CHAPTER-6

RESULTS AND DISCUSSIONS

6.1 Problem Statement

The idea behind our work is to perform different experiments on the work by Weiming Hu [7], and obtain answers for the following questions:

1. What would happen if we perform wavelet analysis on the frames and then apply foreground segmentation and tracking?
2. Another idea that struck our mind is that, if we increase the dimensions of our tensor by one then how will it affect the whole procedure?

6.2 Challenges faced

While performing the above mentioned experiments, we have confronted various challenges:

1. Software limitations, although the wavelet analysis has considerably reduced the size of the data to be analyzed , but due the introduction of a new dimension the whole method is still computationally expensive for the platform we worked on i.e matlab 7.0.6.
2. Another challenge in our work is related to the machine used, the earlier method i.e [7], is executed on a 3.25GB RAM and Intel Core2 Quad CPU at 2.83 GHz . Whereas, our work is executed on a 2 GB RAM, Intel Core2Duo CPU at 2 GHz. Our method is advanced to the earlier one, so the results would have been much more improved if some advance machine would have been used.

6.3 Our Proposed Foreground Segmentation method (3D-ITSL)

In this section we present details of the proposed 3-Dimensional Incremental Tensor Subspace learning algorithm i.e 3D-ITSL for object tracking. First, we propose an efficient method that incrementally updates an eigenbasis as new observations arrive, which is used to learn the appearance of the target while tracking progresses. Next, we describe our new approach for drawing particles in the motion parameter space and predicting the most likely object location with the help of the learned appearance model. Collectively, we show how these two modules work in tandem to track objects well under varying conditions. The proposed 3D-ITSL is shown in Figure 6.1.

Here we extend the method proposed by Weiming Hu et. al [6] , referred to as Hu's method which develops an incremental tensor subspace learning algorithm based on subspace analysis within a multi linear framework. The method uses the 3-order tensors for the scene representations which is performed in a block-wise fashion and the changes in the appearances over time are modeled by incrementally learning a low dimensional tensor subspace representation which is updated with the arrival of new image frame. This incremental tensor subspace learning algorithm is applied for the foreground segmentation and object tracking. An advantage of choosing tensors for the subspace representation is that it has the capability to capture the intrinsic spatiotemporal characteristics of the gray as well as RGB scenes. The proposed work deals with the foreground segmentation and object tracking directly on RGB scenes based on likelihood function that is constructed on the basis of the learned tensor subspace model.

We have brought two major changes in the work by Weiming Hu et. al [6], which is highlighted as under :

1. One modification and improvement in Hu's method is that the wavelet analysis of the image scenes is performed and the approximate wavelet image thus obtained is utilized to capture the spatial correlation of the image scenes. Thus, the novelty of our work relies on the fact that the above incremental subspace learning algorithm is applied on the wavelet coefficients rather than directly on the intensity values of the image pixels. The wavelet decomposed image generated has considerably smaller size as compared the original image which is the major conception behind this new modified method. With this modification in Hu's method, the following benefits are produced:

(i) The method is comparatively faster.

(ii) During foreground detection only the significant changes are detected.

(iii) The non-significant moving objects/background is neglected.

(iv) Lesser computational complexity.

In our method the level of the wavelet decomposition must be adjusted properly keeping in mind the size and details of the objects in the scenes. However the wavelet to be used should be decided according to the application. We apply wavelet decomposition in such a manner so as to preserve the spatio-color-temporal (SCT) characteristics of the image scenes. The tensor created is the 3^{rd}-order tensor of the same size as the original one. A tensor subspace is similarly created and incrementally learned. The wavelet coefficients of the newly arrived image is projected on the subspace created, and

we obtain the reconstruction error from it and foreground and background is separated based on the likelihood criterion computed using this reconstruction error. The background and the foreground thus obtained are reconstructed using the inverse wavelet transform in order to obtain the original sized image.

2. Another novel feature of our method is that in spite of using a third-order tensor as represented in Hu's method, we are using fourth- order tensor or in other words a 3-Dimensional tensor, containing spatial information, color information and temporal information in its three dimensions. With the increase in the number of dimensions of the tensor, more and more information could be captured and, thus better accuracy could be achieved. The proposed method is superior to existing techniques as it provides more realistic results that are in close proximity to the actual data for analysis. Thus, using this notion we have further developed another novel technique by increasing the number of dimensions of the Tensor image as compared to Hu's method by incorporating the color content of every pixel. The incremental subspace learning for 4-order tensors is performed to create a (4-dimensional) subspace over which the new test image tensors are projected in order to draw conclusions regarding whether a particular pixel belongs to the foreground or background. The background models created in this manner capture the intrinsic spatio-color-temporal characteristics of the scenes. The appearance information of the target is captured by the tracking algorithm, in which an unscented particle filter is used to estimate the optimal object state. With the consideration of the color content of the image scene, the foreground-background separation can be directly applied on the image. Unlike the channeled way where the input image scene is first converted into RGS form and then individually to each channel the incremental SVD is applied to create the corresponding subspaces. Finally, the corresponding reconstruction errors are combined to form an overall likelihood function and by using a likelihood function, a decision is then made regarding whether the pixel belongs to the foreground or background. Various experiments have been performed on different type of videos containing single and multiple moving objects having brightness and contrast variations, noisy and images from poor source. The results make it evident that our method is more robust to noise or low quality image (as we applied our algorithm to the wavelet compressed image that itself has very low quality), occlusions, lighting changes, scene blurring, lighting changes, scene variations.

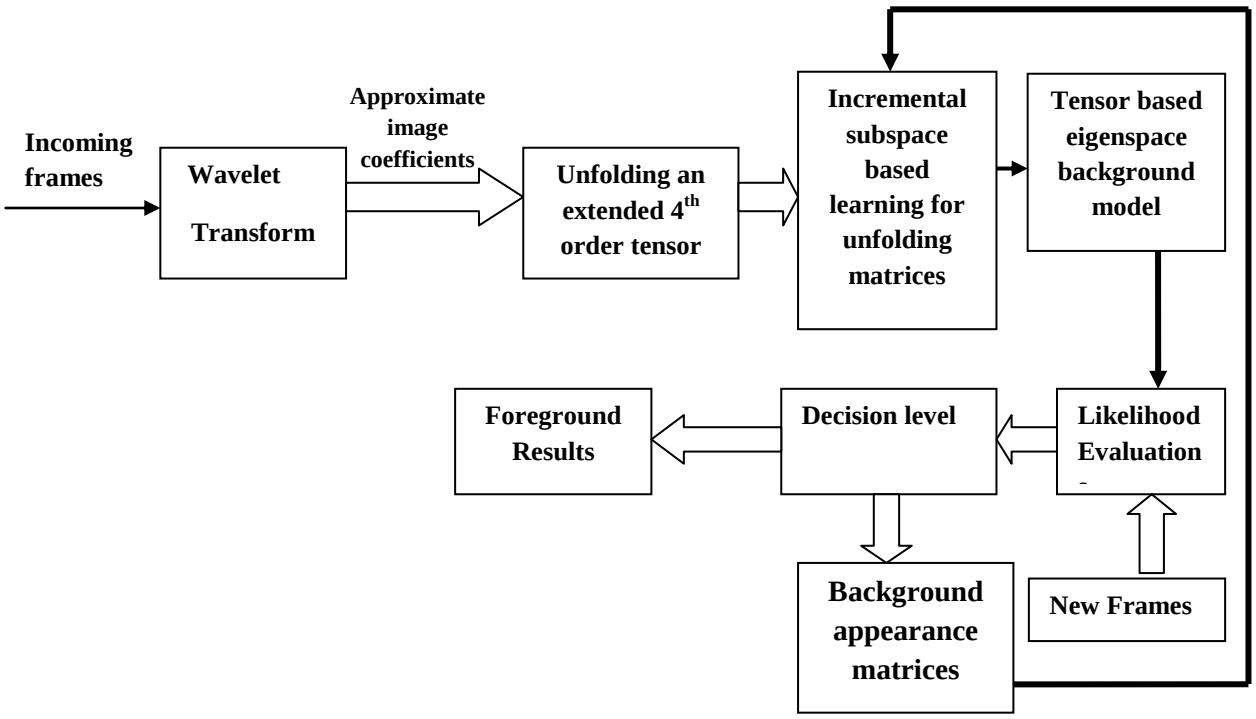


Figure 6.1 the proposed 3D-ITSL Block Diagram

Experimental evaluation for foreground segmentation as well as object tracking is done. The experiment is performed on six videos, in which first six corresponds to foreground segmentation and the next three are for object tracking. And, both sets are included under separate sections i.e section 6.3.1 for foreground segmentation and section 6.3.2 for object tracking. The experimental results for foreground segmentation are shown first followed by the tracking. The experiments are executed on Matlab 7.6.0(R2008a) installed on a computer with 2GB RAM and Intel Core2 Duo CPU.

6.3.1 Results of Foreground Segmentation

For the evaluation of performance of the proposed 3D-Incremental Tensor Subspace Learning (3D-ITSL) algorithm for foreground segmentation on color video sequences, three examples are chosen to demonstrate the behavior of our algorithm. The first three videos are chosen from the PETS2001 data base which is available on http://www.cvg.cs.rdg.ac. uk/slides/pets.html. The first video contains 145 frames and consists of 24-Bit color images. In this video the scene is captured as top-view, by fixing a camera on the ceiling which monitors the motion of the people moving in and out of the scenes every now and then. The second video consists of 159 frames. In this video person is walking in proper

lightened scene and vehicle enters and leaves the scenes. The third video consists of 389 frames. The video that is captured is of a shopping mall. The persons are walking in the corridors and their behavior is analyzed. All the three experiments are performed on the colored videos directly by representing its spatial, temporal and color information by a tensor, which in our algorithm is referred to as SCT information of the scene and thus creating a 3-Dimensional tensor which is actually viewed as a 4-D figure. The 3D tensor subspace based background model for color image sequences is updated after every three frame. The analysis is performed by employing 5x5x3x3 tensor obtained from the rectangular region centered at pixel of size 5x5 pixel across each image channel i.e R,G,B. Hence, in this example I1,I2 is 5 and I3,I4 are 3.

Example 1

In the first example, the input video worked on is the 'Walk1_80x120.avi' the subspace which consists of four dimensions are assigned values R1=3, R3, B3=3, R4=10. The scale factor σ in the computation of the likelihood is set to 15. The threshold value as mentioned is set to value 0.98. The frame rate for

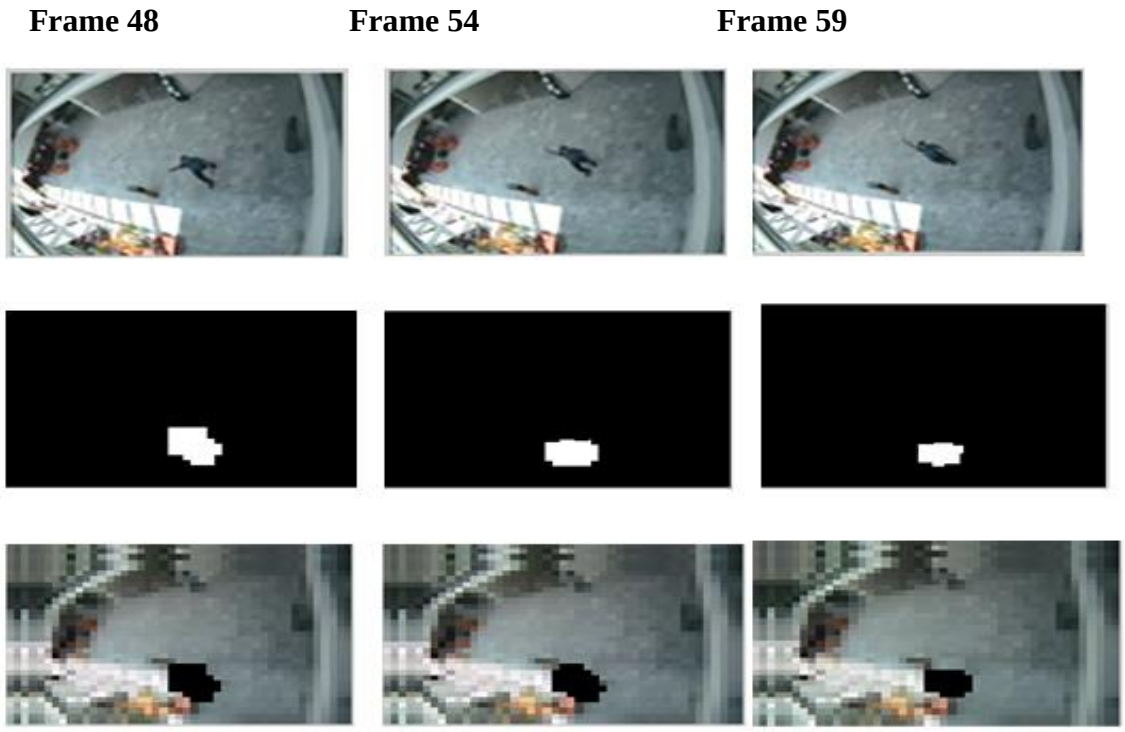


Figure 6.2 (i) The results of our proposed method in foreground segmentation.

this particular application is 5. Although our algorithm has lesser speed but it is still acceptable. The results are shown in Figure 6.2. In the results the segmented objects appears to be larger in size than the original object, the reason behind it is that the results has been in subjected to inverse wavelet transformation. If applied, then the size will come out to be the same.

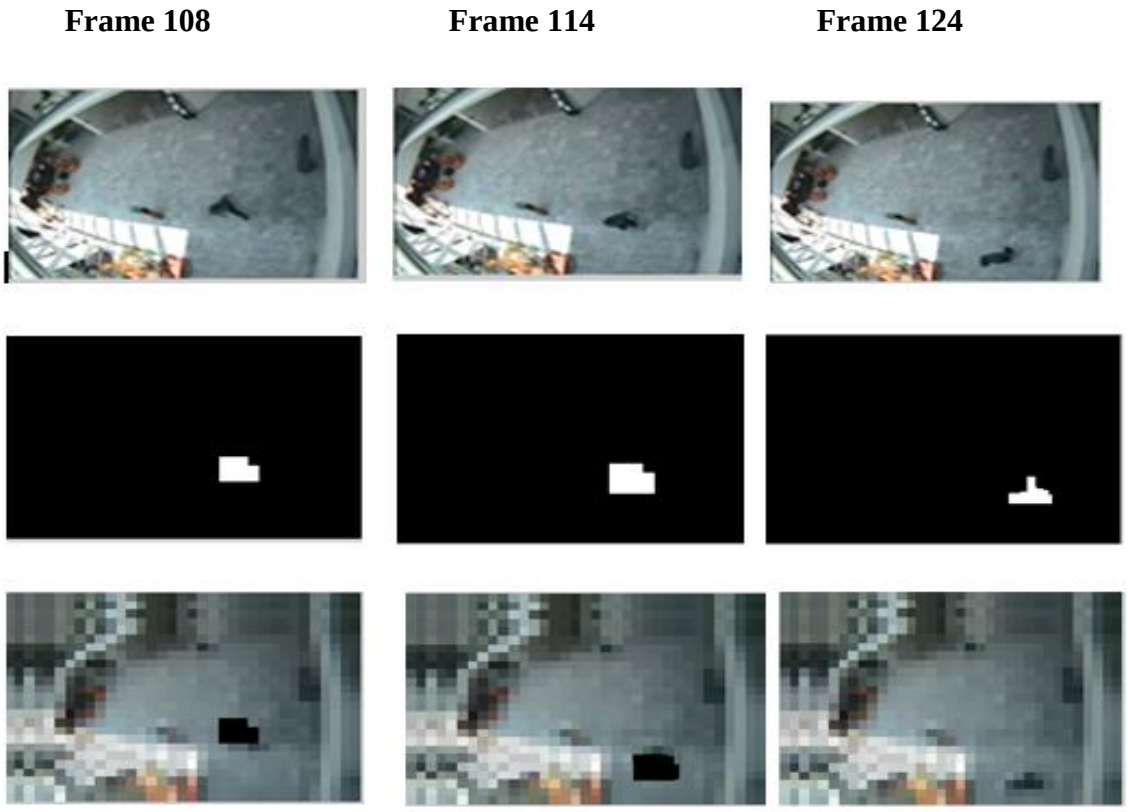


Figure 6.2(ii) The results of our proposed method in foreground segmentation

Example 2

In the second example, the input video worked on is the 'camera_2_112x160.avi' the subspace which consists of four dimensions are assigned values R1=3, R3, B3=3, R4=10. The scale factor σ in the computation of the likelihood is set to 5. The threshold value T_{gray} as mentioned is set to value 0.94. The frame rate for this particular application is 5. Although our algorithm has lesser speed but it is still acceptable. The results are shown in Figure 6.3. The frame rate for this particular application is 5.

Frame 48 **Frame 64** **Frame 104**

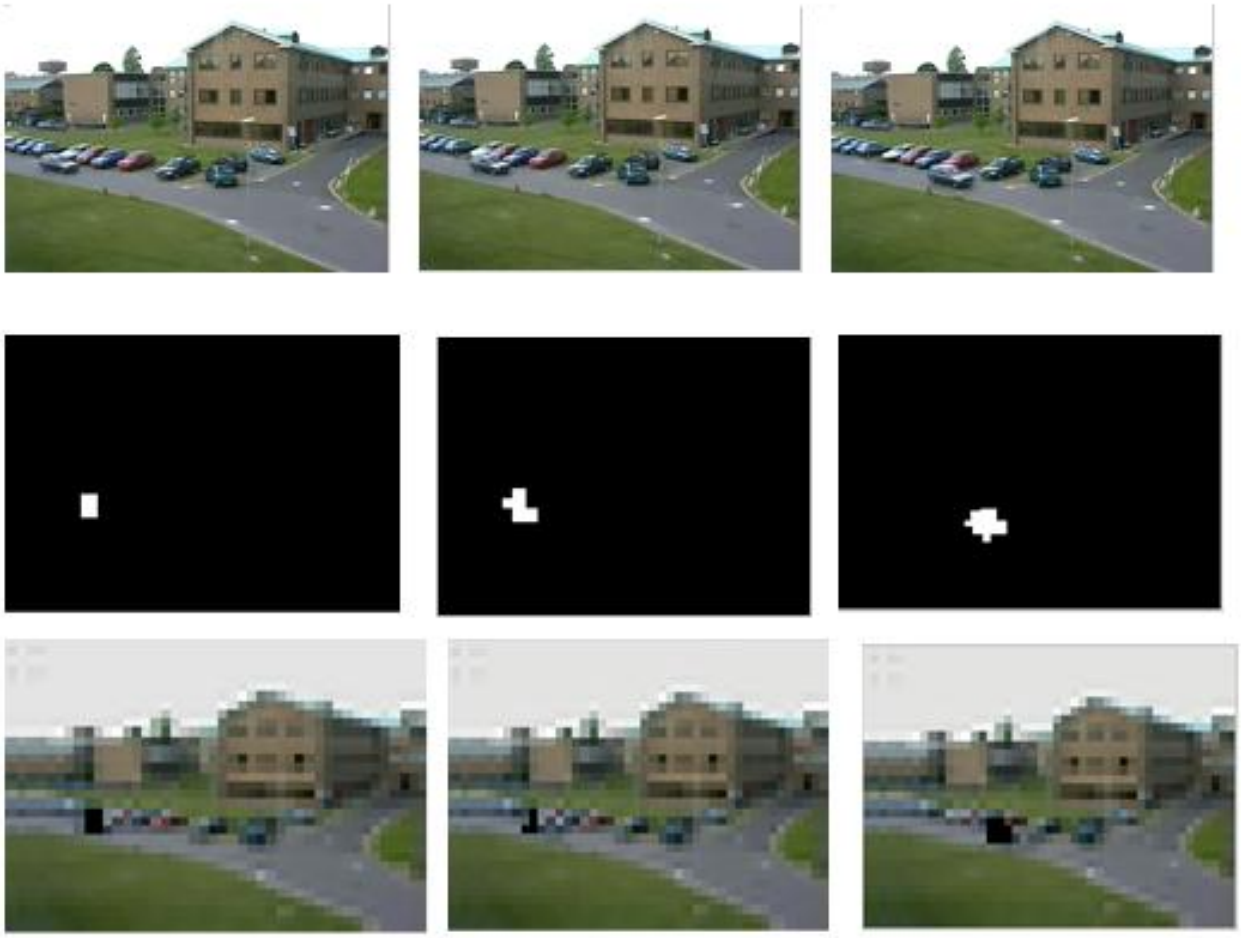


Figure 6.3(i) The results of our proposed method in foreground segmentation for example

Example 3

In the third example, the input video worked on is the "EnterExitspaths.avi" the subspace which consists of four dimensions are assigned values R1=3, R3, B3=3, R4=10. The scale factor σ in the computation of the likelihood is set to 15. The threshold value as mentioned is set to value 0.936. The frame rate for this particular application is 30. As the video contains very little and slow motion of the objects. Although our algorithm has lesser speed but it is still acceptable. The results are shown in Figure 6.3.

Frame 126 **Frame 141** **Frame 151**

Figure 6.3(ii) The results of our proposed method in foreground segmentation for example 2

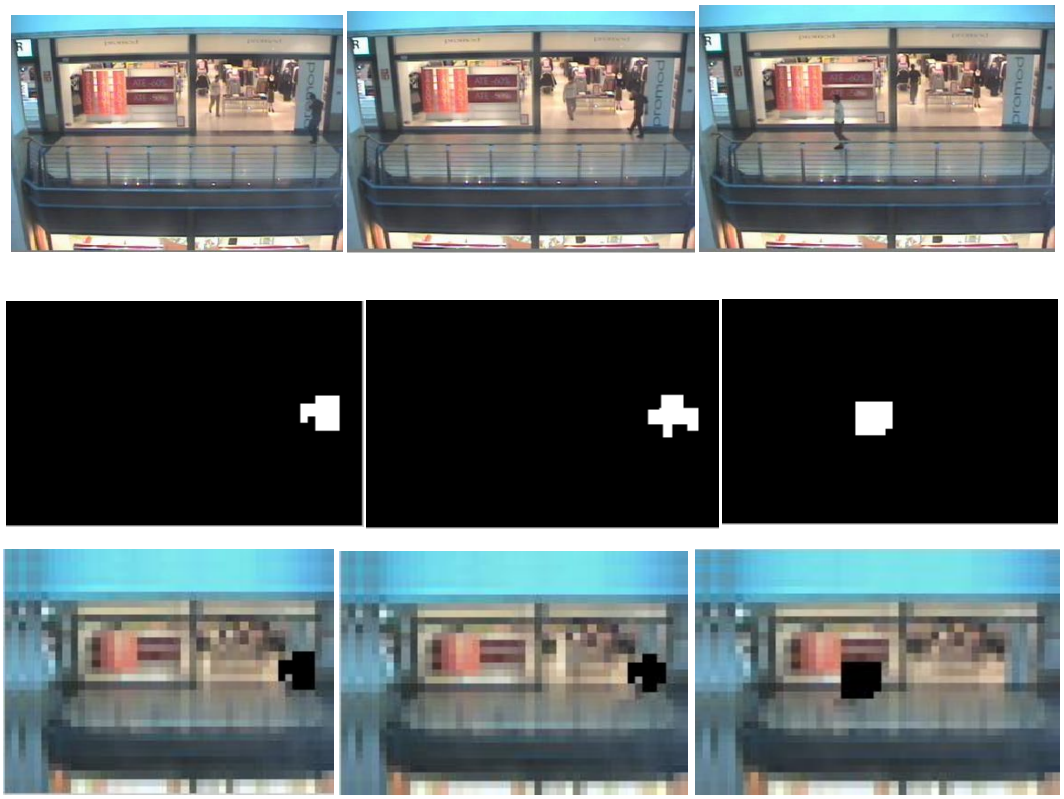


Figure 6.4(i) The results of our proposed method in foreground segmentation for example 3

6.3.2 Remarks

It can be seen that the foreground segmentation results of our algorithm are well indicated, connected for the objects, unique and quite different from rest of the results of the other algorithms due to the following factors:

1. Wavelet transformation is employed thus the non-significant objects doesn't play any role in foreground segmentation.

2. The shape of the object is not retained as the algorithm is applied on the approximated image.

3. Only the presence of the object is given importance, which is well obtained from our algorithm.

4. Despite of the application of foreground segmentation technique on the approximated image where much of the information of the actual image is lost (except the significant changes) the object is located accurately , because a new dimension of the information in the form of color variation has been incorporated thus aiding in obtaining satisfying results.

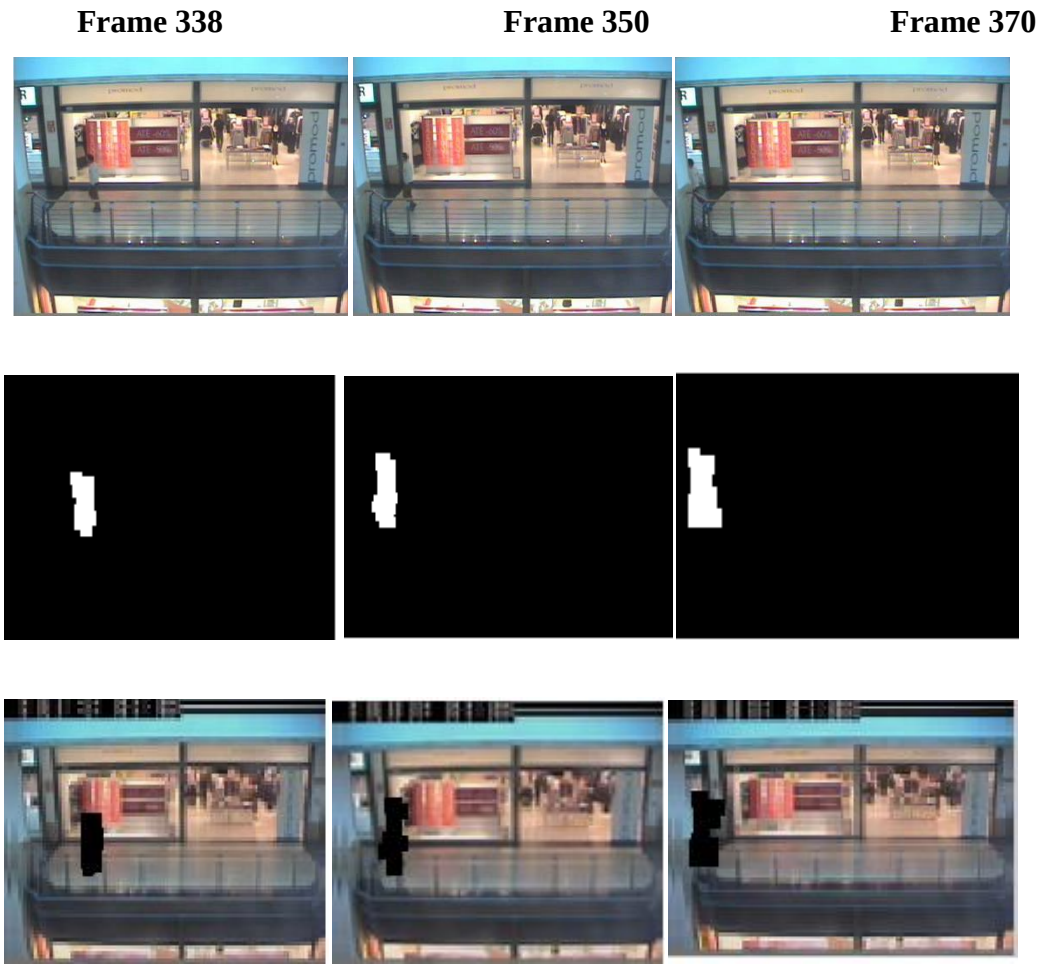


Figure 6.4(ii) The results of our proposed method in foreground segmentation for example 3

The algorithm is capable of being operated under noise as wavelet transform is being applied which can be regarded as an effective noise filter due to its localization property. And, using wavelet transform, the features in the original signal remains sharp. The averaging of the denoising results over all possible shifts is done to obtain the better results.

6.4 Our Object Tracker

To achieve the object tracking we apply the proposed three dimensional incremental tensor subspace learning algorithm to appearance-based object tracking. Figure 6.2 shows the architecture of our object tracking algorithm. In the algorithm, a low dimensional subspace model for the appearance tensor of an object is learned. The model uses the incremental 3-DITSL algorithm to find the dominant projection subspaces of the 3-order tensor of the object appearance. The current object state, which is

initialized according to the previous state and the dynamic model, is optimized using an unscented particle filter (details about particle filter). The appearance region specified by the optimal state is the tracking result which is used to further update the object appearance tensor subspace model.

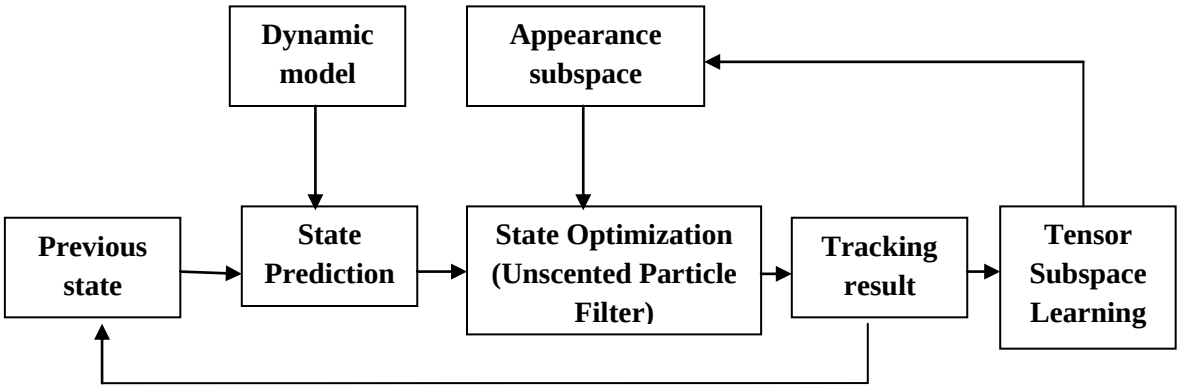


Figure 6.5 the Unscented Particle filter based tracker [6]

These four subspaces are combined via tensor reconstruction to form a tensor subspace based object appearance model. The likelihood for a candidate object appearance region given the tensor subspaces of the object appearance is obtained. The value of the likelihood is a measure of the similarity between the candidate region and the tensor subspaces of the object appearance.

6.4.1 Bayesian Inference for Tracking

A Markov model with hidden state variables is used for motion estimation. An affine image warping is applied to model the object motion between two consecutive frames. The object state variable vector $\boldsymbol{X_t}$ at time **t** is described using the Extrema points of the object of the affine motion transform, where the Extrema points of an object contains the following values

Extrema = [top-left top-right right-top right-bottom bottom-right bottom-left Left-bottom left-top]

So, $\boldsymbol{X_t}$ contains the 8x2 sized matrixes. Given a set of observed image regions $\boldsymbol{O_1, O_2, \ldots O_{t-1}}$, the posterior probability of the object state is formulated by Bayes' theorem:

$$\boldsymbol{p}(\mathbf{X}_t|\boldsymbol{O_1, O_2, \ldots O_t}) \propto \boldsymbol{p}(\boldsymbol{O_t}|\boldsymbol{X_t}) \int \boldsymbol{p}(\boldsymbol{X_t}|\boldsymbol{X_{t-1}}) \times \boldsymbol{p}(\mathbf{X}_{t-1}|\boldsymbol{O_1, O_2, \ldots O_{t-1}}) \boldsymbol{dX_{t-1}} \quad (6.1)$$

Where $\boldsymbol{p(O_t|X_t)}$ is the likelihood for the observation $\boldsymbol{O_t}$ given the object state $\boldsymbol{X_t}$, and $\boldsymbol{p(X_t|X_{t-1})}$ is the probability model for the object state transition. The terms $\boldsymbol{p(O_t|X_t)}$and $\boldsymbol{p(X_t|X_{t-1})}$ determine the tracking process. A Gaussian distribution is employed to model the state transition distribution $\boldsymbol{p(X_t|X_{t-1})}$

$$\boldsymbol{p(X_t|X_{t-1}) = N(X_t: X_{t-1}, \sum)} \quad (6.2)$$

Where $\sum$ denotes a diagonal covariance matrix with six diagonal elements $\boldsymbol{\sigma_1^2\sigma_2^2\sigma_3^2\sigma_4^2\sigma_5^2\sigma_6^2\sigma_7^2\sigma_8^2}$. The observation model $\boldsymbol{p(O_t|X_t)}$ is evaluated using the likelihood for a sample image region given the learned appearance tensor subspaces. An unscented particle filter is used to estimate the object motion state. The components of each particle correspond to the six affine motion parameters. For the maximum a posteriori estimate, the particle which maximizes the observation model is selected as the optimal state of the object in the current frame. The affinely warped image region associated with the optimal state of the object is used to incrementally update the tensor subspace-based object appearance mode.

6.4.2 Unscented Particle Filter

It is a variation of the particle filtering algorithm in which a particle filter uses an unscented kalman filter (UKF) in order to generate the importance proposal distribution. Unscented Particle filter is developed to address some of the short-comings of the extended Kalman particle filter:

(i) As the distributions generated by the UKF generally possesses a bigger support overlap with the true posterior distribution as compared to the overlap obtained from the EKF estimates. The reason behind it is that the UKF calculates the posterior covariance accurately to the 3rd order, whereas the EKF depends on a first order biased approximation and thus more accurately generates the proposal distribution particle filter framework.

(ii) The UKF also the capability to scale the approximation errors in the higher order moments of the posterior distribution like. Kurtosis, etc., and hence allowing for heavier tailed distributions. The unscented Kalman filter (UKF) is able to propagate more accurately the mean and covariance of the Gaussian approximation to the state distribution as compared the EKF.

(iii) The UKF also allows the particle filter to incorporate the latest observations into a prior updating routine.

6.4.3 Object Tracking Results

The performance of the tracking algorithm is evaluated and demonstrated on three videos and separately explained under these examples. The different scenarios include scene blurring, object occlusions, small apparent size, object appearance. The tracking is performed on the scenes, captured from the stationary camera is only considered. In the first example the scenes corresponds to the video captured from the top i.e a camera installed in the ceiling. The person enters in field of view (FOV) of the camera comes in the middle of the scene and goes back. The second example corresponds to a properly illuminated video, where a car enters in the scene and comes till the middle of the scene. In the third video there are two persons who cross their path at the entrance of the store and a couple is also present who is walking on the corridor.

The results of the foreground segmentation from the algorithm 3D-ITSL are used as the input to obtain the tracking results. An unscented particle filter, where number of particle is chosen to be 200 is used as the tracker. The eight diagonal elements σ^1, σ^2, σ^3, σ^4, σ^5, σ^6, σ^7, σ^8 in the covariance matrix $\sum$ (eq. 6.2) are assigned the values corresponding to the eight extrema points of the segmented object, given by

Extrema = [top-left top-right right-top right-bottom
bottom-right bottom-left Left-bottom left-top]

The extrema point of any particular object can be understood as under:

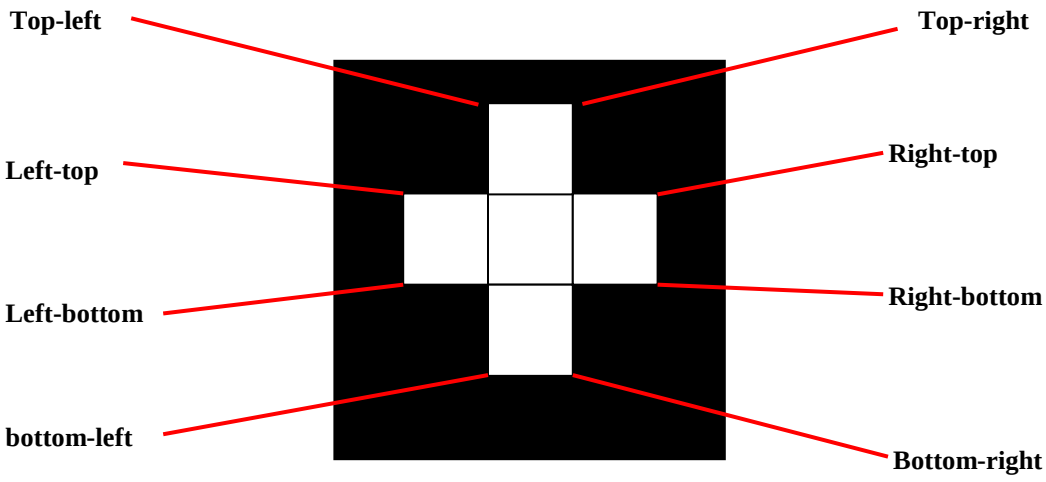
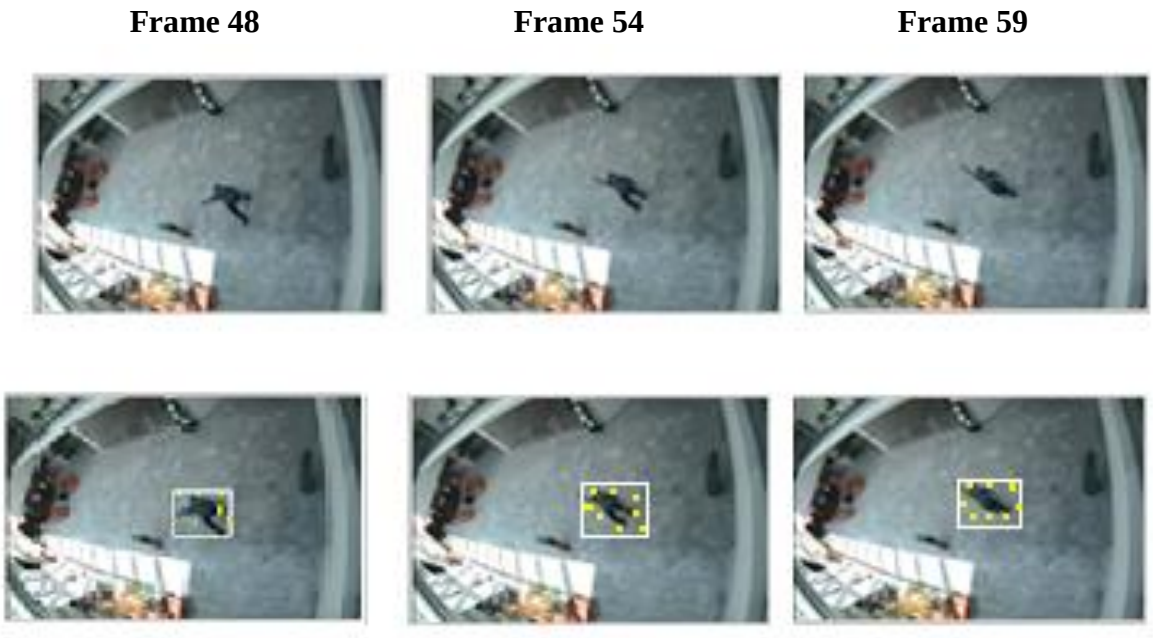


Figure 6.6 illustration of an object feature i.e Extrema points

A remark can be made about our tracking technique that, it wholly depends on the results obtained from the foreground segmentation algorithm. As the tracker runs on the results obtained from this previous stage. So, if the segmentation results are good the tracking performance will be good.

Example 1

The input video chosen is '**Walk_1.avi'**, the threshold T_{gray} is selected as 0.806. The parameter sigma in the likelihood computation is chosen to be 5. The four dimensional subspace created corresponding to our four dimensional tensor are assigned the corresponding Ranks as R1=3, R2=3, R3=3, R4=10. The results obtained are shown below; over the top of the every result the corresponding frame number is written. The first row corresponds to the actual video frame with the mentioned frame number. The second row is the result of tracking and the output has been highlighted with a square box. The actual result of our tracker is denoted by the yellow points marked in the videos as shown in Figure 6.7



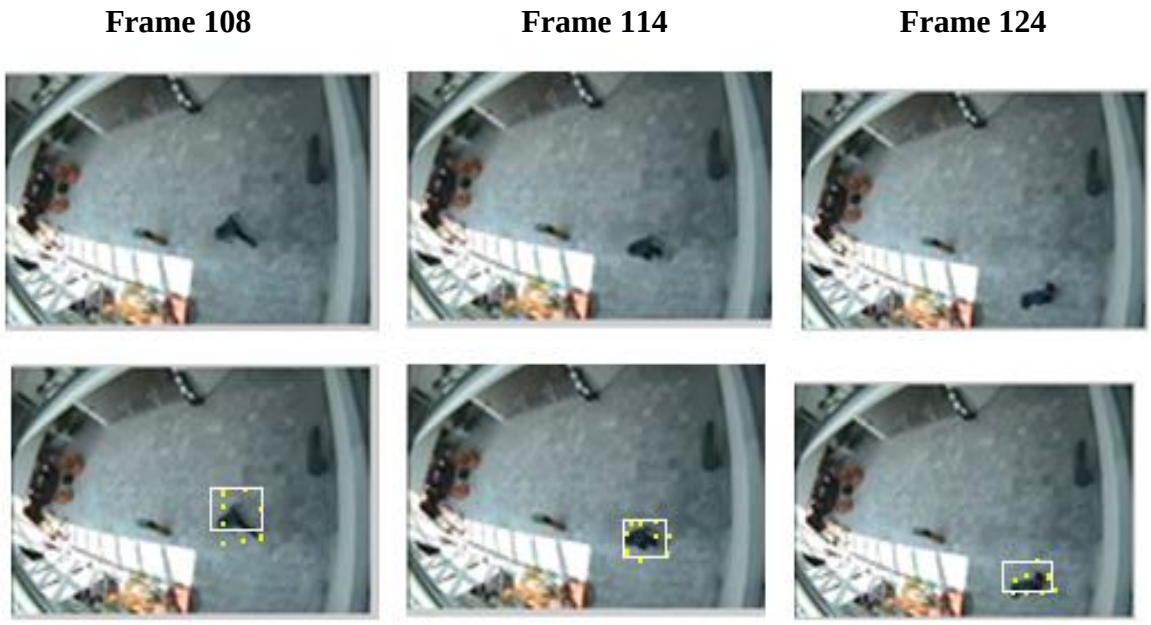


Figure 6.7 The result of tracking algorithm for the example 1

Example 2

The input video chosen is **'camera_2.avi',** the threshold T_{gray} is selected as 0.94. The parameter sigma in the likelihood computation is chosen to be 5. The four dimensional subspace created corresponding to our four dimensional tensor are assigned the corresponding Ranks as R1=3, R2=3, R3=3; R4=10. The results obtained are shown below, over the top of the every result the corresponding frame number is written. The first row corresponds to the actual video frame with the mentioned frame number. The second row is the result of tracking and the ouput has been highlighted with a square box. The actual result of our tracker is denoted by the yellow points marked in the videos

. Frame 48 Frame 64 Frame 104

Figure 6.8(i) Result of tracking algorithm for the example 2

Frame 126 Frame 141 Frame 151

Figure 6.8(ii) result of tracking algorithm for the example 2.

R2=3, R3=3; R4=10. The results obtained are shown below, over the top of the every result the corresponding frame number is written. The first row corresponds to the actual video frame with the mentioned frame number. The second row is the result of tracking and the output has been highlighted with a square box. The actual result of our tracker is denoted by the yellow points marked in the videos. The results are shown in Figure 6.9.

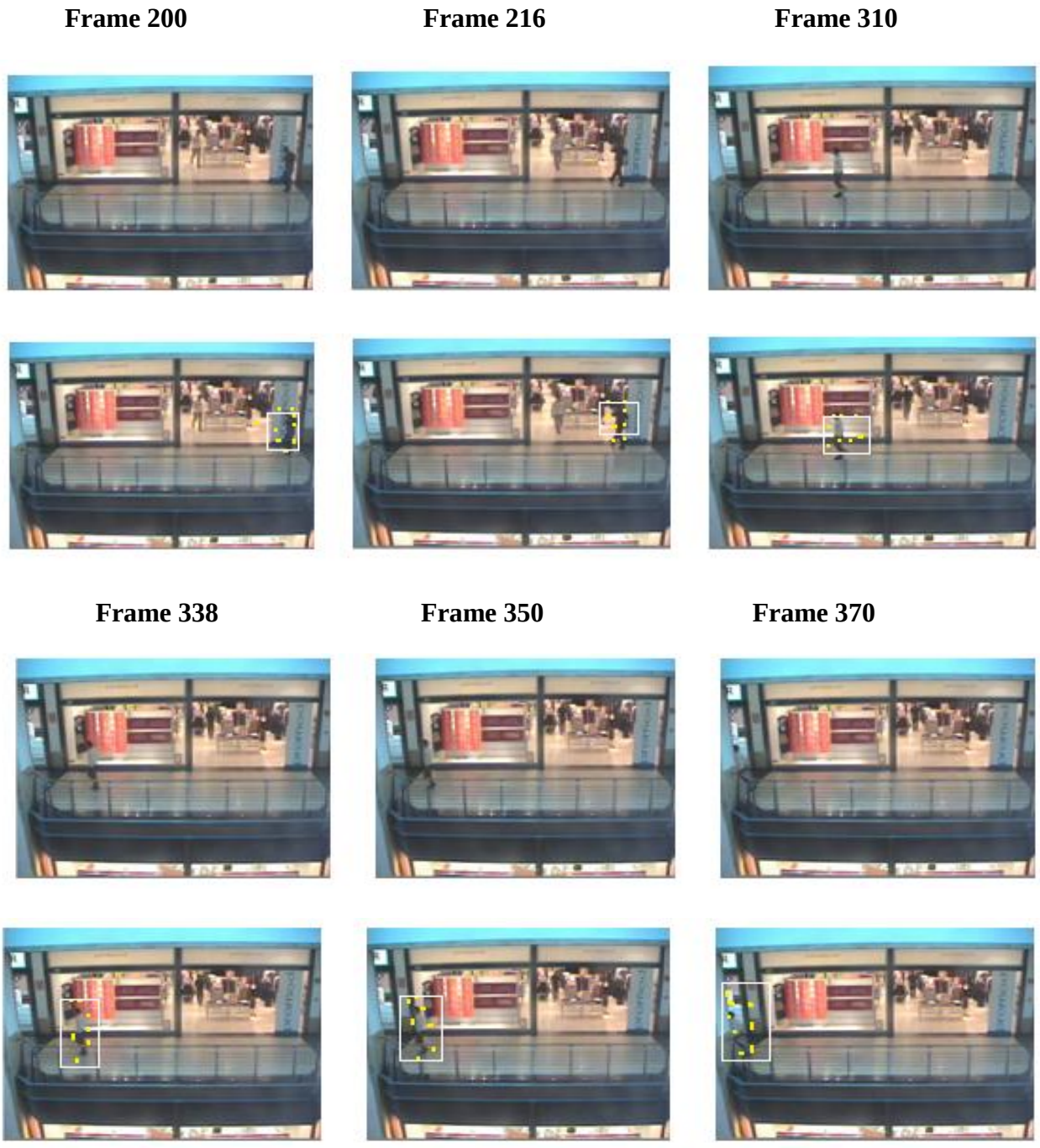


Figure 6.9 result of tracking algorithm for the example 3.

6.5 Conclusion and future scope

In this study we have presented robust appearance base foreground segmentation and tracking algorithm that learns incrementally a subspace of very lower dimensions using multi-linear analysis.

This subspace is kept on updating incrementally as the new image arrives. Our algorithm works directly on the colored video sequences rather than operating separately on the three RGB channels. Firstly the video is undergone from wavelet analysis, so as to remove the insignificant features and reduce the size of the image for lower computation. This subspace based algorithm which learns an ensemble of images in an online manner. The proposed method is named as **3D-ITSL**, as it operates on a 3-D tensor, containing spatial, temporal and color features and thus, very well captures the SCT interactions. A likelihood based classifier specified over the basis of the learned tensor subspace model. A hybrid tracker comprised of unscented Kalman filter and Particle filter is used (UKF-PF) that propagates the distribution of the samples and forms an unscented particle filter.

Experimental demonstrations in the last section proves the effectiveness and robustness of our algorithm to noise, ambiguous lighting conditions, occlusions, blurring effects, pose variations of the objects. Thus, our algorithm **3D-ITSL** is found to be efficient in the complex imaging scenes.

This is an advanced work presented here, but still it can be remarked that this current work has ample future scope, as we already pointed that computer vision area possesses a tremendous opportunity for growth and advancements. So, following can be suggested:

1. A further work could be done in the direction of utilizing some adaptive threshold for a better foreground segmentation results.
2. A further improvement in the quality of the technique could be achieved by extending the Tensor to few more dimensions, containing few more features like texture, pose etc.
3. Work could be done to create parallel processing scenarios in order to further reduce the computation complexity of the algorithm.
4. A better software and, machine could be employed to enhance its performance.