Person Detection and Tracking System for Visual Surveillance

Sreepathi B[1] & Halaharvi keerthi[2]

Abstract: **Detecting moving objects in video sequences is very important in visual surveillance. This describes a method for accurately tracking persons in indoor surveillance video stream obtained from a static camera with difficult scene properties including illumination changes and solves the major occlusion problem. Simple image processing with frame differentiation method is applied to identify multiple human motions. Firstly, a crowd is segmented by frame-difference technique, followed by morphological processing and region growing. Detecting and tracking multiple moving people in a complex environment with indoor surveillance video stream obtained from a static camera. The background subtraction method is to use the difference method of the current image and background image to detect moving objects, with simple algorithm, but very sensitive to the changes in the external environment. The effectiveness of the proposed method is demonstrated with experiments in an indoor environment.**

Keywords: **Video Surveillance, Motion detection, Person Tracking, Background Subtraction, Morphological Processing, Occlusion handling.**

1. INTRODUCTION

Video analysis and video surveillance are active areas of research. Video surveillance systems have long been in use to monitor security sensitive areas. In video surveillance system to provide better solution for security and monitoring problems new technologies are introduced. Visual Surveillance in dynamic scenes attempt to detect, recognize and track certain object from image sequences and more generally to understand the human or any object behavior.

The motivation of this paper is to show how Simple image processing with frame differentiation method is applied to identify multiple human motions.

The Video surveillance is increasing significance approach as organizations seek to safe guard physical and capital assets. At the same time, the necessity to observe more people, places, and things coupled with a desire to pull out more useful information from video data is motivating new demands for scalability, capabilities, and capacity. These demands are exceeding the facilities of traditional analog video surveillance approaches. Providentially, digital video surveillance solutions derived from different data mining techniques are providing new ways of collecting, analyzing, and recording colossal amounts of video data. Motion detection systems will not only be monitoring the areas of interest but will also keep an active lookout for any motion being produced.

In Automated video surveillance applications computer vision system is designed to monitor the movements in an area, identify the moving objects and report any doubtful situation. The system needs to discriminate between natural entities and humans, which require a good object tracking system. There are three key steps in video analysis: detection of interesting moving objects, tracking of such objects from frame to frame, and analysis of object tracks to recognize their behavior.

Human motion analysis helps in solving problems in indoor surveillance applications. Moving Objects Detection and tracking are widely used low-level tasks in many computer vision applications, like surveillance, monitoring, robot technology, gesture recognition, object recognition etc. Many approaches have been proposed for moving object detection and tracking from videos, mainly dedicated to traffic monitoring and visual surveillance. Detection of moving objects in video streams is the first stage in any automated video surveillance. Aside from the intrinsic usefulness of being able to segment video streams into moving and background components, detecting moving blobs provides a focus of attention for recognition, classification, and activity analysis, making these later processes more efficient since only "foreground" pixels need be considered. Tracking aims to describe trajectories of moving objects during time. The main problem to solve for tracking is to find correspondences of the same physical objects in different frames. Since this method uses color information for tracking, blurring causes no loss of data [1]. The frame work of the system starts with the acquiring of video images by means of camera and pre-processing has to be

1. Information Science & Engineering Dept, VTU RYMEC, Cantonment Bellary
sreepathib@gmail.com

2. Information Science & Engineering Dept, VTU RYMEC, Cantonment Bellary
hkeerthi3@gmail.com

done on them for enhancing the quality of frames in the sequences. The video frames have a lot of noise due to camera, illumination and reflections etc. This can be removed and quality of images can be enhanced with the help of preprocessing stages. The suitable steps should be carried out in this stage. The next stage is motion segmentation which separates foreground images from background images and it is followed by Object classification, Tracking and Human pose modeling.

2. **LITERATURE SURVEY**

In recent years, many computer vision based approaches for people counting and tracking to deal various applications are proposed. Researchers are trying to develop robust background modeling algorithm to recognize the background and track the moving object.

The feature based tracking of objects have various methods and techniques for performing detection of objects. For accurate detection, suitable methods have to be followed which may affect the factors such as illumination changes over time and shadows of objects. The contribution of researchers have been made in the areas of both indoor as well as outdoor environment scenes for tracking and detection of objects is high and they provide some solution for the aforementioned unsolved problems. In [1], the motion and features of person is tracked in videos using tag assignment. In [2], the robust and the efficient detection method based mainly on statistical and knowledge based is proposed by R.Cucchiara et.al. This method can handle situations where there is a change in luminance condition. In [3], for detecting the contours of moving objects the researchers used color segmentation and non-parametric approach. In [4], Tang Sze Ling et. al., proposed a method of using color information for differentiating various objects and for handling occlusion. In [5], S.J.Mckenna et.al proposed a method which detects people using mutual occlusions. They form groups using mutual occlusion and separating the groups from one another using the color information. In [6], I.Harintagin et.al., proposed an histogram based approach that locate the human body parts like head. Then using the head information, the number of people detected in the scene can be found out. In [7], for tracking multiple people and vehicles during crowded environment shape and color information are used which is proposed by A.J.Lipton et.al. But in case of occlusion, the parameters such as color, motion and edges guarantee the considerable results in different environment. In case of occlusion, color histogram is efficient but it is more sensitive in case of illumination changes. Along with color and motion, shape of an object also plays an important role in detecting the objects, where the shape is determined using parameterized rectangle or ellipse in [8]. In [9], the changes between two images taken in two different times are calculated using Hopfield Neural Network.

3. **THE PROPOSED METHOD**

In this work, initially foreground objects are segmented from the background. Next, the foreground object motion between the current and previous frames is obtained.

This algorithm aims to identify each object (person) appears in scene when individual merge into or split from the group and involves several methods to obtain the lowest possibility of false tracking. In tracking interested object (human), shadows affect the performance of tracking and leads to false tagging. To avoid this problem, we apply Morphological operations to remove noise. Since we are using color information for tracking, blurring causes no loss of data. The structural design of our proposed method shown in Fig.1

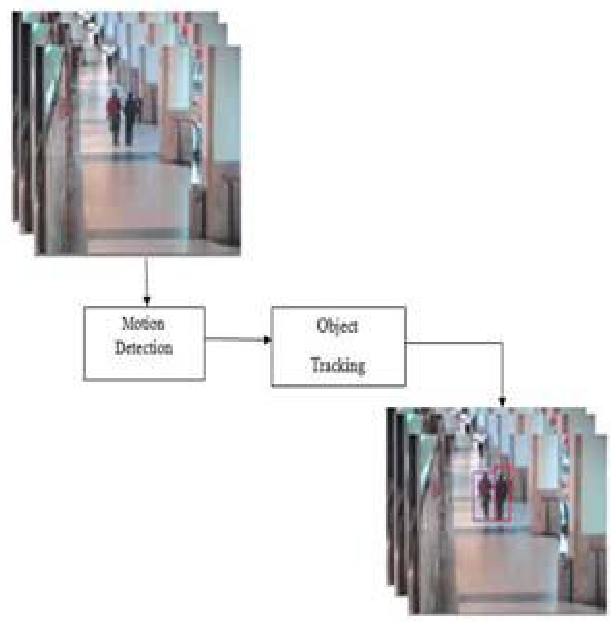


Fig. 1 System architecture

A. IMAGE FRAMING AND IMAGE ACQUISITION

In Data acquisition phase different kind of sensing and capturing input devices used for data gathering from the working environment. For example video capturing and monitoring devices used to record video stream. For Image framing video must be divided into sequence of frames. Video must be divided mostly into 20-30 frames which are sent to the next phases for further processing.

Most probably background subtraction method applies on the static background images. System can detect “motion area” (only useful data) by comparing reference frame with current frame and get the extraction of a (new) moving object from the background. This phase attempt background subtraction comparing reference image and current image in a pixel-by-pixel fashion as shown by equation below:

$$|P(X_1, Y_1)_C - P(X_2, Y_2)_{bk}| > T \quad (1)$$

Where $P(X_1, Y_1)_C$ is current frame pixel present at (X_1, Y_1) co-ordinate $P(X_2, Y_2)_{bk}$ is reference frame pixel present at (X_2, Y_2) co-ordinate and this threshold value of difference and T is the threshold which suppress shadow depending on the value assigned. Here foreground pixels are extracted from the static background image so the threshold is fixed. When the threshold value is less than the pixel difference, it is considered as foreground image. In this way the foreground objects present in the scene are detected. Only patterns are preprocessed for selection of significant features.

B. OBJECT SEGMENTATION

Most of the work on foreground objects segmentation is based on three basic methods, namely frame differencing, background subtraction and optical flow. Only background subtraction requires modeling of background. Building a representation of the scene called the background model and then finding deviations from the model for each incoming frame can achieve object detection. Any significant change in an image region from the background model signifies a moving object. Usually, a connected component algorithm is applied to obtain connected regions corresponding to the objects. This process is referred to as the background subtraction. An alternate approach for background subtraction is to represent the intensity variations of a pixel in an image sequence as discrete states corresponding to the events in the environment. Nearly every system within human motion analysis starts with segmentation, so segmentation is of fundamental importance. Although current motion segmentation methods mainly focus on background subtraction, how to develop more reliable background models adaptive to dynamic changes in complex environments is still a challenge.

A feature-based person-tracking algorithm requires useful feature selection, feature extraction, feature matching and proper handling of object's appearance and disappearance. Most of the works on tracking use a prediction on features in the next frame and compare the predicted value with estimated value to update the model. Selected feature then passed through the classification phase where classified objects are compared with the features of object using any mapping function. Optical flow segmentation can be used to provide the velocity of the moving object is distinguish from that of the background, and has expected characteristics.

C. MOTION DETECTION

The motion detection algorithms are broadly classified into two main categories. They are feature based and optical flow based. In video surveillance, the first step is to detect the moving objects. Next step is to segment the video streams into moving components and background components, where moving objects provide attraction for object recognition. Then classification and activity analysis processes are done for making this process more efficient.

The video surveillance system usually has two major components, one is detecting moving object the other one is to tracking them in sequence from video images as shown in fig.2.

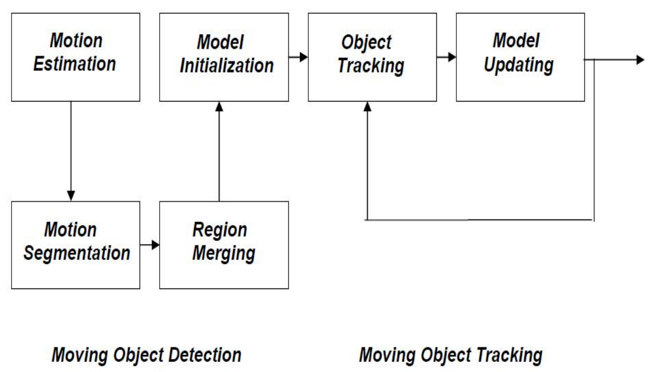


Fig . 2 General model for object detection and tracking

The accuracy of these components largely affects the accuracy of overall surveillance system. Detecting moving regions in the scene and separating them from background image is a challenging problem. In the real world, some of the challenges associated with foreground object segmentation are illumination changes, shadows, and camouflage in color, dynamic background and foreground aperture. Foreground object segmentation can be done by three basic approaches: frame differencing, background subtraction and optical flow.

The main objective of tracking is to describe trajectories of moving objects during time analysis. But it is very difficult to track such objects because it is necessary to provide correspondence of objects in different frames.

The most basic form of motion detection is the method of subtracting background image containing no objects from an image under test. Motion detection systems will not only be monitoring the areas of interest but will also keep an active lookout for any motion being produced. Frame differencing technique does not require any knowledge about background and is very adaptive to dynamic environments. There are several methods to background subtraction, including averaging background frames over time and statistical modeling of each pixel. Preprocessing based on mean filtering is done on the input video (i.e., image sequences) to equalize the light illumination changes and also to suppress the presence of shadows.

D. BACKGROUND SUBTRACTION

Basically background subtraction done subtracting reference image from current image, here in this algorithm temporal differencing is used to perform background subtraction.

In this method the reference image is the previous images, and it is subtracted from the current image. Background subtraction delineates the foreground from background in the images.

$$D_k(x,y) = \begin{cases} 1 & \text{if } |F_k(x,y) - B_{k-1}(x,y)| > T \\ 0 & \text{Otherwise} \end{cases} \quad (2)$$

Where $D_k(x,y)$ is the resultant difference, $F_k(x,y)$ is the current frame and $B_{k-1}(x,y)$ is the background initialized frame and T is the threshold value. In our implementation we will be keeping static background and hence will use the simple background subtraction method of human detection. The Fig.3 illustrates the result of background subtraction.

(a) (b) (c)

Fig. 3 Background Subtraction
(a) Background image initialization
(b) Current frame with Moving objects.
(c) Resultant background subtracted image

E. PERSON TRACKING

Each nd every frame is tracked to trace objects. The proposed methodology feature is to separate objects from the frame and identify the person who is present in that image frame. The motion blocks in the current frame are grouped as clusters. The matching information of motion blocks is compared between the current frame and the previous frame. By this comparison, people present in that frame are traced.

The objective of tracking is to establish correspondence of objects and object parts between consecutive frames of video. Object tracking is the method of detecting moving objects of interest and plotting its route by analyzing them. The object tracking algorithm utilizes extracted object features together with a correspondence matching scheme to track objects from frame to frame. Object detection in a video sequence is the method of detecting the moving objects in the frame sequence using digital image processing techniques. Background subtraction is the most commonly used technique for object detection. Background subtraction techniques for object detection from video sequence use the concept of subtracting the background model or a reference model from the current image. The methods considered in tracking of objects use various techniques for building the background model. It has been found that the methods require different time for execution and their performance differs in speed and memory requirements. The techniques involved in these algorithms are based on the intensity values of the pixels constituting the image. The background and illumination changes of the image influence the intensity values to a great extent, ultimately affecting the overall performance. In such situations, these methods fail to give accurate outputs and so there is no single algorithm that performs well in all conditions. An analysis of all these methods based on perturbation detection rate is used to evaluate the performance.

Any tracking method requires an object detection mechanism in each frame or in the first appearance of the object in the video. An ordinary approach for object detection is to use information in a single frame. But, some object detection methods utilize the chronological information computed from a sequence of frames to lessen the number of false detections. This temporal information is usually in the form of frame differencing, which highlights changing regions in consecutive frames. Given the object regions in the image, it is then the tracker’s task to perform object correspondence from one frame to the next to generate the tracks.

Once the object areas are determined in each frame, the tracking is performed to trace the objects from frame to frame. The color information from each blob is derived and tracking is performed by matching blob color. To handle occlusion, each motion blob is the key feature of proposed method is the color information of each object is extracted cluster-by-cluster. Each cluster has its own weight age for comparison. The color information is extracted from the motion blocks in the current frame to categorize matching color information between motion blocks in the current frame and previous frames. The first sub-task in object tracking is, each motion block in the current frame is segmented into areas of almost similar color as clusters (head, torso and feet). The second sub-task is, to identify matching color information between motion blocks in the current frame and motion blocks in the previous frames. After detection of an object as human being same object will track using its speed and velocity.

4. TESTS ON PETS DATASET

The above methods are implemented using Matlab. The motion detection can be accurately performed for efficient tracking. Morphological operations can be performed for noise removal and the occlusion problem can be handled.

A. MOTION DETECTION

Accuracy in motion detection is important for efficient tracking. The threshold should be set in such a way to avoid shadow to a greater extent also the blob size should be maintained properly and it depends on the application. The figure 4 shown below illustrates the results with various threshold values.

B. MORPHOLOGICAL OPERATIONS FOR NOISE REMOVAL

Due to environment factors (for example, humidity or fog in the area under surveillance, during transmission of video from the camera to the processing unit the illumination changes) the image is expected to contain noises. To get proper boundary lines (edges) without any noise portion morphological operations like opening and closing on the subtracted image are very useful. Opening is a combination of erosion and dilation operations with erosion followed by dilation whereas closing is dilation followed by erosion.

Human detection is based on the Navneet Dallal and Bill Triggs‟ Histogram of Oriented Gradients (HOG) for Human Detection. They used small region (cell) structure which accumulates a local 1-D histogram of edge orientations of the pixel value. A unique representation for the image and better invariance to illumination changes is given by combine entries of the histogram with contrast normalization. By simple 1-D masking the Gradients in the image is calculated, in the vertical (↑) and the horizontal (→) direction. A histogram of gradient direction is computed for multiple regions of fixed rectangular size which generates a 9 component feature vector. Gradient sign shows the contrast difference between the foreground and the background in the image so it is ignored. The histograms clearly capture the edge or gradient structure that is very characteristic of local shape. The SVM classifier decides whether the region of interest contains a human figure or not.

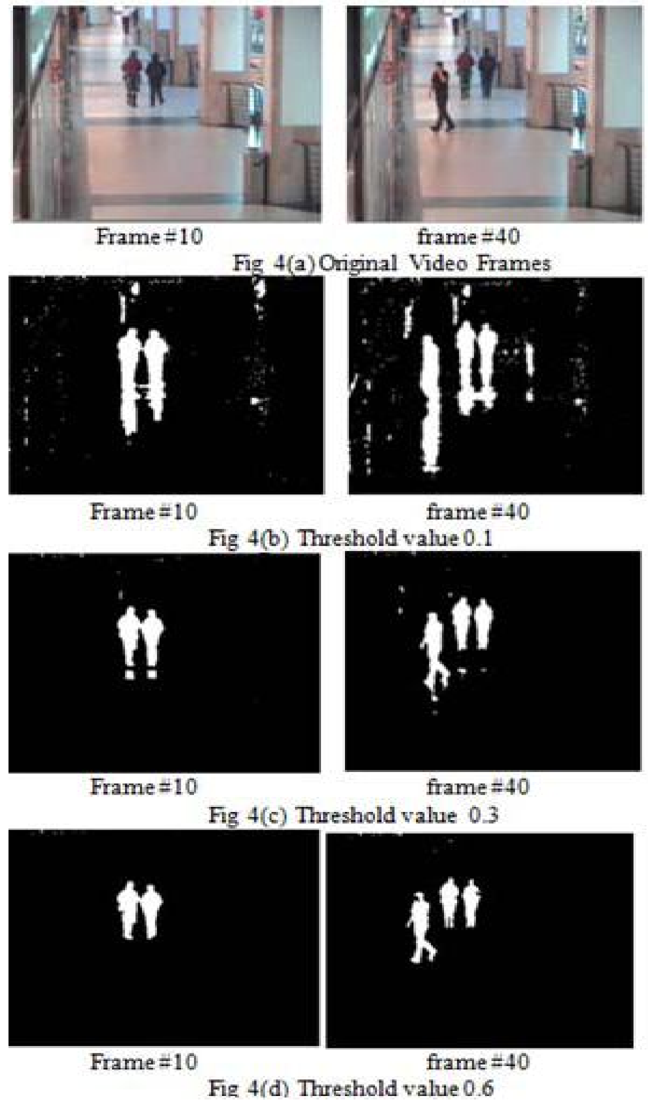


Fig.4 Motion detection

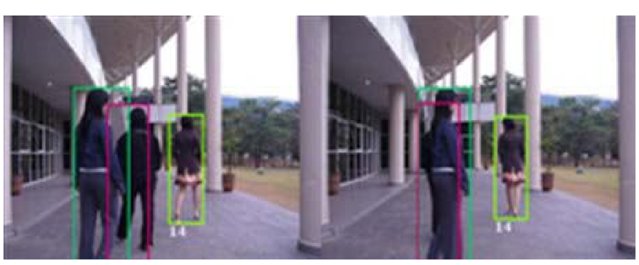
C. *HANDLING OCCLUSION PROBLEM*

To track same object between two consecutive frames some correspondence should be there, if correspondence found then mark that object with same color rectangle and classify that object as human being. To find correspondence few features of previous frame are stored and compared with the next frame, if features matches then same object found in both frames. And these features can be color, orientation, speed, Intensity etc. For tracking any object these features plays an important role. Tracking can be categories in two types one is region based tracking and second one is contour based tracking. In region based tracking features of the blob detected from two consecutive images if both matches then these two frame object gets related. In contour based tracking the energy of the boundary/ contour of blob detected from both frames and if that energy of the boundary matches then same object present in both images is declared. In this paper static background images are used and region tracking is very efficient with stationary background. So for tracking human being region tracking method is used. Region tracking stores the features of whole object for matching the features with the features of the object in the next frame. But in this procedure required so much time for comparison, to reduce this wastage of time another approach can be applicable that is salient region tracking.

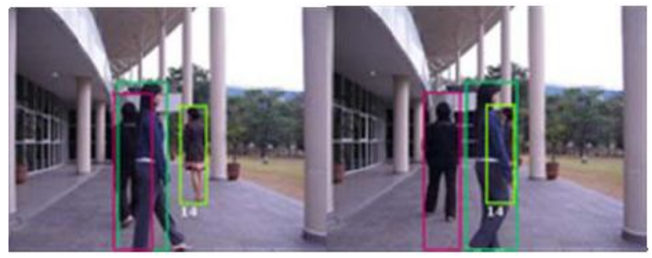
In most salient region method only matches the most salient region of the previous frame with the most salient region of the current frame. Hence, in less amount of time required to match the whole image. The most salient region tracking fetches the basic features from the image like color, orientation and intensity. Then from these fetched features, the feature vector is calculated for color, orientation and intensity using centre-surround method. After the feature map has been created to find out which feature more uniquely identifies the object features are weighted. Using their weight the saliency map created for the detected object. Finally from the previous image feature weight vector is calculated for this most salient region and this feature weight vector is matched with the subsequent next frames feature weight vector. If match is above the threshold value then there is a match and can predict the same object is present in both images. If match not found the search area is doubled and the same salient feature matching procedure is repeated. If the detected object match is not found after doubling the search area, then object is expected to be occluded by some other object or any stationary background object.

Occlusion handling is done with the tracking of object. When motion area of the moving object has been detected then its centroid of an object can easily obtained from both images.

Even after doubling the search area if object not found then algorithm can conclude that object is occluded by some other object area. Using previously calculated velocity and direction of motion system can predict the next possible area of an object, minimum 5 frames should be processed to calculate the velocity and the direction of motion. The fig.5 shows handling occlusion during tracking.



Frame 325 Frame 340



Frame 350 Frame 360

Fig.5 Occlusion handling

Each of the object in the group is identified which can be observe in frame 325 to frame 360. Frame 325 to frame 340 performed individual merge into group while frame 350 to frame 360 showed the individual split from the group.

5. CONCLUSION

This paper presents and discusses a two methods namely background subtraction and frame difference for surveillance system to detect moving objects in the field of view. The advantages of using color as feature to achieve object's similarity are robust against the complex, deformed and changeable shape (i.e. different human profiles). In addition, it is also scale and rotation invariant, as well as faster in terms of processing time. From the experimental results, it shows the ability to discriminate various occluded color objects.

This proposed method provides a potential to the surveillance system which required wide area observation and tracking persons over multiple cameras, for example airport, train station or shopping malls. It is not possible for a single camera to observe the complete area of interest as the structures in the scene constraint the visible areas and the devices resolution is restricted. Therefore, surveillance system of wide areas required multiple cameras for person tracking. Our future work is to implement our proposed method in multiple camera system in order to measure the robustness of the method in assigning a similar identifier for the similar person moving from one camera to another camera.

6. REFERENCES

[1] Besita Augustin.M, Sujitha Juliet, Palanikumar.S, ―Motion and Feature based Person Tracking in Surveillance Videos‖, in Proceedings of ICETECT, 2011.

[2] R. Cucchiara, C. Grana, G. Neri, M. Piccardi and A. Prati, ―The Sakbot system for moving object detection and tracking,‖ Video-based Surveillance Systems-Computer vision and Distributed Processing, pp. 145-157, 2001.

[3] L. Qiu and L. Li, ―Contour extraction of moving objects,‖ in Proc. IEEE Int'l Conf. Pattern Recognition, vol. 2, 1998, pp. 1427–1432.

[4] Tang Sze Ling, Liang Kim Meng, Lim Mei Kuan, Zulaikha Kadim and Ahmed A. Baha_a Al-Deen, ―Colour-based Object Tracking in Surveillance Application‖ in Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I IMECS 2009, March 18 - 20, 2009, Hong Kong.

[5] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld and H. Wechsler, ―Tracking group of people,‖ Comput. Vis. Image Understanding, vol. 80, no. 1, pp. 42-56, 2000.

[6] I. Haritaoglu, D. Harwood, and L. S. Davis, ―W4: Real-time surveillance of people and their activities,‖ In IEEE Trans. Pattern Analysis and Machine Intelligent, vol. 22, no. 8, 2000, pp. 809-830.

[7] A. Lipton, H. Fujiyoshi and R. Patil, ―Moving target classification and tracking from real-time video,‖ In DARPA Image Understanding Workshop, pp. 129-136, November 1998.

[8] C. Shen, A. van den Hengel, and A. Dick, "Probabilistic multiple cue integration for particle filter based tracking," in Proc. of the VIIth Digital Image Computing: Techniques and Applications. C. Sun, H. Talbot, S. Ourselin, T. Adriansen, Eds., 10-12, (2003).

[9] Beulah David.D, B.V.Krishna, ―Comparison of Analog and Discrete Hopfield Neural Network for Image Change Detection‖, CIIT International Journal of Digital Image Processing , 2010.