International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 4 Issue 2, Feb.-2015, pp: (14-21), Impact Factor: 1.252, Available online at: www.erpublications.com

Survey on Object Motion Estimation in case of video

Dharmendra Mehta[1], G. A. Kulkarni[2]
Electronics and Communication Engineering Department,
S.S.G.B.C.O.E.T, Bhusawal (Maharashtra), India

Abstract: The fast evolution of the digital video technology has opened new areas of research. Motion estimation in case of video is important application in this direction. The goal of motion estimation is segmenting a region of interest from a video scene and keeping track of its motion, positioning .It has two key steps first detect the moving object in video scene and track such object motion in subsequent video sequence. The detected object can be classified in various categories such as humans, vehicles, birds, floating clouds, swaying tree and other moving objects. Motion tracking is performed using monitoring objects' spatial and temporal changes during a video sequence, including its presence, position, size, shape, etc. Motion estimation is used in several applications such as computer vision, video surveillance, robot vision, traffic monitoring etc. This paper presents a brief survey of different methods of object detection in video and motion tracking algorithms available in the literature including analysis and comparative study of different techniques used for various stages of tracking.

Keywords: Background Modeling, background subtraction, frame difference, motion detection, motion Tracking, video object plan.

Introduction

With the development of information technology, people apply segmentation to digital video more and more extensively. Inside the video, there is much meaningful information in video object plane for example; the MPEG-4 multimedia communication standard enables the content-based functionalities by using the video object plane (VOP) as the basic coding element. Each VOP includes such as color, texture and shape information of an object in the scene [16]. Such information can be extracted for motion estimation of moving object in video. Moving target detection technology distinguishes moving objects and background, and extracts the moving target from digital video, which apply to the bank, highways, residential and other intelligent monitoring system. The results of object detection will directly affect the result of object motion tracking.

Motion detection consists on identifying the foreground and background objects into the scene and tracking is the method of following image objects in their movement through an image sequence. Many methods for motion detection and tracking have been proposed, each one having its own strengths and weaknesses. Motion detection of object generally vital part because background changes in dynamic environment condition such as change in weather condition, variation in lighting, waving tree etc. Foreground objects are generally moving objects like people, boats and cars and everything else is background [11]. Many a times shadow is classified as foreground object which further increase complication of motion tracking. General steps involve in object motion estimation from video are as follow:

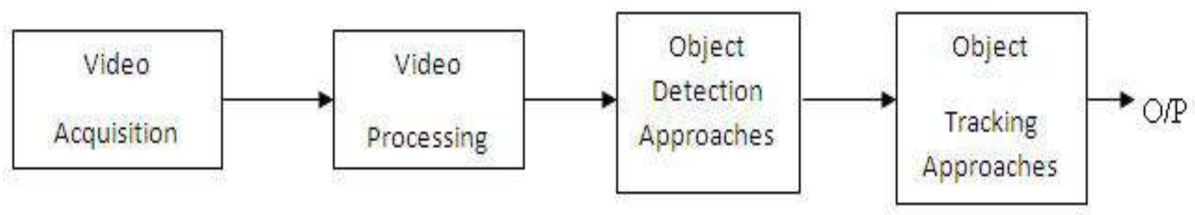

Figure 1. General Approach of motion Detection

A. Video acquisition and pre-processing

In Video acquisition video can be acquired by live video or prerecorded video. Generally it involves the acquisition of video by means of optical and electronic techniques. There is a great variety of physical phenomena used to acquire video: from 2D conventional images to 3D cameras or laser range sensors. Images are generated by: measurement of light intensity in one or several spectral bands, depth, sonic reactance, electromagnetic waves or nuclear magnetic resonance. Illumination is also a key element to provide optimal images, even in difficult situations encountered in outdoor navigation or industrial applications. Pre- processing techniques is applied to video image before prior to the extraction of information. Such techniques aimed to improve the quality and contrast of images, reduce the level of noise, re-sample the image into more convenient coordinates and obtain the area in the image where the subsequent processes was performed. Filtering methods can be applied to spatial, temporal and frequency domains, according to the nature of the images.

B. Motion Detection

Object Detection is to identify objects of interest in the video sequence and to group pixels of these objects. Object detection can be done by various techniques such as frame differencing, Optical flow and Background subtraction.

C. Motion Tracking

Tracking can be defined as the problem of approximating the path of an object in the image plane as it moves around a scene. The approaches to track the objects are point tracking, kernel tracking and silhouette. Following are some of the challenges that should be taken care in object tracking as described in [10]: The background subtraction methods tend to be sensitive to illumination changes, periodic motion of objects on the scene, occlusion, noise, shadow, camera jitter, etc.

Output of such system can be used for further analysis such as object classification, gesture recognition, object activity detection etc. Object can be classified as vehicles, animals, swaying tree and other moving objects. The approaches to classify the objects are Shape-based classification, Motion-based classification, Color based classification and texture based classification.

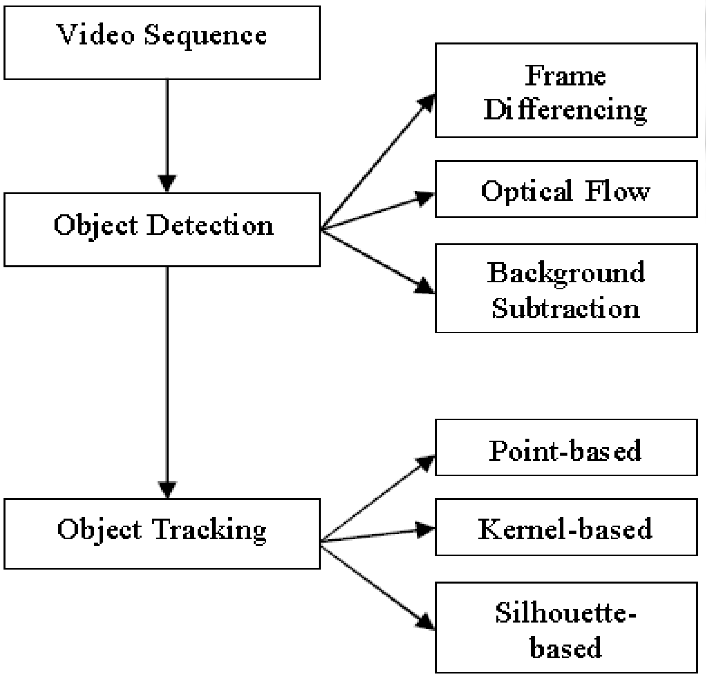

Figure 2. Basic steps for tracking an object [8]

Related Work

Motion estimation of moving object is one of the most important techniques in digital image processing. Image detection and segmentation is generally the first step in computer vision system, such as: video surveillance, traffic monitoring in sports intelligent, perception and object-based video compression. The result of image detection and segmentation is input to the other modules in machine vision application, so higher accuracy and lower error rate is needed. Then it could provide better foundation for subsequent processing. There are many algorithm proposed for object motion estimation and lots of research work done in same direction.

There are primarily two sources of information in video which are used in object detection and motion tracking. First is a visual feature (such as color, texture and shape) and second is motion information. The first step in video processing is to identify the objects present in a scene. The next step is to see how these detected objects move with respect to each other. The above two steps combine termed as "Object Tracking". The results of object detection directly affect moving object tracking. At present, there are many methods of moving object detection, such as background modeling [1], background subtraction [3] and optical flow [4]. In background modeling reliable background from input video sequence is constructed which is used in further video processing steps. The background subtraction method gets the moving target by comparing each frame image and background image. The optical flow method deals with object detection by computing the optical flow field, i.e. under the conditions of smoothness conditions, estimates moving field by the space-time gradient of image sequence and separates the moving target from background image by analyzing the change of the moving field.

However, background modeling is one of difficult tasks due to the changing environmental condition (such as illumination changes), periodic motion of objects on the scene, object shadow, camera noise, etc. If reliable background is not constructed then it made ghost in foreground image [5]. A ghost is false detection of foreground object. A ghost appears when the background images are not updated rapidly. The ghost effect and occlusion can be solved by adaptive background modeling [8] that rapidly update background image. However, in adaptive background modeling, the moving object that stops for a while can be absorbed into the background. Thus, the stopped object cannot be tracked successfully. Several approaches [4-9] have been proposed to construct and update the background information from the video sequence. In these approaches complex operations were involved to generate the background image. These approaches were developed for enhancing the coding efficiency. However, these approaches are not suitable for real-time application where fast processing is required.

Background subtraction is the process of detecting foreground pixels from the background image. The background subtraction approach required background information to get motion information of the object. This approach detects the moving target by comparing each input frame with background image. The main idea behind a background subtraction model is to find the difference between the background model and the current image. Most background subtraction algorithms are not robust against changing lighting conditions, non-static backgrounds, camera jitter, shadows, etc. There are several background subtraction methods have been developed to cope with these challenges such as frame difference [2], single Gaussian [12], Mixture of Gaussian [7] and Kernel Density Estimate [9] etc. These methods were differing in the way they judged whether a pixel belongs to the background or not.

In frame difference method the background model was an image supplied by the user or simply the first image of an image sequence. There are chances that this image may lag information related to foreground objects. For each time step the current frame was compared with the given background image. If for a certain pixel the difference in intensity was above a user-defined threshold, the pixel was appointed to be a foreground pixel. This method requires less computation time but it is not adaptable to any changes in the background.

The intensity distribution of the background at each pixel location was modeled by a single Gaussian [12] with a mean and a variance for every color channel. Since it was assumed that a foreground object is only present in the image for a short while, the Gaussian was determined by the intensity of the background pixels. If a pixel intensity of the current frame does not lie within a certain threshold determined by the variance of the Gaussian, the pixel was identified as a foreground pixel. In this method assumed a static background, since the intensity distribution of the background is modeled by one Gaussian only. It can handle only gradually changing backgrounds because the distribution parameters are the result of several consecutive frames. However, this method handles noise much better than the straightforward frame difference background subtraction method, since the threshold is determined by the noise level. Despite its shortcomings, the Single Gaussian background subtraction method can be used in practice for a controlled indoor environment.

Jiang Peng et al. [3] proposed foreground detection using a statistical representation of the scene background. In this system the weighted kernel density estimation was applied for each pixel by the analysis of temporal distribution in background initialization phase. Based on kernel density estimation, an adaptive threshold approach was demonstrated to estimate foreground threshold automatically. The main limitation of most traditional statistical solutions is their need for a series of training frames absent of moving objects. However, in some situations, e.g., public areas, it is difficult or impossible to control the area being monitored. In such cases it is necessary to train the model using a sequence which contains foreground objects. Another limitation of these methods is that, most schemes determine the foreground threshold experimentally. The threshold is quite arbitrary and required manual tuning.

In the Mixture of Gaussians, the intensity value of a color channel of a particular pixel is modeled. The pixel was modeled as a mixture of Gaussians [7]. By using multiple Gaussians for each pixel, this method handles a periodically moving background like weaving trees. The method is originally developed to cope with lighting changes, repetitive motions of scene elements, tracking through cluttered regions, slow-moving objects, and introducing or removing objects from the scene. Hence, it is evident that for static backgrounds the single Gaussian background subtraction model perform similar to the mixture of Gaussians. The latter it is only slow down the computation due to its computational complexity. Adding to these challenges, most applications require a background subtraction model that can perform in real-time.

Motion objects are often the surveillance objects in a visual surveillance system and motion segmentation is necessary for object-based video coding and motion analysis. Therefore, motion segmentation technique is of great research significance and application value and is a difficult and hot field. Liu Xuedong et al [11] proposed a fast motion segmentation algorithm based on hypothesis test. Although some algorithms of motion segmentation based on hypothesis test have been proposed, the tradeoff between complexity and efficiency has not been solved very well.

T. Sikora et al. [13], exploit the motion vectors available directly from the MPEG stream for object detection and tracking. Video segmentation, which extracts the shape information of moving object form the video sequence, is a key operation for content-based video coding, multimedia content description and intelligent signal processing. For example, the MPEG-4 multimedia communication standard enables the content-based functionalities by using the video object plane (VOP) as the basic coding element. Each VOP includes such as color, texture and shape information of a object in the scene. However, the shape information of moving objects may not be available from the input video sequences; therefore, segmentation is an essential tool to gain from this newly developed coding scheme. In addition, many multimedia communication applications have real-time requirement such as they requires fast computation and an efficient algorithm for video segmentation is very desirable.

Zhiwen Chen et al [4] proposed a way, which is based on optical flow, to track object by using the object contour. This algorithm achieves the effective object tracking in spatial position. The experiment result shows that moving object tracking is accurate, rapid and stable by using this algorithm.

Object detection and tracking remains an open research problem even after research of several years in this field. A robust, accurate and high performance approach is still a great challenge today. The difficulty level of this problem highly depends on how one defines the object to be detected and tracked. The typical challenges of background subtraction in the context of video surveillance have been listed below:

1) Illumination Changes:
It is desirable that background model adapts to gradual changes of the appearance of the environment. For example in outdoor settings, the light intensity typically varies during day. Sudden illumination changes can also occur in the scene [2], [6]. This type of change occurs for example with sudden switching on/off a light in an indoor environment. This may also happen in outdoor scenes (fast transition from cloudy to bright sunlight). Illumination strongly affects the appearance of background, and cause false positive detections. The background model should take this into consideration.

2) Dynamic Background:
Some parts of the scenery may contain movement (a fountain, movements of clouds, swaying of tree branches, wave of water etc.), but should be regarded as background, according to their relevance. Such movement can be periodical or irregular (e.g., traffic lights, waving trees). Handling such background dynamics is a challenging task [1-2].

3) Occlusion:
Occlusion (partial/full) may aspect the process of computing the background frame. However, in real life situations, occlusion can occur anytime a subject passes behind an object with respect to a camera [8], [12].

4) Presence of Shadows:
Shadows cast by foreground objects often complicate further processing steps subsequent to background subtraction. Overlapping shadows of foreground regions for example hinder their separation and classification. Researchers have proposed different methods for detection of shadows [6].

5)Video Noise:
Video signal is generally superimposed with noise. Background subtraction approaches for video surveillance have to cope with such degraded signals affected by different types of noise, such as sensor noise or compression artifacts [6].

6) Challenging Weather:
Detection of moving object becomes a very difficult job when videos are captured in challenging weather conditions (winter weather conditions, i.e., snow storm, snow on the ground, fog), air turbulence etc [6].

Object Detection Methods

First step in the process of object tracking is to identify objects of interest in the video sequence and to cluster pixels of these objects. Since moving objects are typically the primary source of information, most methods focus on the detection of such objects. Detailed explanation for various methods is given below.

D. Frame differencing

In which difference of two frames are taken for detection and tacking of moving object. Its calculation is simple and easy to implement. For a variety of dynamic environments, it has a strong adaptability, but it is generally difficult to obtain complete outline of moving object, responsible to appear the empty phenomenon, as a result the detection of moving object is not accurate [2].

E. Optical Flow

The optical flow method deals with object detection by computing the optical flow field, i.e. under the conditions of smoothness conditions, estimates moving field by the space-time gradient of image sequence and separates the moving target from background image by analyzing the change of the moving field. However this method involves computationally intensive operation and cannot be used for real time application [4].

F. Background subtraction:

In which reliable background is mathematically modeled. Once background is model moving object is detected by comparing current frame with background frame. However the method is very sensitive to the changes in the external environment and has poor anti- interference ability [3],[12]. However, it can provide the most complete object information in the case background is known. Some of the commonly used approaches are as follow

1) Mean filter:
In this approach for calculating the image containing only the background, a series of preceding images are averaged.

2) Running Gaussian average:
In this approach, fitting a Gaussian probabilistic density function (pdf) on the most recent N frames. In order to avoid fitting the pdf from scratch at each new frame time t, a running (or on-line cumulative) average is computed.

3)Background mixture models:
In this technique, it is assumed that every pixel's intensity values in the video can be modeled using a Gaussian mixture model. A simple heuristic determines which intensities are most probably of the background. Then the pixels which do not match to these are called the foreground pixels. Foreground pixels are grouped using 2D connected component analysis

The [2],[4],[12] excellent review have been published to report studies done in this area during past few decades.

TABLE 1: Comparative study of object detection methods [12]

Methods		Accuracy	Computation Time	Comments
Background Subtraction	Gaussian of Mixture	Moderate	Moderate	+ Low memory requirement – It does not cope with multimodal background
	Approximate Median	Low to Moderate	Moderate	+ It does not require sub sampling of frames for creating an adequate background model – It computation requires a buffer with the recent pixel values
Optical Flow		Moderate	High	+ It can produce complete movement information – Require large amount of calculation
Frame differencing		High	Low to Moderate	+ Easiest method. perform well in static background – It require background without moving object

Motion Tracking

The aim of an object tracker is to generate the path of an object over time by locating its position in every frame of the video [1]. But tracking has two definition one is in literally it is locating a moving object or multiple object over a period of time using a camera. Another one in technically tracking is the problem of estimating the trajectory or path of an object in the image plane as it moves around a scene. The tasks of detecting the object and establishing a correspondence between the object instances across frames can either be performed separately or jointly. In the first case, possible object region in every frame is obtained by means of an object detection algorithm, and then the tracker correspond objects across frames. In the latter case, the object region and correspondence is jointly estimated by iteratively updating object location and region information obtained from previous frames [1]. There are different methods of Tracking.

G. Point Tracking

Tracking can be formulated as the correspondence of detecting objects represented by points across frames. Point tracking can be divided into two broad categories, i.e. Deterministic approach and Statistical approach. Objects detected in consecutive frames are represented by points, and the association of the points is based on the previous object state which can include object position and motion.

1) Kalman Filter:
The Kalman filter is a recursive estimator. This means that only the estimated state from the previous time step and the current measurement are needed to compute the estimate for the current state. Thus, no history of observations or estimates is required. They are based on Optimal Recursive Data Processing Algorithm. The Kalman Filter performs the restrictive probability density propagation. Kalman filter [14] is a set of mathematical equations that provides an efficient computational (recursive) means to estimate the state of a process in several aspects: it supports estimations of past, present, and even future states, and it can do the same even when the precise nature of the modeled system is unknown. The algorithm works in a two-step process. In the prediction step, the Kalman filter produces estimates of the current state variables, along with their uncertainties. Once the outcome of the next measurement (necessarily corrupted with some amount of error, including random noise) is observed, these estimates are updated using a weighted average, with more weight being given to estimates with higher certainty. Because of the algorithm's recursive nature, it can run in real time using only the present input measurements and the previously calculated state and its uncertainty matrix; no additional past information is required. Kalman filters always give optimal solutions.

2) Particle Filtering:
The Kalman and Particle filters both are algorithms that recursively update an estimate of the state and find the innovations driving a stochastic process given a sequence of observations. However, the limitation of the Kalman filter is the assumption that the state variables are normally distributed (Gaussian). Thus, the Kalman filter will give poor estimations of state variables that do not follow Gaussian distribution. This limitation can be overcome by using particle filtering. The particle filtering [12] generates all the models for one variable before moving to the next variable. Algorithm has an advantage when variables are generated dynamically and there can be unboundedly numerous variables.

This algorithm typically uses contours, color features, or texture mapping. The particle filter [12] is a Bayesian sequential importance Sample technique, which recursively approaches the afterward distribution using a finite set of weighted trials. It also consists of fundamentally two phases: prediction and update as same as Kalman Filtering.

3) Multiple Hypothesis Tracking (MHT):
MHT is an iterative algorithm [12], in which several frames have been observed for better tracking outcomes. Iteration begins with a set of existing track hypotheses. Each hypothesis is a group of disconnect tracks. For each hypothesis, a prediction of object’s position in the subsequent frame is made. The predictions are then compared by calculating a distance measure. MHT is capable of tracking multiple object, handles occlusions and calculating of optimal solutions.

H. Kernel Based Tracking

Kernel tracking [9] is typically performed by computing the motion of the object, which is represented by a primitive object region, from one frame to the next. Object motion is in the form of parametric motion or the dense flow field computed in subsequent frames. Kernel tracking methods are divided into two subcategories based on the appearance representation used i.e. Template and Density-based Appearance Model and Multi-view appearance model.

These algorithms differ in terms of the presence representation used, the number of objects tracked, and the method used for approximation the object motion. In real-time, illustration of object using geometric shape is common. But one of the restrictions is that parts of the objects may be left outside of the defined shape while portions of the background may exist inside. This can be detected in rigid and non-rigid objects.

1) Simple Template Matching:

Template matching [9][4] is a brute force method of examining the Region of Interest in the video. In template matching, a reference image is verified with the frame that is separated from the video. Tracking can be done for single object in the video and overlapping of object is done partially. Template Matching is a technique for processing digital images to find small parts of an image that matches, or equivalent model with an image (template) in each frame. The matching procedure contains the image template for all possible positions in the source image and calculates a numerical index that specifies how well the model fits the picture that position. It can capable of dealing with tracking single image and partial occlusion of object.

2)Mean Shift Method:

Mean-shift tracking tries to find the area of a video frame that is locally most similar to a previously initialized model. The image region to be tracked is represented by a histogram. A gradient ascent procedure is used to move the tracker to the location that maximizes a similarity score between the model and the current image region. In object tracking algorithms target representation is mainly rectangular or elliptical region. It contain target model and target candidate. To characterize the target color histogram is chosen. Target model is generally represented by its probability density function (pdf). Target model is regularized by spatial masking with an asymmetric kernel.

I. Silhouette Based Tracking Approach

It provides an accurate shape description of the target objects [9]. The goal of silhouette tracker is to find the object region in each frame by means of an object model generated using the previous frames. Silhouette trackers can be divided into two categories i.e. Shape matching and Contour tracking.

Object tracking consists in estimating of the trajectory of moving objects in the sequence of images. The most important is the automation of object tracking is a challenging task. Dynamics of multiple parameters, changes representing features and motion of the objects and temporary partial or full occlusion of the tracked objects have to be considered.

1)Contour Tracking:

Contour tracking methods [9], iteratively progress a primary contour in the previous frame to its new position in the current frame. In the contour-based tracking algorithm, the objects are tracked by considering their outlines as boundary contours. Thereafter these contours are updated dynamically in successive frames. The discrete version of this approach is represented in active contour model. The discrete version of this approach takes the advantage of the point distribution model to limit the shape. However, this algorithm is highly sensitive to the initialization of tracking, making it difficult to start tracking automatically. The most significant advantage of silhouettes tracking is their flexibility to handle a large variety of object shapes.

2)Shape Matching:

These approaches inspect for the object model in the existing frame. Shape matching performance is similar to the template based tracking in kernel approach.

Another approach to Shape matching [12] is to find matching silhouettes detected in two consecutive frames. Silhouette matching, can be considered analogous to point matching. Detection based on Silhouette is carried out by background subtraction. Models object are in the form of density functions, silhouette boundary, object edges. Capable of dealing with single object and Occlusion handling will be performed in with Hough transform techniques.

Conclusion

Along with the increasing popularity of video on internet and versatility of video applications, availability, efficiency of usage and application automation of videos will heavily rely on object motion detection and tracking in videos. Although so much work has been done, it still seems impossible so far to have a generalized, robust, accurate and real-time approach that will apply to all scenarios. This will require, combination of multiple complicated methods to cover all of the difficulties, such as noisy background, moving camera or observer, bad shooting conditions, object occlusions, object shadow etc. Of course, this will make it even more time consuming. Research may go more directions, each targeting on some specific applications. Some reliable assumption can always be made in a specific case, and that will make the motion detection and tracking problem much more simplified. More and more specific cases will be conquered, and more and more good application products will appear.

Acknowledgements

I would like to express profound gratitude to my guide Prof. G. A. Kulkarni for his invaluable support, encouragement, supervision and useful suggestions throughout this work. His moral support and continuous guidance enabled me to complete my work successfully. I am also thankful to Prof. D. O. Sakle who provided me a stimulating environment, encouragement and support for preparing this paper.

References

[1]. T. Aach, A. Kaup, and R. Mester, "Statistical model-based change detection in moving video", Signal Processing, vol. 31, pp. 165–180, Mar. 1993.

[2]. Shigang Wang, Xuejun Wang, Hexin Chen, "Video object segmentation based on frame differences and its implementation on DSP", IET, pp. 618-621, 2008.

[3]. Jiang Peng, Jin Wei Dong, "Statistical Background Subtraction with Adaptive Threshold", IEEE, pp. 123-127, June 2012.

[4]. Zhiwen Chen, Jianzhong Cao, Yao Tang, Linao Tang, "Tracking of Moving Object Based on Optical Flow Detection" , IEEE trans. on Computer Science and Network Technology, IEEE, pp. 1096-1099, December 2011.

[5]. Fatih Porikli, "Achieving Real-Time Object Detection and Tracking under Extreme Conditions," Journal of Real-Time Image Processing, vol 1, pp. 33-40, 2006.

[6]. Shao-Yi Chien, Shyh-Yih Ma, and Liang-Gee Chen, "efficient moving object detection algorithm using background registration technique", IEEE trans. on circuits and systems for video technology, vol. 12, no. 7, pp. 577-586, July 2002.

[7]. Feng Wang, Shuguang Dai, "Adaptive Background Update Based on Mixture Models of Gaussian", IEEE trans. pp. 336-339, 1999.

[8]. Tao Yang, Stan Z.Li, Quan Pan, Jing Li, "Real-time Multiple Objects Tracking with Occlusion Handling in Dynamic Scenes," CVPR 2005, vol 1, pp 970-975, 2005.

[9]. Mr. Joshan Athanesious J; Mr. Suresh P, "Implementation and Comparison of Kernel and Silhouette Based Object Tracking", International Journal of Advanced Research in Computer Engineering & Technology, pp 1298-1303, March 2013.

[10].C.R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-Time Tracking of the Human Body", IEEE trans. Pattern Analysis and Machine Intelligence, Vol. 19, no. 7, pp. 780-785, July1997.

[11].Liu Xuedong, Wang Hong, "A Fast Motion Segmentation Algorithm Based on Hypothesis Test for Surveillance Video Coding", IEEE, 2010.

[12].J.Joshan Athanesious, P.Suresh, "Systematic Survey on Object Tracking Methods in Video",International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 242-247, October 2012.

[13].T. Sikora, "The MPEG-4 video standard verification model", IEEE trans. Circuits Syst. Video Technol., vol. 7, pp. 19–31, Feb. 1997.

[14].Greg Welch, Gary Bishop," An introduction to the Kalman Filter", In University of North Carolina at Chapel Hill, Department of Computer Science. Tech. Rep. 95-041, July-2006.

[15].P. Salembier and F. Marques, "Region-based representations of image and video: Segmentation tools for multimedia services", IEEE trans. Circuits Syst. Video Technol., vol. 9, pp. 1147–1169, Dec. 1999.

[16].Rahul Mishra, Mahesh K. Chouhan, Dr. Dhiiraj Nitnawwre, "Multiple Object Tracking by Kernel Based Centroid Method for Improve Localization", International Journal of Advanced Research in Computer Science and Software Engineering, pp 137-140, July-2012.