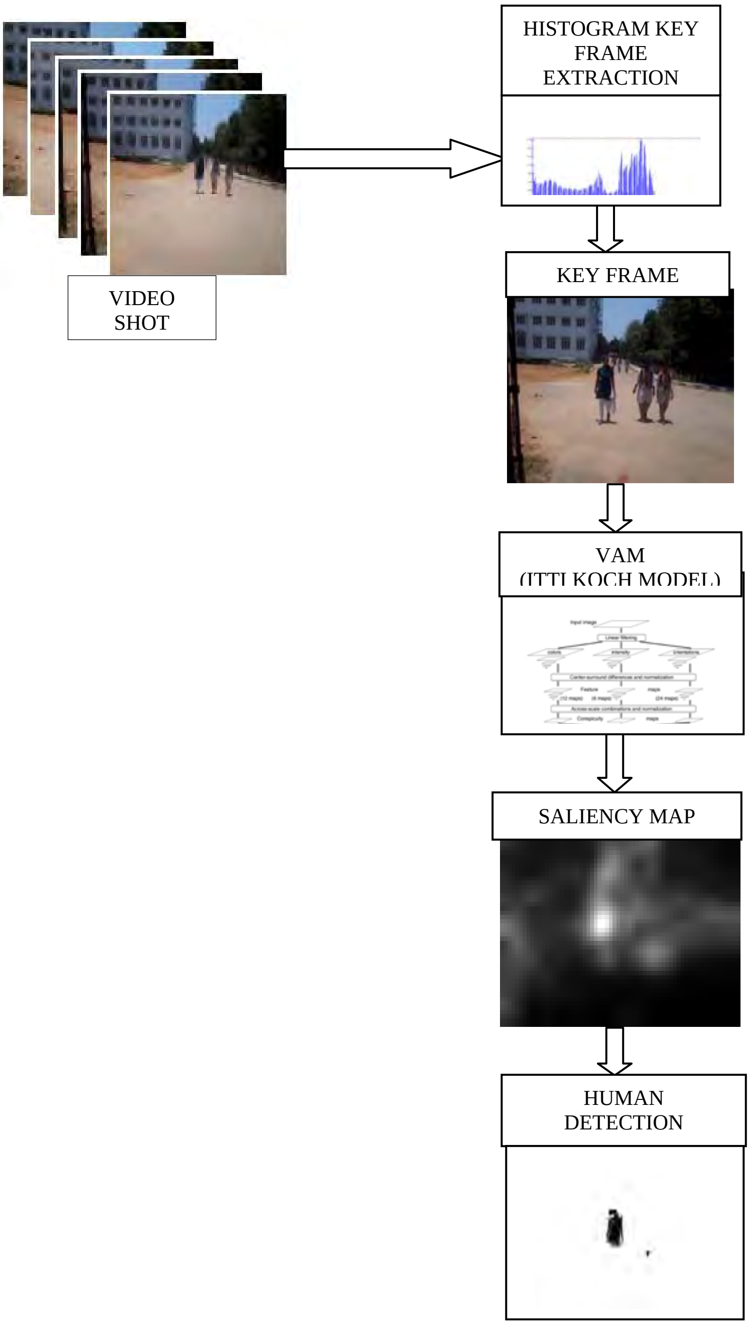


Fig. 1: Detection of unauthorized Human entity System Model

Fig. 1 describes the flow of the system. There are three important modules in the system. The first module is the key frame extraction module which generates the key frame. The key frame is passed on to the visual attention model module which generates a saliency map. The saliency map is given as input to the human detection module which finally generates the output i.e. the unauthorized entry. The 3 modules are explained in detail below.

A. Key Frame Extraction

A video can be discretized to finite still images. Each still image is called a “frame”, which is the basic unit of the video. The image sequence is naturally indexed by the frame number. All the frames in one video have same size and the time between each two frames is equal, typically 1/25 or 1/30 seconds [11]. A frame which can summarize the content of the video shot is considered as key frame. There may be more than one key frame depending on the content of shots. Appropriate methods can be applied to consider one among them. Key frame extraction refers to finding a set of salient frames taken from a full length video that represent the visual content in the video efficiently. The proposed key frame extraction is as follows-

Step 1: Frames ($f_{1,}$ f_2...f_n) are extracted from the given input video.

Step 2: For each consecutive frames, histogram difference (h_1, h_2,…h_{n-1}) is calculated.

Step 3: Mean and standard deviation for all the histogram differences are calculated.

Step 4: Threshold is computed using the formula

Threshold T_h=mean + standard deviation*α [12]

where α is a coefficient which is equated to 4.

Step 5: The frame f_i with histogram difference h_i greater than Threshold T_h will be selected as key frame.

Step 6: If none of the histogram difference is greater than threshold, the frame with highest difference is selected as key frame.

Step 7: Each key frame is converted to .jpg format.

B. Visual Attention Model

The key frame extracted is given to the Itti Koch Model which gives the saliency map of the frame. Low-level vision features (color, orientation and intensity) are extracted from the original color image at several spatial scales created using Gaussian pyramids which consist of progressively low-pass filtering and sub-sampling the input image [13]. The intensity channel(I) is obtained using(1) is used to create a Gaussian pyramid with a scale ranging from [0..8].Four broadly tuned color channels for red(R),green(G),blue(B),yellow(Y) are created using (2),(3),(4),(5). The orientation channels are obtained from I using oriented Gabor pyramids O(σ,θ) where σ ∈[0..5] is the scale, and θ {0 ,45 ,90 ,135 } is the preferred orientation[14].

$$I = \frac{r + g + b}{3} \qquad (1)$$

$$R = \frac{r - (g + b)}{2} \qquad (2)$$

$$G = \frac{g - (r + b)}{2} \qquad (3)$$

$$B = \frac{b - (r + g)}{2} \qquad (4)$$

$$Y = r + g - 2(|r - g| + b) \qquad (5)$$

For each of the channels, center-surround ‘‘feature maps’’ are constructed and normalized. Center surround is implemented in the model as the difference between fine (c) and coarse (s) scales. The center is a pixel at scale c =2,3,4 and the surround is the corresponding pixel at scale s = c + δ where δ =3,4[10]. Six set of intensity feature maps are created by absolute center surround difference (6). For the color channels, the feature map is based on the idea of chromatic opponency for red/green and blue/yellow such as in the human primary visual cortex [15]. Two maps are created from the new color channels by first computing(red-green) at the center, then subtracting (green-red) from the surround and finally outputting the absolute value. Accordingly maps RG(c,s) are created in the model to simultaneously account for red/green and green/red double opponency and BY(c,s) for blue/yellow and yellow/blue. double opponency(7),(8)[14].12 set of color feature maps are created. A total of 24 maps are created for the orientation (6 maps for each angle) (9).

$$I(c,s) = |I(c) \ominus I(s)| \qquad (6)$$

$$RG(c,s) = \left|\big(R(c) - G(c)\big) \ominus (G(s) - R(s))\right| \quad (7)$$

$$BY(c,s) = \left|\big(B(c) - Y(c)\big) \ominus \big(Y(s) - B(s)\big)\right| \quad (8)$$

$$O(c,s,\theta) = |O(c,\theta) \ominus O(s,\theta)| \qquad (9)$$

Feature maps of each feature are linearly combined to get three conspicuity maps, intensity $\bar{I}$, color $\bar{C}$, and orientation $\bar{O}$, at the saliency map's scale (σ =4).These maps are computed through normalization and across scale addition(⊕),where each map is reduced to scale four and added point-by-point[14].

To compute the orientation conspicuity map, four intermediary maps are created by first combining the six feature maps. These intermediary maps are then combined into a single orientation conspicuity map [14].

The conspicuity maps related to each feature are linearly combined to compute the saliency map. The final input S into the saliency map can be determined by averaging the three normalized conspicuity maps(10).The N(.) represents the non-linear Normalization operator.

$$S = \frac{1}{3}[N(\bar{I}) + N(\bar{C}) + N(\bar{O})] \qquad (10)$$

From the saliency map, the most attention regions are identified in the order of decreasing saliency [14]. To detect the most salient location and direct attention towards it, a Winner Take All (WTA) network is implemented. After the currently attended location is suppressed, attention is directed to the next most salient location in the image and repeating this process generates attentional scan paths. Such inhibitory tagging of recently attended locations is called Inhibition of Return (IOR).

C. Human Detection System

The saliency map generated by Itti Koch model may include any salient objects that are prompted from the frame. The salient entity which is human will be detected by this module. Here we are considering the school ground scenario where the background scenes contain road, vehicles, buildings and other inert objects. In order to remove background objects from the scene we are performing the aspect ratio analysis. The procedure for this analysis is as follows: The saliency is plotted over actual image. Some of the redundant salient regions are removed using the 4-connected graph.eg:-roads, buildings etc. Perimeter pixel coordinates(x, y) of remaining salient regions (R_1,R_2…$R_{n)}$ are extracted and aspect ratio of each region R_i (11)is calculated using the formula

$$\text{Aspect\_Ratio}(R_i) = \frac{\Delta y}{\Delta x} \qquad (11)$$

Δy = Difference between two y extremes for i^{th} region in the image

Δx = Difference between two x extremes for i^{th} region in the image

Since we are trying to detect the humans from the scenes, the objects whose aspect ratios are at the higher extreme end are ignored. Threshold factor T_h(12) is calculated to satisfy this condition-

$$T_h = \frac{\sum \text{Aspect\_Ratio}(R_i)}{N} \qquad (12)$$

The Aspect_Ratios(R_i) that exceed T_h are considered as most salient in our system. The regions containing Aspect_Ratios that cross this threshold T_h highlighted in the actual image. The threshold factor might change depending on the scenes which are being considered. Here we are taking the school ground scenes; hence we ignore the Aspect_ratios which are less than average Aspect_ratio.

Some of the experimental results are shown in below.

.

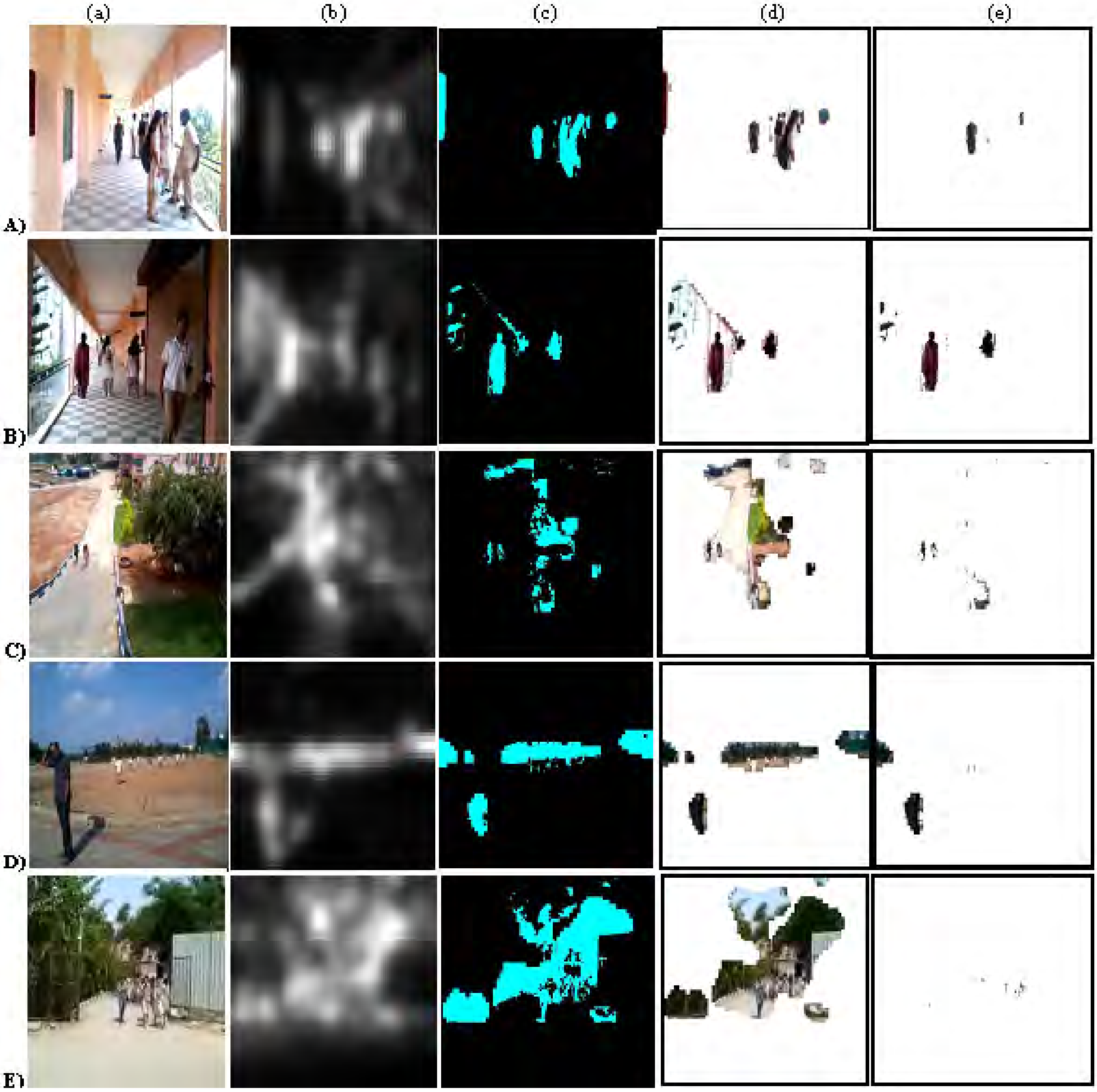


Fig 2 (a) Key frame extracted; (b) Saliency map using Itti Koch model; (c) Objects extracted after 4 connected algorithm; (d) Objects extracted using Itti Koch model (e) Detection and extraction of distinct entity using the proposed model based on aspect ratio analysis for sample images A-E.

Fig. 2.A and 2.B are indoor successful cases while Fig. 2.C, 2.D and 2.E are outdoor successful, outdoor partial successful and outdoor unsuccessful cases respectively.

III. RESULTS AND ANALYSIS

The video sequences used for testing is sampled at a resolution 320 x 240. The typical length of video sequences is around 200 frames. Real time videos have been taken for both indoor and outdoor case. Indoor case includes windows, doors, walls pillars etc. along with humans which should not be detected. Outdoor case includes trees, vehicles, gate, road, buildings along with humans.

The result from the video sequences has been divided into 3 categories:- successful, partially successful and unsuccessful. Some of the experimental results are shown above. In Fig. 2, A and B are the indoor case whereas Fig. C, D, E are the outdoor cases.

The analysis has been done in two ways:- a) On the basis of aspect ratio, b) On the basis of overall performance of the system .

A. Analysis Based On Aspect Ratio

Here the aspect ratio of some cases has been taken and analyzed to find the approximate range for human's aspect ratio. On the basis of this analysis the following results have been obtained. Table I contains the ratios calculate from all the successful cases where the human aspect ratio exceeds the threshold value T_h.

TABLE I. ASPECT RATIO DETAILS FOR SUCCESSFUL CASES

SAMPLE NO.	AVERAGE ASPECT RATIO (Th)	HUMAN'S ASPECT RATIO	HIGHEST ASPECT RATIO
1	1.1335	3.5714	4
2	1.1966	1.3590	3
3	0.9101	1.0506	2.5000
4	0.8796	1.6164	2
5	1.0906	1.4545	4.3333
6	2.0292	2.3750	8.4444
7	2.4519	2.6786	8
8	1.1390	1.7241	3.3125
9	1.2863	2	3.1739
10	1.0466	1.2551	3.5000
11	1.1404	2.5345	Infinity
12	1.1045	3	5

Fig. 3(a) is the Itti Koch output of the video sample 1 of the successful case. Here the human aspect ratio is 3.5714, average aspect ratio of all the objects is 1.1335 and the highest aspect ratio is 4. As the human aspect ratio is more than the average aspect ratio, it is detected and shown in Fig. 2.C.

It has been observed that in some of the cases the proposed algorithm has found other objects along with distinct human entities present in the scene as they exceed the threshold value T_h. Table II has listed three such cases.

TABLE II. ASPECT RATIO DETAILS FOR PARTIALLY SUCCESSFUL CASES

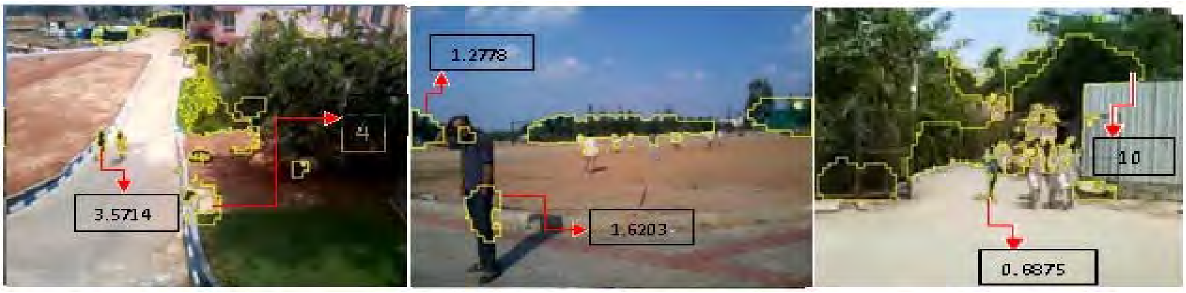
SAMPLE NO.	AVERAGE ASPECT RATIO (T_h)	HUMAN'S ASPECT RATIO (h-human, o-other objects)	HIGHEST ASPECT RATIO
1	1.7067	1.6203 (h)	2
		1.2778 (o)	
2	0.6004	1.8333 (h)	4
		1.1685 (o)	
3	1.3338	1.1288 (h)	3
		1.5000 (o)	

Fig. 3(b) is the Itti Koch output of the video sample 1 of the partially successful case. Here the human aspect ratio is 1.6203 and the other object ratio is 1.2778. The average aspect ratio is 0.6004. As both the human and the object's aspect ratio is greater than average, both are detected and shown in Fig. 2.D.

A few cases have been found where the Itti Koch has identified the distinct entity but it was ignored in the later stages due to the condition enforced in the aspect ratio analysis in the form of threshold factor. The aspect ratio details of these cases have been enlisted in Table III.

TABLE III. ASPECT RATIO DETAILS FOR UNSUCCESSFUL CASES.

SAMPLE NO.	AVERAGE ASPECT RATIO (T_h)	HUMAN'S ASPECT RATIO	HIGHEST ASPECT RATIO
1	1.4501	0.6875	10
2	2.2067	1.5439	4
3	1.7792	1	7



(a) (b) (c)

Fig. 3. (a) Aspect ratio of Sample No. 1 in Table I, (b)Aspect ratio of Sample No. 1 in Table II, (c)Aspect ratio of Sample No.1 in Table III.

Fig. 3.(c) is the Itti Koch output of the video sample 1 of the unsuccessful case. Here the human aspect ratio is 0.6875 and the average aspect ratio is 1.4501. As the human aspect ratio is below the average aspect ratio, the human is ignored in the final output and shown in Fig. 2.E. The highest aspect ratio here is 10.

From the above analysis, it can been concluded that most of the human ratios calculated from the recorded scenes lie in the range of (1,4) if the aspect ratio is calculated as per the above mentioned formula. These ratios highly depend on salient regions obtained from Itti Koch approach. Therefore, from the above results the proposed model is successful in most of the cases.

The Fig.s 3(a), 3(b) and 3(c) are outdoor images. As we can see that although the shadows of humans and other objects eg: trees in Fig..3(c) are present in the key frames extracted through the histogram key frame extraction technique these are ignored or are not detected in our final output due to the application of Itti Koch model. This is a huge advantage as other human detection algorithms are not able to eliminate shadow in the image. Some techniques like blob analysis apply additional methods for shadow elimination. Thus our proposed system has lesser add-ons compared to other systems and qualifies itself as a much simpler procedure for its use in a real time system.

B. Overall System Performance

A total of 70 test cases were taken i.e. 70 videos for analysis. Out of 70 cases:- 52 cases are tested as successful, i.e. all the distinct human entities are detected successfully; 11 cases are tested as partially successful i.e. distinct human entities are detected along with certain other objects; 7 cases are tested as unsuccessful, i.e. distinct human entity is not detected. In few cases the system considers objects with higher intensity and aspect ratio similar to humans, also as human. These are considered as unsuccessful cases. Table IV gives the confusion matrix. Each row represents the instances in a predicted class while each column represents the instances in an actual class. In 63 cases, distinct human entity has been detected. Apart from distinct human entity, there are 18 cases where some other entities are also detected. The overall accuracy achieved is 82.14% which is the ratio of correct classifications to all the classifications. Precision achieved is 90% which is the correct classifications penalized by the number of incorrect classifications. Recall is 78% which is the number of correct classifications penalized by the number of missed items. F-measure is found out to be 0.83. The system is found to be more robust to shadow elimination, environmental noise and illumination changes than other techniques. It is found to be quite efficient in detecting distinct entity in complex natural scene; both indoor and outdoor.

TABLE IV. CONFUSION MATRIX

	DISTINCT HUMAN ENTITY	OTHERS
DETECTED	63 (true positive)	18 (false negative)
UNDETECTED	7 (false positive)	52 (true negative)

IV. CONCLUSION

In this paper, a robust distinct entity detection system has been proposed. We have extended the Itti Koch model using aspect ratio analysis and have applied it for our application. The Itti Koch model detects the distinct human entity along with many other objects. A big advantage of using this approach is that the shadow is ignored. It overcomes the problem of detection of shadow which is a disadvantage in many other approaches. Once the human entity is detected using the Itti Koch model, aspect ratio analysis is applied to eliminate objects other than the distinct human entity which were also detected in the Itti Koch model. The system has been tested on many real time videos taken from school grounds. The analysis of result derived from these videos strongly suggests that our system is quite efficient in detecting any distinct human entity in terms of dress code in the video. The same concept can be applied to detect any unauthorized entry in any organization. In future, further work can be done for better distinction of objects having aspect ratio similar to humans. Hardware and software part can also be included where automatic alarm will be generated whenever any distinct unauthorized entity is detected.

REFERENCES

[1] Shih-Chia Huang, "*An Advanced Motion Detection Algorithm with Video Quality Analysis for Video Surveillance Systems*", IEEE Transactions On Circuits And Systems For Video Technology, Vol. 21, No. 1, January 2011.

[2] Lionel Carminati, Jenny Benois-Pineau, "*Gaussian Mixture Classification For Moving Object Detection In Video Surveillance Environment*", 2005.

[3] E.Komagal ,Arthy Vinodhini, Archana and Bricilla, "*Real time Background Subtraction Techniques for Detection of Moving Objects in Video Surveillance System*", 2008.

[4] Kehuang Li, Yuhong Yang,"*A Method for Background Modeling and Moving Object Detection in Video Surveillance*", 4th International Congress on Image and Signal Processing,2011.

[5] Xiaoshi Zheng, Na Li, Huimin Wu, Yanling Zhao, "*An automatic moving object detection algorithm for video surveillance application*" , May, 2009.

[6] Henan Guo, Yanchun Liang, Zhezhou Yu, Zhen Liu, "*Implementation and Analysis of Moving Objects Detection in Video Surveillance*", June, 2010.

[7] Ping Wang,"*Moving Object Segmentation Algorithm Based on Edge Detection*", December 2010.

[8] Sen-Ching S. Cheung and Chandrika Kamath,"*Robust techniques for background subtraction in urban traffic video*".

[9] Arun Hampapur, Lisa Brown, Jonathan Connell, Ahmet Ekin, Norman Haas, Max Lu, Hans Merkl, Sharath Pankanti, Andrew Senior, Chiao-Fe Shu, and Ying Li Tian, "*Smart Surveillance: Applications, Technologies and Implications*", IEEE Signal Processing Magazine, March 2005.

[10] Laurent Itti, Christof Koch, Ernst Niebur, "*A Model of Saliency-Based Visual Attention for Rapid Scene Analysis*", November, 1998.

[11] Haiyan Xie,"Key Frame Segmentation in Video Sequences – Applied to Reconstruction of 3D Scene", MS Thesis, University of Kalmar.

[12] Prajesh V. Kathiriya, Dhaval S. Pipalia, Gaurav B, Vasani, Alpesh J. Thesiya, Devendra J. Varanva," X^2 *(Chi-Square) Based Shot Boundary Detection and Key Frame Extraction for Video*", International Journal Of Engineering And Science2278-4721,Vol.2,January 2013.

[13] Laurent Itti, Christof Koch,"*A saliency-based search mechanism for overt and covert shifts of visual attention*", Vision Research 40 (2000) 1489–1506, 1999.

[14] Amudha.J, Soman.K.P,Kiran.Y," *Feature Selection in Top-Down Visual Attention Model using WEKA*", International Journal of Computer Applications Volume 24,(0975 – 8887),No.4, June 2011.

[15] Nathan Funk,"*Implementation of a Visual Attention Model*", April 14, 2004.