

A Survey on Moving Object Detection and Tracking Techniques

Shilpa #1, Prathap H.L #2, Sunitha M.R #3

#1,2 P.G Scholar, #3 Associate Professor, Department of CS&E

A.I.T, Chikkamagalur, Karnataka, India

shilpaygowda@gmail.com, prathap.jan18@gmail.com, sunithamr2310@gmail.com.

Abstract

Moving object detection and tracking are the more important and challenging task in video surveillance and computer vision applications. Object detection is the procedure of finding the non-stationary entities in the image sequences. Detection is the first step towards tracking the moving object in the video. Object representation is the next important step to track. Tracking is the method of identifying, the position of the moving object in the video. Identifying the position is much more challenging task then detecting the moving object in a video. Object tracking is applied in numerous applications like in robot vision, monitoring the traffic, Video surveillance, Video in-painting and Simulation. Here we are going to present a brief review of numerous object detection, object classification and object tracking algorithms available.

Keyword: Object detection, Object representation, Object tracking, Performance Analysis.

1. Introduction

Video scrutiny is a more challenging task in today's environment. Tracking the moving object has attracted many researchers in the field of computer vision and image processing. Video surveillance is the process of observing the behaviour, events and other necessary evidence, typically of the people for supervision and protecting them. Surveillance is mainly used by governments for gathering information, for investigating and preventing the crime. Video surveillance activity can be divided into three types namely. Manual video surveillance, Semi-autonomous surveillance and Fully-autonomous surveillance. In manual surveillance human is responsible for analysing the contents of the video. Semi-autonomous surveillance system involves video processing by system with the interaction of the human when necessary. Fully-autonomous surveillance system, the system will perform every task such as detection of motion, tracking etc., without the need of human interaction.

Automated monitoring procedure also known as the Intelligence Visual Surveillance (IVS) involves the scrutiny and interpretation of objects activities, in addition to object detection and tracking to recognize the visual actions of the scene. The main task of IVS consists of wide area surveillance control and scene interpretation. Object tracking must deal with several illumination changes and well-known challenges. Mainly video analysis is categorized into three basic phases: moving entity detecting, finding the trajectory of object from one frame to another frame and scrutiny of entity tracks to identify their performance. Tracking objects

in a static environment is much easier than tracking objects in dynamic environment.

In common, in a dynamic ecological system together background and entity varies [1]. In principle, to resolve this common unconstrained problem is rigid. One can place a group of impulsion to make this problem answerable. The more the impulsions, the problem is informal to solve.

The rest of the paper is systematized as follows. Section II presents methodology in which object detection, representation and tracking Techniques is illustrated. Performance analysis is described in Section III. Conclusion about the paper is presented in Section IV.

2. Methodology

Detection of objects in motion is the first step towards non-stationary object tracking. Object detection is the method of finding the non-stationary object in a video sequence. Some of the major and important methods of detecting the moving objects are Frame differencing, Optical flow, Background subtraction and Double difference etc. Object representation is the process of demonstrating the objects. Object representation can be categorized as shape representation, color representation, texture based representation and Motion oriented representation. Object tracking is the process determining the position of the moving entity in a sequence of video.

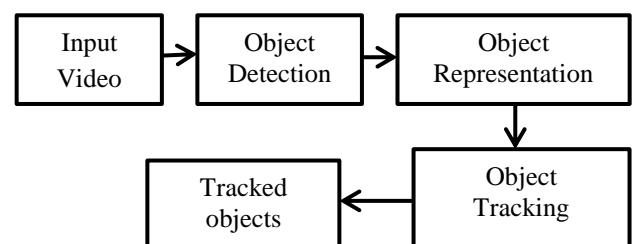


Figure 1: Phases of Moving Object Tracking

Certain types of tracking algorithms are point tracking, Motion tracking, Shape tracking, Feature tracking

and Kernel based tracking. Step by step detail of the non-stationary object tracking is as shown in figure 1.

2.1) Object detection

Object detection is the method for recognizing the non-stationary or moving object in a video sequence. This is the primary and main step towards moving object tracking.

Different detection techniques are as despite in figure 2

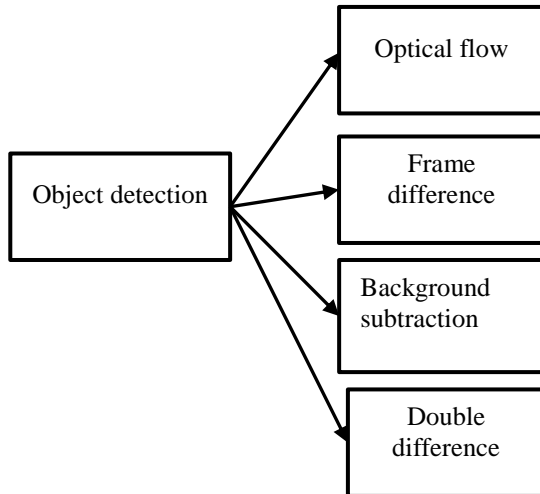


Figure 2: Object Detection techniques.

2.1.1) Optical Flow

Optical flow is substitute standard from of object detection in which the optical flow arena of the image is calculated and grouping of those arenas is done rendering to appearances of the image. The motion among dual video frames occupied at time t and $t + \delta t$ at every single location is estimated in optical flow process. This technique gives the broad information regarding the movement of the object. And also detects the object accurately compared to that of background technique. This method is not widely used because of its huge calculation and it is very sensitive to noise. It is not good for real-time occlusion condition.

2.1.2) Background subtraction

Background subtraction is the most widely used method for moving object detection. It can be of two types firstly by considering first frame as the reference frame or background image. Secondly by considering average of 'n' frames as the background image. In this background subtraction method every pixel of on-going frame is subtracted with the pixels of the background image. The equation (1) and (2) shows the background subtraction method for first frame as the background image.

$$B(a, b) = A(a, b) \quad (1)$$

where $B(a, b)$ represents background image pixel by pixel. The background subtraction method splits the video frames into foreground and background object, where the foreground object is discovered by matching the current frames $A(a, b)$ through the background image $B(a, b)$. The equation used is

$$C(a, b) = \begin{cases} 1 & \text{if } B(a, b) - A(a, b) > \text{threshold} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where $C(a, b)$ is the foreground pixel, threshold value can be set manually or can selected automatically as per video input.

This method consumes less memory. Accuracy of detection is moderate. But it will not suit for multimodal backgrounds. Result of the background subtraction methods are as depicted in figure 3.

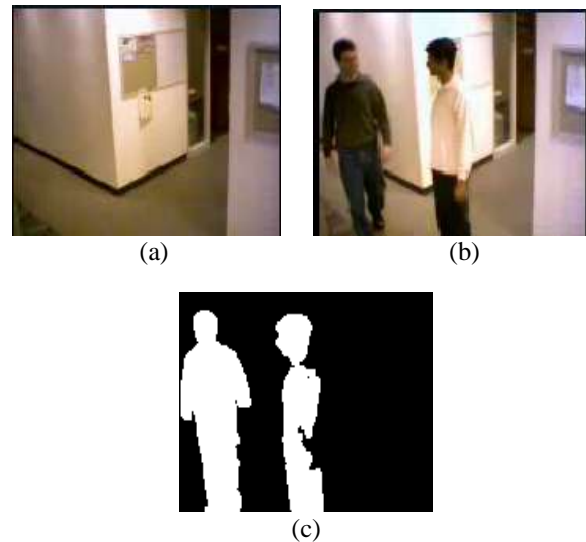


Figure 3: Background subtraction result for hall monitoring video (a) Background model (b) Current frame #89 (c) Resulting frame.

2.1.3) Frame difference

The frame difference scheme is also known as the temporal difference, in which each current frame pixel is subtracted with its prior frame pixel. If the transformation is superior to the manually set threshold value than that pixel is reflected as the foreground pixel else the pixel is reflected as the background pixel. Equation (3) presents the way for frame difference

$$F(a, b) = \begin{cases} 1 & \text{if } I_n(a, b) - I_{n+1}(a, b) > T \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Where I_n is the prior frame pixel and I_{n+1} is the pixel value of the current frame. T will be the threshold value which is manually defined by the user. Calculation of this process is modest and easy. For non-static environments, it is very challenging to achieve whole outline of the moving entity. So it is very cumbersome to obtain accuracy. Results of the frame difference methods are as depicted in figure 4.





(c)

Figure 4: Frame difference result of hall monitoring video. (a) frame #47 (b) frame #48 (c) Resulting frame.

2.1.4) Double difference

The frame difference scheme is also known as the temporal difference, in which each current frame pixel is subtracted with its prior frame pixel and immediate next frame pixel. If the renovation is more than defined threshold value then that pixel is reproduced as the foreground pixel else, the pixel is replicated as the background pixel.

$$C_n(a, b) = D_n(a, b) - D_{n+1}(a, b) \quad (4)$$

$$C_{n+1}(a, b) = D_{n+1}(a, b) - D_{n+2}(a, b) \quad (5)$$

$$DD(a, b) = C_n(a, b) - C_{n+1}(a, b) \quad (6)$$

Where $C_n(a, b)$ is the resulting foreground pixel. D_n denotes the current frame of the video sequence. D_{n+1} indicates the next frame. Similarly in equation (5) D_{n+1} is the current frame, D_{n+2} is the next frame. Finally $DD(a, b)$ specifies the resulting double difference frame pixel value.

$$R(a, b) = \begin{cases} 1 & \text{if } DD(a, b) > Th \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

where Th is the threshold value. If the pixel of the absolute difference is larger than the threshold value than the pixel is reflected as black otherwise it is reflected as white pixel.

This method produces accurate movement of the objects. But it consumes large memory and its takes more time to calculate.

2.2) Object representation

The extracted moving object may be of any types like human, vehicles, trees, floating clouds, birds and other non-stationary objects. Strategies to represent moving objects are shown in figure 5.

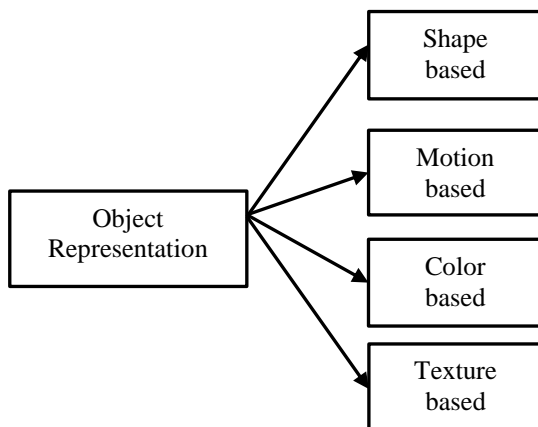


Figure 5: Object Representation Methods.

2.2.1) Shape-based representation

Dissimilar imageries shape data of motion regions such as depictions of points, Blob and boxes are accessible for categorizing objects in motion [2]. Mixture of image and scene object constraint like blob region of image, deceptive feature ratio of blob rectangle box area will form the input feature for network of images. Ordering is accomplished on every blob at respective frame and outcomes are retained as histogram.

2.2.2) Motion-based classification

Non-rigid object motion illustrates a interrupted assets, since it is been used as a robust indication for dynamic object organization. Optical flow process is convenient method for object grouping [3]. Residual flow may be cast-off to scrutinize inflexibility and periodicity of moving objects. It is predictable that inflexible entities may describe minute lasting flow whereas nonrigid objects like human being has composite average residual flow and even displays a intermittent component.

2.2.3) Color-based Representation

Unlike several features of an image the relatively constant feature is color information which does not change and easy to be developed [4]. Even though color is not permanently applicable as the individual resources of identifying and tracking objects, even the processes that contain small computational cost makes color as an significant feature to practice when suitable. The most significant technique is color histogram for detecting and tracking non-stationary objects in real time[5]. A Gaussian mixture model is recycled to illustrate the color scattering surrounded by the structure of images.

2.2.4) Texture-based Representation

Texture based system counts the existences of gradient alignment in confined parts of an image, then calculates the information on a condensed grid of consistently spread out cells and uses overlapping narrow disparity standardization for enhanced accuracy. Texture feature is important to measure the intensity disparity of surfaces and are apprehensive with object pattern demonstration [1].

2.3) Object tracking

Object representation is the process of finding the route of the moving objects in a sequence of the images [4]. Object tracking is accomplished to find or produce, the path for entity by discovering its location in each frame. Major categories of the tracking are Point based tracking, Kernel based tracking, Shape based tracking, Motion based tracking etc. as shown in figure 6.

Some of the most widely used tracking techniques are Mean shift tracking, CAM-Shift tracking, KLT tracking etc.

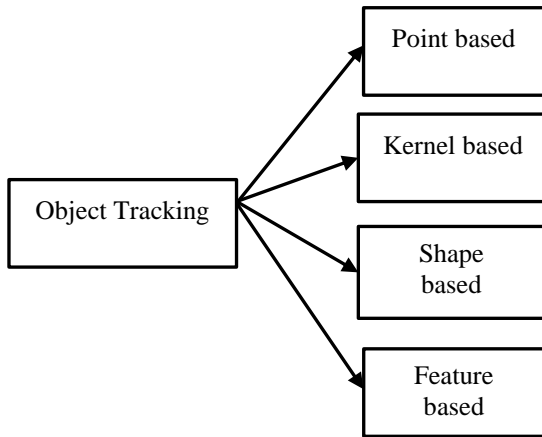


Figure 6: Object tracking methods.

2.3.1) Point based tracking

Moving objects are usually denoted by their feature point in the image sequence during tracking. Point tracking does not suit for occlusion condition and false object detection. Some of the points tracking algorithms are as described below.

a) Kalman Filter

The Kalman filter is valuable for tracking distinct types of moving objects [4]. It was initially created by Rudolf Kalman at NASA to track the path of spacecraft. Kalman filters can be utilized by various different types of linear dynamical systems. They are constructed on Maximum Periodic Documents Dispensation Algorithm. The Kalman Filter accomplishes the obstructive probability compactness propagation. It uses a group of scientific equations that provides an economical computational way to evaluate the objects state of occurrence. The Kalman filter measures a procedure through a form of opinion control.

The Kalman filter is a recursive two-stage filter. The two phases are predict and update [5]. In the predict phase, current location of the moving object is estimated or predicted based on the previous observation made. For illustration, if an object is moving with consistent dispatch, current location of object at X_t can be based on its prior location, X_{t-1} . In the update phase the capacity of the object's current location is pooled with the predicted location and acquires the posteriori projected current position of the object.

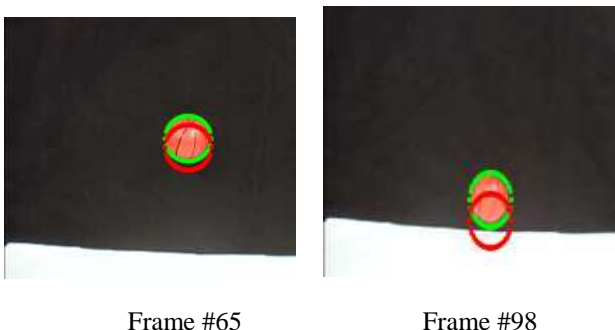


Figure 7: Kalman filter tracking results for video with ball moving in frame #65 and #98.

b) Particle Filtering

Particle Filter is most widely used filter to track single and multiple moving objects [2]. It is a hypothesis tracker, filtered subsequent distribution are estimated by a set of

weighted particles. One restriction imposed by Kalman filter is the predefined state variables are normally dispersed (Gaussian). This algorithm commonly uses, or texture feature mapping, contours, color features etc. The particle filter is a Bayesian chronological Trial procedure, which recursive approach. It also contains two phases similar to that of Kalman filter, predict and update phase.

2.3.2) Kernel based tracking

Kernel based tracking is typically performed to determine objects in motion, which is symbolized by elementary object area, since frame to frame[3]. The object movement is typically in the method of parametric movement such as transformation, affine, etc. This type of tracking can be applied for both rigid and non-rigid moving object. Kernel based tracking usually works based on the object representation, appearance, silhouette of the object.

a) Mean shift tracking

In mean shift tracking the position of the non-stationary object is tracked centred on the histogram [5]. In this method region of the object can be traced constructed on the similarity computation of the object. A gradient rise scheme is utilized to transfer the tracker to position that exploits a similarity value among the model and the current frame region. Target region are usually selected in rectangle form or elliptical form. This tracking method contains target model and candidate model [6]. Color histogram is usually chosen to characterize the target. Probability Density Function is used to represent the target model. If the similarity scores matches than the object is tracked. Results of mean shift tracking, first figure shows the target selection and second shows its tracking outcome.



Figure 8: Mean-shift Tracking results.

b) Simple Template Matching

Template matching is a heuristic search process of scrutinizing the important areas in the video [6]. A background image is matched with the sequence of images that are detached from a video. This algorithm suite well for single moving object tracking. Template matching is the method for handling sequence of images to discover minor portions of an image that matches model with a template image in every frame. The matching technique encompasses of image pattern for all probable locations in the basic image and computes a geometric key that identifies how accurately the model suits the image position. This process is widely used when single object or particle occluded object is to be tracked.

c) Layering based tracking

This is the most common mechanism for tracking multiple non-stationary objects in a video [7]. Every layer comprises of shape illustration, movement like rotation and translation, and layer presence, depending upon power. Layering is attained mainly by recompensing the background movement, such that the entity's movement can be calculated by the content image via 2D parametric gesture. All pixel's prospect is estimated based on the entity's previous gesture and shape structures. It can also track multiple images and complete occlusion condition of the item.

2.3.3) Shape based tracking

Certain objects would have composite shapes such as hand, finger, shoulder, nose etc. which cannot be demonstrated by exact geometric shapes [6]. The goal of this technique is to identify the moving object state in each and all frame through object model produced in the prior frame. This technique is proficient of tracking complex shape objects. Occlusion condition, split and merge condition can also be handled in this method. Information of the object region is utilized by this technique to track the objects.

a) Contour Tracking

Contour tracking methods, repeatedly process a principal contour of the prior structure to its new area in the present image sequence [6][7]. This silhouette procedure needs specific measure regarding items present in the current frame intersection with the entity district in the prior frame. Contour Tracking may be implemented utilizing two distinct methodologies. The principal methodology utilizes formal space models to demonstrate the form, shape and movement. The subsequent method straightforwardly develops the shape by reducing the contour vitality utilizing direct reduction methods, likewise gradient descent. This is flexible to handle huge variation of item shape.

b) Shape Based Matching

These methodologies inspect for the object model in the surviving frame. Shape identical algorithm works analogous to that of template based tracking method. This method is used to identify the matching shapes detected in 2 consecutive frames. Shape matching, can be measured comparable to point matching [3][4]. Background subtraction mechanism is utilized for detecting the shape of the object. Object reproductions are in the structure of density gatherings, shape frontier, edges of the objects. Shape based matching are proficient of dealing with solitary object and Hough renovate technique is utilized to handle occlusion condition.

2.3.4) Feature Based tracking

Feature based tracking is the most widely used tracking technique now a days. It is mainly divided into 2 steps. First phase is to extract the features of the entity like centroid, shape, color etc. Second step is to match those features in every frame. One or more feature can be combined to obtain better results or outcome. Cluster.

a) Color feature based tracking

The color information of every object in a video frame is assembled group-by-group [8]. Specific weights are assigned for every group for further evaluation process. The color evidence is obtained from the movement chunks in the current frame are divided into regions of nearly alike color as group.

Color information can be either RGB or HSV [1]. The next step is to recognize matching color information. This is accomplished by matching the group of color data of the motion chunk in the existing frame with the cluster color evidence motion chunks in the prior frames using weighted matching. The maximum evaluation score of every group in the existing frame is achieved and tracker is initialized to that position [8]. This process repeats until end of the video sequence.

b) Centroid feature tracking

The key feature centroid of each object is extracted. The centroid information is accessed from the gesture chunks in the recent frame to classify identical centroid information

among gesture blocks in the existing frame and prior frames [8]. Consequently, a bounding box with centroid is allocated for the gesture blocks in the existing frame. New centroid is detected by averaging the sum of previous centroid as in equation (8)

$$Cen_n = \frac{\sum_{i=1}^{n-1} cen_i}{n} \quad (8)$$

where Cen_n is the new centroid point and Cen_i prior centroid and n is the overall number of frames processed. Results of the centroid feature based tracking is as shown below



Figure 9: Tracking results for centroid based object tracking frame # 85 and frame # 131

c) Edge feature tracking

Edge feature tracking is similar to the color based feature tracking [8]. Edge of the moving object can be determined by various edge detection technique such as canny edge detection, sobel edge operator etc. in every frame. The next step is to identify matching edge information. This is obtained by matching the weight information of the gesture block of the present frame with the weight data of the gesture chunks in the prior frames. The highest score comparison of every group in the current frame is achieved and tracker is initialized to that position. This method is repeated until last frame of the video sequence.

III. Performance Evaluation

Performance of the video analysis is achieved by extracting manually non-stationary object as ground-truth image and compares it with the results obtained. Some parameter like True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) are estimated. TP refers to total amount of pixels where both ground-truth and proposed outcome overlap. TN refers to total number of pixels that does not comprise object, both in ground truth and proposed. FP is sum of pixels in which proposed system

contains object but ground truth does not. FN refers to total amount of pixel where ground-truth contains entity and projected system doesn't have objects.

Precision and Recall are some of the important factors for performance evaluation. Recall provides the percentage of the true positive rate that is similar to ground-truth image. It is given by the equation

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (9)$$

Precision offers the proportion of the true positive that does not match with the ground-truth image. The precision is found by using the equation

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (10)$$

A) MOTP (Multiple object tracking precision)

Total faults in identified area for compared object hypothesis among all frames, averaged via sum of comparisons made. It illustrates the capacity of tracker to determine specific object area, liberated of its proficiency at identifying objects formation.

$$\text{MOTP} = \frac{\sum_i t_i d_i^t}{\sum_i c_i} \quad (11)$$

MOTA (Multiple object tracking accuracy)

$$\text{MOTA} = 1 - \frac{\sum_i (m_i + fp_i + mme_i)}{\sum_i g_i} \quad (12)$$

Where m_i , fp_i indicates number of misses related to false positive. Time mismatches at an interval of 'i' is demonstrated by mme_i .

MOTA provides various error percentages that are

$$m' = \frac{\sum_i m_i}{\sum_i g_i} \quad (13)$$

Is the proportion of errors in the group, and is calculated around the sum of quantity of object existing in every frame.

Percentage of false positive and percentage of disparities is given by the equation (14) and (15) respectively

$$fp' = \frac{\sum_i fp_i}{\sum_i g_i} \quad (14)$$

$$mme' = \frac{\sum_i mme_i}{\sum_i g_i} \quad (15)$$

IV. Conclusion

In this paper different phases of object tracking technique like object detection, object representation and object tracking are summarized. Among various detection methods Background subtraction is the simplest method which provides complete information. A recent researcher chooses texture and color for representing objects. Object tracking can be achieved

via numerous algorithms which uses point, shape, feature techniques. These are some efficient algorithms which decrease calculation time. And also reduces the cost, used for tracking the entities for various kinds of the sequence of frames containing differentiated appearances. These are some of basic techniques which are commonly used. These techniques works well for static camera.

Despite the fact that amid the most recent centuries, has seen a significant advancement against object detection and tracking. However, tracking objects in a noisy, informality, and vigorous situations creates this issue as the research topic. Another key factor influencing is evaluation of reference frame powerful and proficient to ecological variations is still a inspiring work. Manipulation of previous and circumstantial information in tracking is quiet in its primary phase. A serious attention need to be paid towards tracking objects in noisy and dense video data.

References

- [1] Anand Singh Jalal, Vrijendra Singh " *The State of-the-Art in Visual Object Tracking*", International journal of computing and informatics, Vol. 36, Issue-3, pp. 227-248, May 2012.
- [2] J.Joshan Athanesious, P.Suresh, "Systematic Survey on Object Tracking Methods in Video", *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* pp. 242-247, October 2012.
- [3] A Yilmaz, O. Javed, M. Shah, "Object Tracking: A Survey", *ACM Computing Survey*, Vol. 38, Issue 4, pp. 1-45, 2006.
- [4] Gandham Sindhuja, Dr Renuka Devi S.M, "A Survey on Detection and Tracking of Objects in Video Sequence", *International Journal of Engineering Research and General Science* Volume 3, Issue 2, March-April, 2015.
- [5] Hitesh A Patel, Darshak G Thakore, "Moving Object Tracking Using Kalman Filter", *International Journal of Computer Science and Mobile Computing* , pg.326 – 332, April 2013.
- [5] S.Bhuvaneshwari, T.S.Subashini, "Tracking Manually Selected Object In Videos Using Color Histogram Matching", *International Journal of Theoretical and Applied Information Technology*, Vol. 67, Issues 3, pp. 562 -568, September 2014.
- [6] 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.
- [7] Kinjal A Joshi, Darshak G. Thakore, "A Survey on Moving Object Detection and Tracking in Video Surveillance System", *International Journal of Soft Computing and Engineering*, Vol. 2, Issue-3, July 2012.
- [8] Sunitha M.R, H.N Jayanna, Ramegowda, "Tracking Multiple Moving Object based on combined Color and Centroid feature in Video Sequence", in IEEE International

Conference on Computational Intelligence and Computing Research, pp. 846-850, Dec 2014.

[9] M.Besita Augustin, Mrs. Sujitha Juliet, Mr. S. Palanikumar, "Motion and Feature Based Person Tracking In Surveillance Videos", in Proceedings Of ICETECT, Vol. 19, no. 7, pp. 780-785, Dec 2011.

[10] Rupali S.Rakibe, Bharati D.Patil, "Background Subtraction Algorithm Based Human Motion Detection", International Journal of Scientific and Research Publications, May 2013

[11] Saravanakumar, S, Vadivel A, Saneem Ahmed, C.G., "Multiple human object tracking using background subtraction and shadow removal techniques", International Conference on Signal and Image Processing, pp.79-84, Dec. 2010.

[12] W. M. Hu, T. N. Tan, L. Wang and S. Maybank, "A survey on visual surveillance of object motion and behaviour", IEEE Transactions on Systems, Man and Cybernetics, vol.34, no.3, pp. 334-352, Dec 2004.

[13] Adam, Amit, Ehud Rivlin, and Ilan Shimshoni. "Robust fragments-based tracking using the integral histogram." Computer Vision and Pattern Recognition, Vol. 1. IEEE, 2006.

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

[15] Hitesh A Patel, Darshak G Thakore, "Moving Object Tracking Using Kalman Filter", International Journal of Computer Science and Mobile Computing, April 2013, pp.326 – 332.

Karnataka, India, in 2013. Currently pursuing Master degree in Computer Science and Engineering from Adichunchanagiri Institute of technology, Chikkamagaluru., Visveswaraya Technological University, Belagavi, Karnataka. His current research interests include Image processing and Object tracking.



Sunitha M R received BE degree in Computer Science & Engineering, Kalpataru Institute of Technology, Tiptur, Karnataka, India, Mtech degree in Computer Science & Engineering from Visveswaraya Technological University, RV College of Engineering, Bangalore, Karnataka, India, and currently pursuing her PhD in Image Processing from the department of Computer Science and Engineering, Visveswaraya Technological University, Siddaganga Institute of Technology, Karnataka. She has published number of papers in various national and International journals and conferences. She is currently working as Associate Professor, Department of Computer Science & Engineering, Adichunchanagiri Institute of Technology, Chikkamagalur, Karnataka, India. Her current research interests include Image processing, algorithms and Cryptography.



Shilpa, received the Bachelor degree in Computer Science and Engineering from Rajeev Institute of Technology, Hassan., Visveswaraya Technological University, Belagavi, Karnataka, India, in 2013. Currently pursuing Master degree in Computer Science and Engineering from Adichunchanagiri Institute of technology, Chikkamagaluru., Visveswaraya Technological University, Belagavi, Karnataka. Her current research interests include Image processing and Object tracking, Algorithms.



Prathap H.L, received the Bachelor degree in Computer Science and Engineering from Rajeev Institute of Technology, Hassan., Visveswaraya Technological University, Belagavi,