Live object monitoring, detection and tracking using mean shift and particle filters

Conference Paper · August 2016

DOI: 10.1109/INVENTIVE.2016.7830143

2 authors, including:

Sandip Rahane
Amrutvahini College of Engineering
12 PUBLICATIONS **29** CITATIONS

SEE PROFILE

Live Object Monitoring, Detection and Tracking Using Mean Shift and Particle Filters

S. P. Sangale
Department of Electronics Engineering
Amrutvahini College of Engineering
Sangamner, India.
sgr.sngl@gmail.com

S. B. Rahane
Department of Electronics Engineering
Amrutvahini College of Engineering
Sangamner, India.
sandiprahane@yahoo.com

Abstract— This paper presents efficient object tracking in video sequences using multiple features by embedding mean shift into particle filters. When clutter background and occlusions are present. Particle filtering is used because it is very robust and performs well for non-linear and non-Gaussian dynamic state estimation problems. The image features, such as shape, texture, color, contours, and random motion appearance can be used to track the moving object(s) in videos. We proposed a smart video surveillance system with real-time as well as offline stored video database for moving object detection, classification and tracking capabilities. The graphical user interface of the proposed system using MATLAB 2012b is implemented that operates on both color and gray scale video images from a stationary camera. In this method, we proposed real time moving object detection and tracking system using static webcam that can processes 800*600, 640*480,and 320*240 resolution video sequences for capturing live scene as well as stored database of segmented videos. Issues related with object detection and tracking is jointly employed using mean shift and particle filters. Tracker will estimate the dynamic shape and random appearance of objects. Tracking requires location and shape of object in every segmented frame. Rectangular bounding box is used to improve the estimate of its shape and appearance. Target object may experience partial occlusions, intersection with other objects with similar color distributions, abrupt motion speed changes, and cluttered background. Test result shows improvement in terms of robustness, tracking drifts, accuracy of tracked bounding box to occlusions.

Keywords— Real time object tracking, Unattended object, mean shift, object intersection, live learning of objects, partial occlusion, particle filters, colour features, edge features.

I. INTRODUCTION

Moving object detection and tracking surveillance system provides crucial information about the object behavior, interaction and the relationship between objects of interest to high level processing since automated video surveillance systems has modules from low-level such as object detection, object tracking, classification, event analysis, efficiency and robustness in each module are particularly important. To design smart computer vision based video surveillance systems which can contribute to the safety of people in the home and in public places such as, airports, railway stations, shopping malls and other public related places has now made it possible with the advent of smart multi mega pixel consumer cameras having higher processing capabilities. The critical threat of public safety due to terrorist attacks especially, explosive attacks with unattended packages, luggage's, bags, and vehicles are repeatedly concentrated on such public places. A key function in such a computer vision based video surveillance system is the understanding of human behavior in relation with objects left unattended in public places. In this context, real time smart visual surveillance object detection and tracking systems for human behavior understanding have drawn much attention of researchers and investigated worldwide as an active research topic [2]. Thus, a primary goal of video surveillance is to obtain a live description of what is happening in a monitored area and take (or trigger) appropriate action against pertained object and human misbehavior at public place, shopping malls, banks, railway station, airports….etc.

In video processing, a video can be represented with some hierarchical structure of units, such as video scene, shot and frame. Also, video frame is the lowest level in the hierarchical structure. The content based video browsing and retrieval uses these structure units for video content analysis. In video retrieval, generally, video applications must first partition a given video sequence into video shots. A video shot is defined as an image or video frame sequence that presents continuous action. The frames in a video shot are captured from a single operation of one camera. The complete video sequence is generally formed by joining two or more video shots consecutively. There are two basic types of video shot transitions, namely abrupt and gradual. Abrupt transitions i.e., cuts are the simplest form that can be occur in a single frame when restarting and stopping the camera. Although many kinds of cinematic effects could be applied to artificially combine two or more video shots.

Despite of many research works tracking multiple moving objects through complex scenarios remains a challenging task [1]. An independent robust tracker is used to detect and track each individual object is one category of existing method while using multiple independent trackers are used for multiple object tracking. Based on Gaussian mixture models,

some examples include moving object detection and tracking such as Eigen objects appearance modeling, Kalman filters, optical flow, and particle filters. In our robust visual tracking method that jointly employs an-isotropic mean shift and particle filters, our main objective includes comparison of existing work with mean shift and particle filters such as: 1) The joint particle filters and mean shift tracking scheme introduces the full combination of functionalities and adjustable bounding box parameters; 2) Partitioning a rectangular bounding box and deriving multi-mode anisotropic mean shift; 3) Introducing live learning method for the reference object distribution; and 4) Computing bounding box parameters through decomposition of Eigen vectors for geometry of partitioned areas and weighted average of parameters of bounding box. Following are some of the main merits of the proposed robust scheme which shows: a) Robustness in terms of long-term partial occlusions and intersections of objects; b) Efficient particle filter tracking by using a small number of particles; c) Robust mean shift through tunable parameters, anisotropic kernel and partitioning bounding boxes, tight tracked boxes due to accurate box parameter estimation [3].

II. LITERATURE SURVEY

There have been a number of surveys about object detection, classification, tracking and activity analysis in the literature [4]. There are several applications that benefit from smart video processing has divergent needs, thus requires different treatments. However, they have something in common: moving objects. Thus, detecting regions that correspond to moving objects such as people and vehicles at railway station, airports in video is the first basic step of almost every vision system since it provides a focus of attention and simplified the processing on subsequent analysis steps. Due to dynamic changes in natural video scenes such as sudden illumination and weather changes, repetitive motions and complex background that cause clutter motion detection is a difficult problem to process reliably. Commonly used techniques for moving object detection and tracking are background subtraction, temporal differencing, statistical methods, and optical flow method.

In object tracking first video is converted into number of consecutive frames then processes for locating moving objects from the scene. Real time object tracking is a challenging problem in the field of computer vision such as motion based recognition, automated surveillance, traffic monitoring, and object based video compression etc. Mean shift has drawn much interest in real time object tracking by maximizing the Bhattacharyya coefficient between the reference and the target i.e. objects region of interest. Early work was proposed by Comaniciu [6], Collins extended his work for mean shift by introducing kernel bandwidth normalization. It performs an extensive search of object within a range of bounding box scales and it is computationally intensive. Using anisotropic mean shift method the centre, size, shape and orientation of the bounding box are simultaneously estimated during the tracking was proposed by Sumin in [5]. Early work of particle filters have been proposed for visual tracking by Wang [7], by separating the state vectors into shape and appearance based (Eigen) sub-vectors thereby it reduces the number of particles required in the particle filter in which appearance is treated by linear models. Further improvement is made in visual object tracking by combining mean shift and particle filters is also called as the hybrid method which is integrated as single scheme. To track the human hands motion [5] by embedding the mean shift in particle filters where particles with higher weights from the mean shift are combined in the observation model, and it reduces the degeneracy and requires fewer particles than the conventional particle filter [3]. By applying the mean shift on particles with large weights is also as called elite particles to weight particles by using the observation model which is proposed by Zhong [8].

III. MOVING VISUAL TRACKING USING COMBINED MULTI-MODE AN ISOTROPIC MEAN SHIFT AND PARTICLE FILTERS

In this section we are describing the proposed tracking scheme using anisotropic mean shift embedded in particle filters as shown in fig.1. The basic idea of moving object tracking is to use the object appearance and its shape. The dynamic shape is estimated by the particle filter, and dynamic appearance is allocated by the mean shift. Further live learning approach is employed for updating the reference object distribution. The main objective of the tracker is to allocate the best region of interest i.e. object region (or bounding box) in each image frame is done by using particle filter that tracks the affine bounding box and weights to moving object box locations that are used as particle weights.

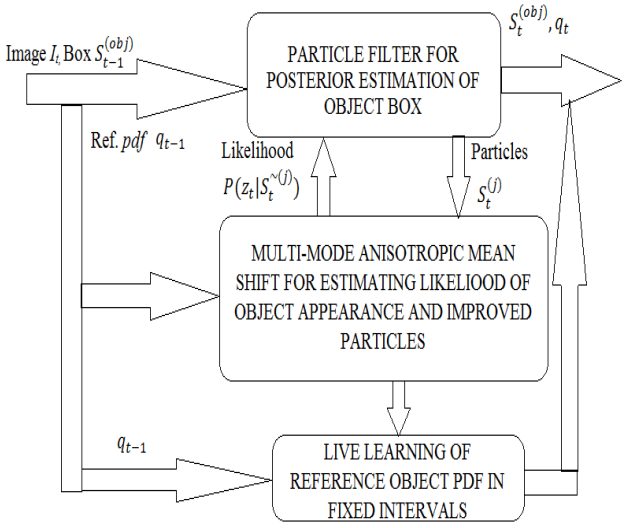

Fig.1. Block diagram of the proposed tracking scheme.

Notations used in the block diagram: I_t -image frame at time t, $S_{t-1}^{(obj)}$ and $S_t^{(j)}$ are the tracked box parameters at (t-1) and t, $S_t^{(j)}$is the jth particle at time t, q_{t-1} and q_t are the estimated pdf of the reference object at (t-1) and t. From top to bottom: Block-1, Block-2, and Block-3, [3].

To compute the parameters from a partitioned box:
To calculate the bounding box orientation according to Cartesian coordinate system, let θ be the angle between the long axis of kernel bandwidth ∑ and the horizontal axis. The height h and width w of the rectangular bounding box can be defined as in terms of its radii along the long and short axes of ellipse of the rectangular bounding box, as shown in Fig.2.

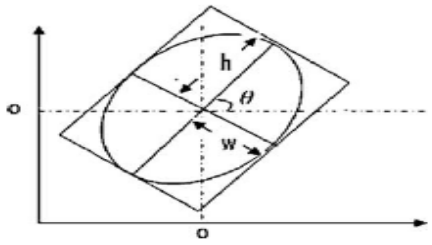

Fig.2. Definition of bounding box w.r.t to Width, Height and Orientation [3].

Given the kernel bandwidth ∑, kernel bandwidth follows that the box parameters h, w, θ, are related by

$$\sum = R^T(\theta)\begin{bmatrix}\left(\frac{h}{2}\right)^2 & 0\\ 0 & \left(\frac{w}{2}\right)^2\end{bmatrix}R(\theta)\ \&\ R=\begin{bmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{bmatrix} \quad (1)$$

Where 'R' is the rotation matrix, to estimate these parameters, Eigen vector decomposition is applied to ∑ such as $\sum = VAV^{-1}$.

$$A=\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix},\qquad V=\begin{bmatrix}v_{11} & v_{12}\\ v_{21} & v_{22}\end{bmatrix}$$

The orientation $\boldsymbol{\theta}$, the height $\boldsymbol{h}$ and the width $\boldsymbol{w}$ of the bounding box are obtained by,

$$\boldsymbol{\theta}=\tan^{-1}\left(\frac{v_{21}}{v_{11}}\right),\boldsymbol{h}=2\sqrt{\lambda_1},\boldsymbol{w}=2\sqrt{\lambda_2} \quad (2)$$

Where v_{11} & v_{21} be the two components from the largest eigenvector, also λ1 & λ2 are the two Eigen values.

The main novelties of the joint proposed scheme are as follows: 1) It uses set five parameter such as (width, height, central location and orientation) of rectangular box as a fully tunable variables; 2) By partitioning a rectangular box into sub-regions, deriving equations for multi-mode anisotropic mean shift; 3) An efficient approach for live learning of reference object distributions is employed; and 4) By relating the rectangular bounding box parameters with the mean shift can be estimated by applying Eigen decomposition, exploiting geometry of partitioned sub-regions, and using weighted average of parameters [3]. The real time moving object detection and tracking is designed to reduce the tracking drifts in offline as well as real time live videos scenes to tackle the problems of single object tracking and to provide further tracking robustness in terms of a) Long-term partial occlusions (poor imaging conditions) and b) Intersections of objects. An efficient moving visual tracking system can be employed by using particle filters with a small number of particles, and live learning of reference object distribution.

IV. PARTICLE FILTER TRACKING BY EMBEDDING MULTI-MODE MEAN SHIFT

This section describes the moving object detection and particle filter tracking through embedding the multi-mode anisotropic mean shift in it. Let the state vector, or particles in the particle filter are defined as the shape of bounding box.

$$S_t=[y_t^1\ \ y_t^2\ \ w_t\ \ h_t\ \theta_t]^T \quad (3)$$

These particles are initially generated according to the Gaussian distribution in under the Brownian motion model [5]. The key connection between the joint particle filter and mean shift tracking is lies in these two methods, where the object appearance within the bounding box is taken into account in for the estimation of posterior box shape by the state vector. Therefore, the state vector of particle filter is simplified to, $X_t=S_t$ which has a small size. Roughly speaking, initial particles first pass through a multi-mode anisotropic mean shift, resulting in the improved particles then by moving the box toward locations where the object of interest appears most similar to the reference object. The posterior estimates of rectangular bounding box for these improved particles are then used by the particle filter to obtain the posterior estimates of box.

A. Improved Particles and Likelihood

To associate the candidate object appearance with the likelihood (or the conditional pdf) of candidate box, the Bhattacharyya distance on the partitioned box described by the five component state vector S_t is defined according to the object appearance similarity [3]. For each initial particle $S_t^{(j)}$ generated under N (0, Q).

$$Q=diag(\sigma_{y(1)}^2,\sigma_{y(2)}^2,\sigma_w^2,\sigma_h^2,\sigma_\theta^2)$$

An improved particle is then replaced. The multi-mode mean shift estimates

$$S_t^{\sim(j)}=[y^1,y^2,w,h,\theta]^T \quad (4)$$

Assuming Gaussian distributed likelihood, the pdf conditioned by $S_t^{\sim(j)}$ (or the likelihood) from the mean shift is

$$P\left(z_t\middle|S_t^{\sim(j)}\right)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{d^{(j)^2}}{2\sigma^2}\right) \quad (5)$$

Where σ^2 the variance determined empirically, and $d^{(j)^2}$ defined as the Bhattacharyya distance for the jth candidate object within the box containing M partitioned sub-regions [3].

$$d^{(j)}=\sqrt{1-\sum_{i=1}^{M}\sum_u p_u^{i,(j)}}\,(y,\in)q_u^{i,(j)} \quad (6)$$

In such case updating the particle filter weights by using the likelihood obtained from the mean-shift

$$w_t^{(j)} = w_{t-1}^{(j)} P\left(z_t \middle| S_t^{\sim(j)}\right) \dots\dots\dots j = 1 \dots\dots\dots Np. \qquad (7)$$

Similarly, updating the posterior state vector of shape is replaced by the weighted average of mean shift predicted particles $S_t^{\sim(j)}$ as follows:

$$S_t = \sum_{j=1}^{Np} w_t^{(j)} S_t^{\sim(j)} \qquad (8)$$

Noting that these weights re-distribute particles according to the distance metric through exploiting the most similar object appearance through using the mean shift, rather than using the random sampling. This leads to more efficient utilizing of particles, hence a significantly reduction of the required number of particles. For example, employing particle filters to track a visual object requires that a state vector contains both the object appearance and shape [3]. Due to a large size state vector, a particle filter usually requires several hundred particles (e.g., ~800 in [9]) to achieve a reasonably good performance. By using particle filters for tracking shape and motion with embedded object appearance from the mean shift, the number of required particles is significantly reduced (Np ~ 16 is used in our tests).

V. ALGORITHM FOR JOINT TRACKING SCHEME

System Initialization: Choose an object region (for frame t=0), compute ***q0*** for the reference object, and compute $\propto_i, m_i, i = 1 \dots\dots\dots M$ for the given partition. **For** frame t =0, 1, ⋯, do:

i) Generate particles :$S_{t+1}^{(j)} \sim P(S_{t+1}|S_t^{(j)})$, set weights $w_t^{(j)} = \frac{1}{P}$; **For** Particles *j=1……Np* do:

ii) Mean shift iterations for $y_{t+1}^{(j)}$ using Eq. (10) from the reference [3].

iii) Iteratively compute: $\sum_{i,t+1,}^{(j)} i = 1 \dots . M$ Using Eq. (11) from the reference [3], until converge;

iv) Compute $w_{t+1}^{(j)}, h_{t+1}^{(j)}, \theta_{t+1}^{(j)}$ by Eq. (14) and (15) from the reference [2].

v) Update $S_{t+1}^{(j)}$in Eq. (4)

vi) Update the particle weight by w_{t+1}^{j}using Eq. (4) and (5).

END {j}

i) Compute the posterior estimate of the state vector s_{t+1} using Eq. (8).

ii) Resampling…

iii) If 't' is the end boundary of the interval sized L (i.e. mod (t, L) = 0) then live learning of q_i^j using Eq. (21), (23) of the reference [3] is then satisfied.

END {t}.

VI. EXPERIMENTAL RESULTS

Experiments have been conducted for set of real time dataset videos and it shows that how particular object can be tracked by the bounding box from the continuous segmented video frames as an input to the system. Following are some of the key performance and accuracy measurement parameters related to object tracking such as Recall, Precision, and F-measure shows that in how many number of frames object can be correctly identified and tracked.

Confusion Matrix:

		Object Detected Frames	
		Positive	**Negative**
Actual Considered Frames	**Positive**	A: True Positive	B: False Negative
	Negative	C: False Positive	D: True Negative

1. Precision: It is the proportion of predicted positive detected object cases that were correct and can be calculated using the equation,

$$Precision\ (P) = \frac{A}{A + C}$$

2. Recall: It is the proportion of positive detected object cases that were correctly identified and it also called as sensitivity/True Positive Rate (TPR), as calculated using the equation,

$$Recall\ (R) = \frac{A}{A + B}$$

3. F-measure: The F-Measure computes some average of the detected information retrieval precision and recall metrics. **Why F-measure?** An arithmetic mean does not capture the fact that a (50%, 50%) system is often considered better than an (80%, 20%) system, F-measure is computed using the harmonic mean:

Given n points, the harmonic x_1, x_2, and x3… x_n mean is:

$$\frac{1}{H} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{xi}$$

So, the harmonic mean of Precision and Recall:

$$\frac{1}{F} = \frac{1}{2}\left(\frac{1}{P} + \frac{1}{R}\right)$$

$$F = \frac{2PR}{P + R}$$

Example:

$$F = \frac{2 * 100 * 91}{100 + 91}$$

$$F = 95.30$$

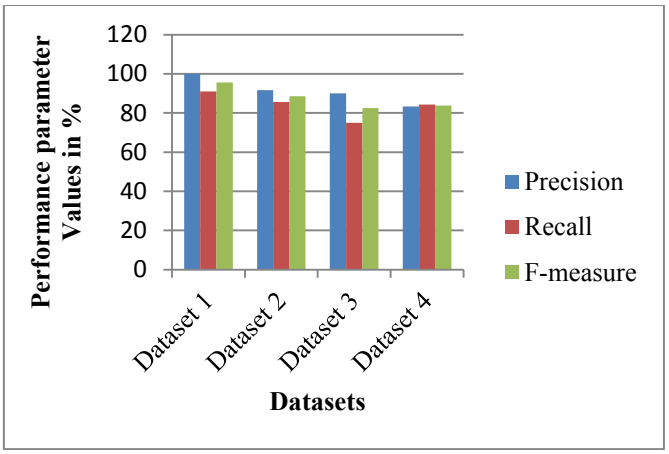

Fig.3. Performance Measurement of the System for Different Video Datasets.

RESULT TABLE I

	Precision	Recall	F-measure
Dataset 1	100	91	95.3
Dataset 2	91.6	85.6	88.6
Dataset 3	90	75	82.5
Dataset 4	83.3	84.3	83.8

A. Matlab Implementation of Moving Object Detection And Tracking Using Static As Well As Dynamic Camera For Offline Stored Video And Live Webcam Scene.

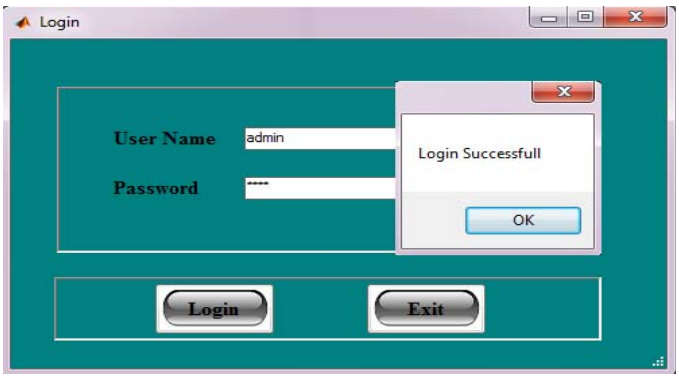

Fig.4. System Login Page

Moving object detection and tracking is one of the important first steps which has attracted a great interest from computer vision and image processing researchers due to its applications in areas, like video surveillance as shown in fig.4, traffic monitoring and image recognition. In case of real time moving object detection from the scene involves identification of an object in consecutive frames but in case of object tracking which uses to monitor the movements with respect to the region of interest of the object. Single object and multiple object movements in a frame with respect to the computed vectors are segmented with the help of specified threshold limit. The extracted movements are tracked using mean shift and particle filter algorithm. In our proposed MATLAB based object detection and tracking system according to algorithm mentioned in section V also fig. 5. Shows the obtained results in which offline stored video captured at railway station using stationary camera is fetched as input to the system. Before applying algorithm to input video is segmented into number of frames and that segmented video frame sequences are fetched to system. Thresholding is one of the most powerful and important tools in image processing of computer vision for image segmentation. The importance of segmented images obtained from thresholding has the advantages as it requires smaller storage space, performs fast processing and eases in manipulation compared with gray level image.

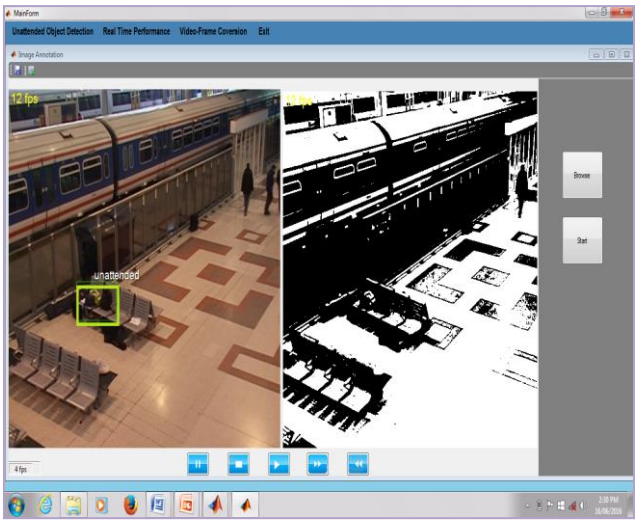

Fig.5. Unattended Object (Travelling Bag) Detection and Tracking at Railway Station from Segmented Continuous Video Frame Sequences (*Offline Stored Video Input to the System*).

Fig.5. shows the travelling bag from the scene as an unattended object is detected; as bag is not attended for long time then system administrator store its template (fig.7) for tracked object using system interface tabs and fires security alarm for tracked object. The results obtained from the four different video datasets is shown in result table I, from that we can see how the performance parameters are varied accordingly videos taken from different scenarios.

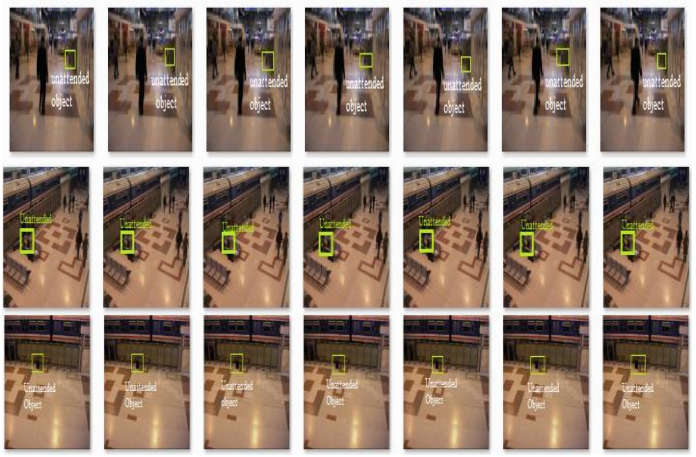

Fig.6. Four Different Dataset of Videos Captured at Railway Station from Different Scenarios.

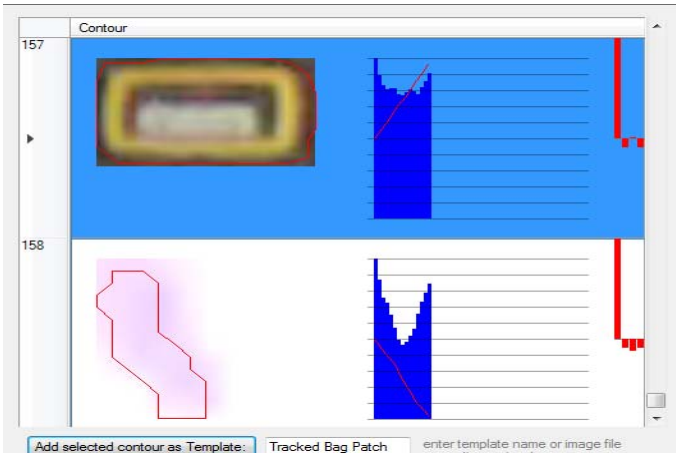

Fig. 7. Live Template Selection and Adding Along the Contour of an Object. (*Bag Patch as Template for Live Tracking*)

Real time moving object detection and tracking system that can processes 800*600, 640*480,and 320*240 resolution video sequences and provide the location of a predefined objects as shown in below (Fig. 8).

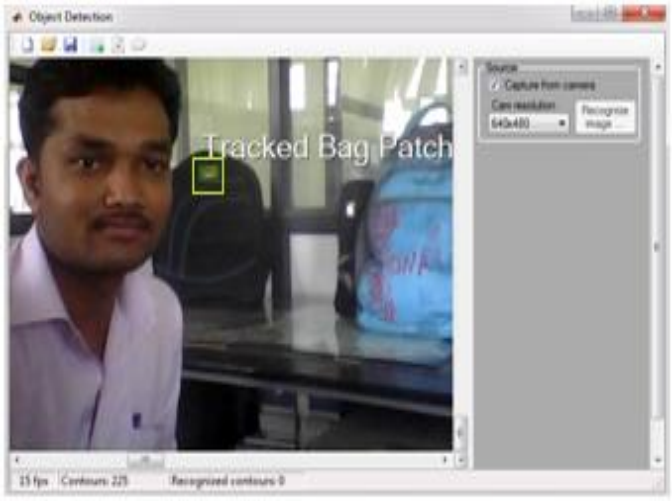

Fig.8. Live Object Tracking Using Webcam (*Real Time*).

In fig.7. Template for the bag patch is added so whenever bag will come across either stationary or moving camera our proposed Mean Shift and Particle Filter algorithm will be get executed in which dynamic shape of object along with rectangular bounding box is calculated by the particle filter, and the dynamic appearance is allocated by the mean shift that is embedded in the particle filter is as shown in fig.8.

CONCLUSION

In our MATLAB based live Object Monitoring, Detection and Tracking system we have tracked single object through multiple complex scenarios from videos captured by a single dynamic as well as stationary cameras with stationary or complex moving background. Test result shows that the real time moving object detection and tracking system is very robust which is resulting in considerable improvement by reducing somewhat tracking drifts and capable of tracking the object intersections, long-term partial occlusions in scene, object pose or shape changes, fast motion changes, and cluttered background. The proposed live as well as stored dataset tracking scheme nearly reaches its limit when video scenes contain too many similar objects in terms of their appearance distributions in video scene, frequent intersections of objects at background and occlusions. Other limitation such as the computational speed and tracking drifts of the system requires future improvement.

REFERENCES

[1] T. B. Moeslund, A. Hilton, and V. Kruger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, pp. 90-126, 2006.

[2] A. Hampapur, L. Brown, J. Connell, A. Ekin, N. Haas, M. Lu, H. Merkl, S. Pankanti, A. Senior, C. F. Shu, and Y. L. Tian., "Smart video surveillance," IEEE Signal Process. Mag., vol. 22, no. 2, pp. 38–51, Mar. 2005.

[3] Z. H. Khan, I. Yu-Hua Gu, and G. Andrew, "Robust visual object tracking using multi-mode anisotropic mean shift and particle filters", IEEE Trans. Circuits and System for video tech., vol. 21, no.1, January 2011.

[4] L. Wang, W. Hu, and T. Tan. Recent developments in human motion analysis. Pattern Recognition, 36(3):585–601, March 2003.

[5] Q. Sumin and H. Xianwu, "Hand tracking and gesture recognition by anisotropic kernel mean shift," in Proc. IEEE Int. Conf. Neural Netw. Signal Process, vol. 25. pp. 581–585, Jun. 2008.

[6] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 5, pp. 564–577, May 2003.

[7] T. Yang, S. Z. Li, Q. Pan, and J. Li, "Real-time multiple objects tracking with occlusion handling in dynamic scenes," in Proc. IEEE Conf. Comp. Vision Pattern Recognition, pp. 970–975, Jun. 2005.

[8] S. Zhong and F. Hao, "Hand tracking by particle filtering with elite particles mean shift," in Proc. IEEE Workshop Frontier Comput. Sci. Technol., pp. 163–16, Dec.2008.

[9] T. Wang, I. Y. H. Gu, A. Backhouse, and P. Shi, "Face tracking using Rao Black wellized particle filter and pose-dependent probabilistic PCA," in Proc. IEEE Int. Conf. Image Process, pp. 853–856, Oct. 2008.