An Adaptive Particle Filtering Technique for Tracking of Moving Multiple Objects in a Video

Raksha Shrivastava
Department of Electronics and Communication
NRI Institute of Information Science and Technology
Bhopal (M.P)

Rajesh Nema
Department of Electronics and Communication
NRI Institute of Information Science and Technology
Bhopal (M.P)

ABSTRACT

Tracking moving objects in a video is of critical importance in various fields such as traffic monitoring, video surveillance, human motion capture, etc. However, tracking multiple objects in a video is very challenging. To meet that, authors proposed a new adaptive technique for object localization through local and global appearance of target. Local layer concentrates on the target's geometric deformation, i.e. the target structure is updated through adding and removing the local patches. The deformation information is constrained through the global layer, which concentrates on the shape, appearance, and color. The deformation information is passed from local to global layer through particle filter initialization and Hidden Markov Model (HMM). Particle filter is used to detect the local layer patches, and the sequence of deformation information is stored using HMM at global layer. This enhancement to the global layer improves multiple objects tracking efficiency. The efficiency of the proposed technique is evaluated through experimenting with a video containing multiple moving objects. Result analysis shows that the proposed method efficiently tracks multiple moving objects in the video.

Keywords

Multiple object tracking, Local layer, global layer, particle filter, HMM.

1. INTRODUCTION

Tacking an object in a video is a major task in the field of computer vision. The range of higher end computers, inventions of prominent high quality, inexpensive cameras, and the increased requirement for video analysis has generated a significant interest in object tracking algorithms. Normally, three steps are followed to analyze a video (1) interested movie objects are detected (2) track the detected object from frame to frame (3) recognize object's behavior. Following are the task where the object tracking is pertinent.

- Human-computer interaction
- Vehicle navigation
- Traffic monitoring
- Video indexing
- Motion-based recognition
- Automated surveillance

In simplest form, object tracking is referred as, track single or multiple objects over a sequence of image. A label can be assigned to the objects that are tracked in different frames of the given video. Based upon the tracking domain, the tracker can also offer object-centric information like shape, area, orientation of that object. Generally, the object tracking task is complex due to the following reasons.

- Noise presented in the image

- Motion of the object
- Non-rigid nature of the object
- Illumination changes in the scene
- Projection of 3D world on 2D image cause some loss in information
- Requirement for real time processing

The process of tracking is carried out in high-level applications which need the shape and/or location of an object in every frame. The tracking process can be simplified through assigning constraints on the appearance and on the motion of the objects. For instance, maximum of tracking techniques assumes the motion of the object is smooth with no sudden change. Different types of tracking mechanism exist, which are represented in Fig 1.
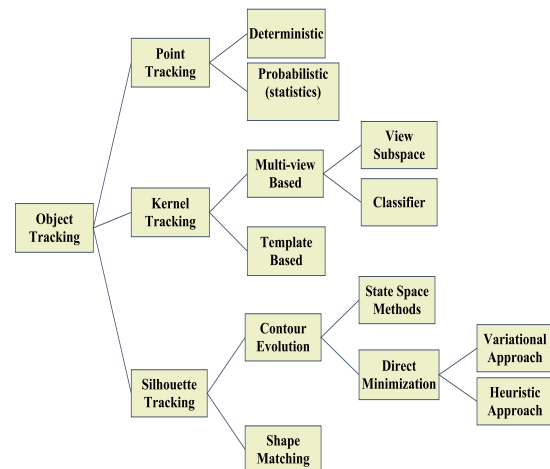

Fig 1: Types of Object Tracking

The detected objects can be represented in points in the consecutive frames of a video and the correspondence is depended on the preceding object state i.e. the motion and position of the object. Point association is a convoluted problem, especially in the presence of misdetection, occlusion, and entries and exists of the objects. The point correspondence problem can be further divided into two categories as (1) deterministic and (2) probabilistic. Qualitative motion heuristics technique is used by deterministic method to confine the correspondence problem. Whereas, statistical method uses object measurement and uncertainties to ascertain correspondence. Fig 2 represents an example for multipoint correspondence.
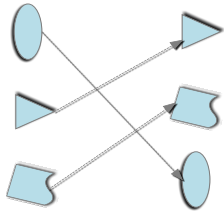
Fig 2: Multipoint correspondence

Kernel type object tracking refers to the appearance and shape of the object. Consider an example; the kernel can be either rectangular or elliptical template along with a related histogram. Objects are tracked through manipulating the kernel's motion in consecutive frames. Motion of the kernel is typically in parametric transformation such as rotation, translation and affine.

Parametric transformation of kernel pointing for rectangular parch is represented in Fig 3.
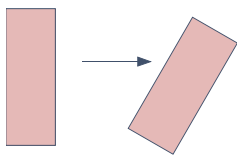

Fig 3: Parametric transformation

Encoded information that is presented inside the object region is used by silhouette tracking technique. The encoded information contains the details about the shape and appearance density that are normally in the form of edge maps. For a given object, either contour evolution or shape matching are tracked by silhouettes as shown in Fig 4. Aforementioned methods can be essentially considered as segmentation implied in the domain of temporal using the previous frames that are generated earlier.
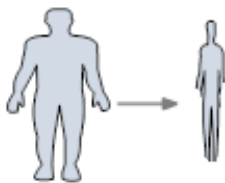

Fig 4: Contour evolution

Many researchers use holistic approaches, which globally tracks the object's appearance. Also this method is proven to be the successful technique. Even though, this technique faces problem during rapid changes in structural appearance changes since changes in appearance leads to drifting and reduced match that ultimately results in tracker's failure. To overcome this problem, techniques using set of local parts is suggested by various research scholars. Therefore, the combination of the global and local layers will works well in tracking the objects. Moreover, this technique can be used to track single object. Still there is a requirement to track multiple objects in an efficient way.

This paper focuses on tracking multiple objects in an efficient way. Authors proposed a new technique which consists of both the local and global layers. The local layer concentrates on the target's deformation whereas the information about the deformation of the objects is stored at the global layer. Information among these layers is passed through the particle filter initialization and Hidden Markov

Model (HMM). The combinations of these techniques are used to track multiple objects in a video effectively. Detailed explanation is presented in the corresponding subsections.

Rest of this paper is organized as follows. In section 2, the works that are related to the technique proposed in this paper are discussed. Tracking of multiple objects using the above said techniques are explained in detail in the section 3. A video stream is experimented using the proposed technique and their results are given in section 4. Finally, the work is concluded along with some ideas to enhance the proposed technique in future is presented in section 5.

2. RELATED WORKS

This section deals with the works that were related to tracking objects using various techniques. Here, the related works were presented in three different subsections. Each subsection explains the major methods used for tracking moving objects.

2.1 Point tracking technique

Initially, Greedy technique was used for tracking the moving objects, which was based on the rigidity and proximity constraints. This technique considered two consecutive frames and they were initialized by the nearest neighbours criterion. Algorithms based on this technique failed to handle occlusions, entries or exists. To overcome such problem, authors of [3] introduced a formulation of an optimization and the correspondence problem. This technique, established a correspondence for the points that were detected and also added hypothetical points for tracking missing objects. A different approach was introduced in [12] that have common motion constraints. This method was suitable for situations where conflict occurred during assignment. The framework also introduced a tracking algorithm based on greedy technique, which overcome the occlusion and detection errors. Moreover, automatic initialization of point tracking was given. With this, the overall cost required was reduced through using Hungarian assignment algorithm.

Tracking normal image sequence was not too difficult when they were compared with the noisy video. Huge research was carried out to track the objects that were presented in the image sequence that were noisy. To track such a noisy image many techniques were proposed few of them were discussed in this subsection. This section concentrates on the techniques proposed in the article [1, 2, and 5]. Authors of [5] proposed probabilistic MHT algorithm to reduce the communication overhead. This technique has an assumption on the association that, they were statistically independent random variables, which eliminates the need for exhaustive enumeration of associations. To track objects in noisy video [1, and 2] used Kalman filter. In [1] the filter is used for the recursive solution. Likewise, in [2] a probabilistic state space technique was proposed in order to provide optimal solution. In addition, assumptions were made on the noise and distribution, which was extensively used to manipulate the parameters of objects that were in motion.

2.2 Kernel Tracking

To represent the objects in an image sequence, [14] computed weighted histogram through the circular region of the frames based on the mean-shift procedure. Appearance similarity was minimized iteratively through comparing the window that exists around the location of hypothesized object and the histogram of an object. Similarity of the histogram was defined in terms of Bhattacharya coefficient. As the process consisted of multiple iterations, the value for mean-

shift vector was computed for all the iteration so that the histogram similarity was increased. The aforementioned process was repeated till the convergence was achieved that normally requires four to five iterations. Histograms were generated through the weighting scheme defined through a spatial kernel. This provides greater weight to the pixels that were closer to the center of the object.

KLT tracker was proposed in [6] that iteratively computed the translation of a given region centered on an interest point. Once the interest point's new location was obtained then the tracker computes the quality of the tracked path. The quality was computed through manipulating the affine transformations that exists among the patches that present in the consecutive frames. The tracking is continued only when the sum of the squared difference between the projected and current path is small. Otherwise the feature that was selected for tracking was eliminated. A new method was proposed in [8], which considers the whole image as a set of layer. It treats the background as single layer and considered individual layers for each moving object. The layers consist of the following parameters namely (1) shape priors, (2) appearance of the layer and (3) motion model. The background motion was first compensated for performing layering through projective motion. Then the probability of a pixel that belongs to a particular layer was computed. The computation was carried out based on the earlier motion and characteristic of the corresponding object. Pixels that were far away from the layer were assigned with the uniform probability that coincides with the background. Authors estimated the parameters separately due to difficulty in estimating it simultaneously.

Subspace based approach, i.e. eigenspace technique was proposed in [9] for computing the transformation from the image's object, that is, currently presented to the reconstructed image. Principal Component Analysis (PCA) was used to represent the subspace appearance of a frame's object. Followed by this, subspace constancy was computed to determine the difference between the input image and the reconstructed image through eigenvector. This process was carried out in two stages.

(1) Coefficient of subspaces were computed by fixing the affine parameters
(2) Depending on the newly computed subspace, parameters of affine were computed

Based on this process, tracking is performed until the difference between the projected image and the input image was minimized.

Scholars of [13] used Support Vector Machine (SVM) for tracking of the objects in an image stack. In general SVM for tracking was used to provide score to the test data, which was given to indicate the test data's membership degree to the positive class. SVM based trackers contain positive and negative examples. Positive examples were consists of the object in an image that was to be tracked. Whereas, the negative examples consisted of the remaining objects that were not to be tracked like background, etc. In [13] author reduced the score of SVM classification over image region for estimating the object's position. The details of background were explicitly incorporated into the tracker, which acted as an added advantage to this tracking mechanism.

2.3 Silhouette Tracking

Object's state was defined in terms of affine motion parameters and spline shape parameters in [10]. They are measured through computing the edges of an image in a direction that is normal to the contour. States of the objects were updated by using a particle filter. State variables were

computed from the contours to obtain the initial samples that are used by the filter. Current state variables were eliminated during testing phase by particle filter. The elimination process was carried out based on the observation of edges of the normal lines at the control points presented on the contour.

Curves of the given images were deformed to a position that was desired in a second image was introduced in [11]. Authors used PDE to deform the curves in the first image. With this deformation tracking was performed through extrapolating the velocities of the first into second equation. Here the curves of interest in both the images were used for tracking. Proposed technique in this paper can be applied to track the object and for sequential segmentation. A region based energy criterion was introduced in [7] for active contours. It also presented the estimation of its implications. In addition to that an optimized scheme was proposed for accounting external and internal energy in separate steps. Authors used heuristic based optimized technique.

The problems in tracking the non-rigid objects was considered in [4] and proposed a model-based tracking method. This technique used two-dimensional geometric model to localize the objects presented in all the frames of an image sequence. The method proposed here can track the objects that were far in the image of one frame to the next and also that has arbitrary motion. Research scholars of [15] addressed the problem faced by tracking through modeling the appearance of the objects that were moving. This method can track many moving objects with total and partial occlusion and permits to reacquire the objects that have been formerly tracked.

3. PROPOSED METHOD

The sequential particle filtering on patches is implemented as a general framework that includes two applications; i.e., video tracking using high-order Markov chains and distributed multiple-object tracking. Predication of local layer patches to initialize the adaption of visual model is required during tracking. The center of the target can be identified as a weighted average of the patches positions. The allocation of new patches in the local layer is constrained by global layer. Authors applied Markov Model (HMM) for global layer prediction and Particle filter for initializing the local layer. This initialization is done by particle filter which makes use of the information from HMM which predicts the location of patches from previous frame. To reduce the time for prediction, allocation of new patches in the local layer is done with the help of a probabilistic model (HMM) which was constrained by global layer which encodes the target's visual features. The particle filter localization combined with HMM prediction improves the prediction performance thereby limits the time consumption. The position of the detected objects can be stored using memory allocation, which is predetermined. Based on the particle filter information the HMM provides information to local layer patches which in turn extends to construct a global layer. Thus this cyclic communication decreases computational complexity there by reduces time consumption. Hence it is more robust than any other system. Fig 5, represents the overall flow of the proposed framework.

Following subsections describes the loading and conversion of video into frame, the implementation of particle filter, and HMM model.

3.1 Load and convert video into frames

The image sequence also named as image stack or video is usually created through collection of images, which are related by time such as frames in a movie or through slices.

The image stack is normally denoted as four-dimensional array, where each image in the image sequence is called frame. The frames that are contained in a video are of same size and they are concatenated along the four dimension. The function named mmreader in Matlab is used for loading and displaying the image sequence. The same function can be used to convert the video into frame. This object creates a reader object for image sequence files. The created object can read the image stack that is given as input. There are no peculiar restrictions on the extensions of video file. It supports all type of file names. Object created through the mmreader function is able to support the below properties.

- Bits/pixel of the image stack that is given as input
- Number of frames per second
- Duration
- Total number of frames in the give video
- Height and width of the frames in the video (measured in pixels)

These are the basic and essential properties required for the proposed process that are supported by the above function. In addition to the above mentioned property it also contains some other property.
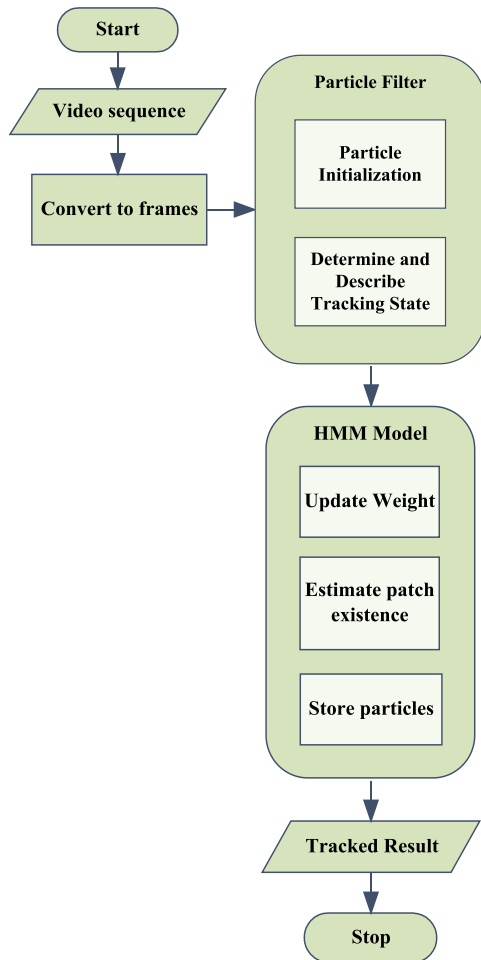

Fig 5: Overall flow of proposed Framework

3.2 Particle filter

Particle filtering is a hypothesis tracker, which is concerned with the problem of tracking multiple objects. This filter approximates the filtered posterior distribution by a set of weighted particles. In the proposed framework, local layer patches are initialized by particle filter through predicting the

location of the patches from previous frame. They are weighted based on a motion model.

The state of the system, x_t at time t is estimated using Markov model and mathematically represented as in equation 1.

$$E(t) = P(x_t|y_{0-t}) \qquad (1)$$

Here, $E(1)$ can be initialized using prior knowledge. According to Markov model, current state is independent of past and future state. Therefore, the observations are dependent only on the current state as expressed in equation 2.

$$E(t) = P(y_t|x_t) \cdot P(x_t|x_{t-1}) \cdot E(t-1) \qquad (2)$$

Where, $P(y_t|x_t)$ is the observation model and $E(t-1)$ is the proposal distribution. In addition to these parameters the proposed framework also requires the following attributes namely (1) Motion model (2) Observation model and (3) Initial model. Proposed framework, samples from the proposal and not the posterior information for estimation. However, the samples should be likelihood weighted by ratio of posterior and proposal distribution. Thus, weight of particle should be changed depending on observation for current frame. In the proposed method, particle filtering have used sequential Monte Carlo simulation for the below two reasons.

(1) For every particle at time t, the proposed method sampled from the transition priors
(2) For each particle, the framework proposed in this article evaluates and normalize the importance weights

To obtain it for N particles, multiply or discard particles with respect to high or low importance weights. The process of selection is used to track the multiple objects proficiently. The work flow of the particle filter present in the proposed frame work is portrayed in the Fig 6. Following steps carried out by the particle filter in order to accomplish the tracking of multiple objects.

(1) Initialize x_t for first frame
(2) Generate particle set of consisting of N particles i.e. $\{x_t^m\}_{m=1,2,\ldots,N}$
(3) Predict each particle (Use 2nd order auto-aggressive dynamics)
(4) Compute distance
(5) Weigh each particle depending on distance
(6) Select the location of target as a particle, which has minimum distance

These are the most important steps that are followed by particle filtering to track objects.

3.3 HMM Model

Global layer prediction is carried using HMM. Markov process used in the above filtering technique is not suitable for prediction at global layer since that is a simple stochastic process. Where, the distribution of future states depends only on the present state and not on how it arrived in the present state. The Markov model can be represented in finite state machine as in Fig 7.

As mentioned earlier, Markov process past state has no influence on present state. Let $\{x_t: t \ is \ in \ T\}$ be a stochastic process with discrete-state space S and discrete-time space T satisfying Markov property $P(x_{n+1} = j|x_n = i_{n-1}, \ldots\ldots\ldots, x_0 = i_0) = P(x_{n+1} = j|x_n = i)$ for any set of state $i_0, i_1, \ldots\ldots, i_{n-1}$, i, j in S and $n \geq 0$ is called Markov chain. In a hidden Markov Model (HMM), we don't know the state sequence. However we know some probabilistic function of it.

The allocation of new patches in the local layer is constrained by global layer, which encodes the target's global visual features. This process does not have a direct observation of the changes of the states. Therefore, this scenario can be referred as the model is hidden. However, framework is still able to observe an emission of the state changes and emission.
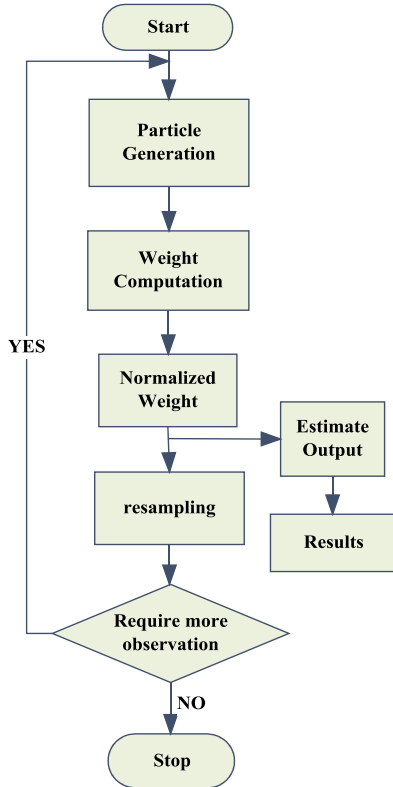

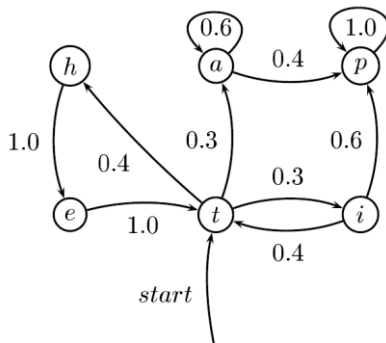**Fig 6: Flow of Particle Filtering**

Fig 7: Finite state machine

For this purpose it maintains a probabilistic HMM. HMM model uses the following representation.

- A set of states over time, denoted by **STATES**
- A set of *emissions*, or observations over time, denoted by **SEQ**
- An *M*-by-*M transition matrix TRANS* whose entry (i, j) is the probability of a transition from state i to state j.
- An *M*-by-*N emission matrix EMIS* whose i, k entry gives the probability of emitting symbol s_k given that the model is in state i.

Information about the patches presented in the frames is stored in memory, which is already predefined.

Different patch information is stored separately and they are tracked using the HMM model. HMM model generally represented using the finite machine; states present in the finite state machine are used to track the information of different patch information of a video. Memory locations of stored patch information are denoted as states in the finite state machine. The transaction details are denoted in the edges that exist between the states. To track the previous step i.e. the previous frames patch information, the states are analyzed to locate the memory location of the required patch information.

4. EXPERIMENTAL RESULTS

Performance of the proposed framework is analyzed using a video sequence that contains objects that undergoes a momentous appearance change. This framework is implemented in Matlab and runs at 30 frames per second on an Intel Core 2 Duo processor. Initially, the input video is loaded and converted into frames. These frames are processed by the proposed framework to track the moving objects. The input image sequence has 101 frames that run in 26.2seconds. Fig 8, shows the tracking of moving objects.
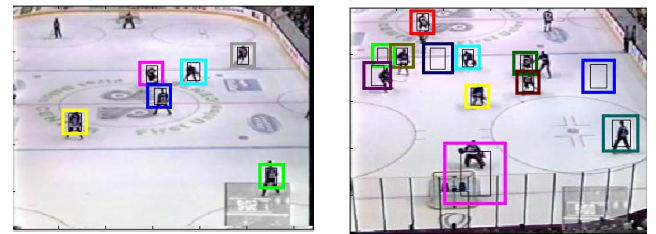

Fig 8: Tracking of moving objects, frames 5 & 91 are shown

The efficiency of the proposed framework is analyzed through two different criteria namely (1) Time and (2) False reduction.

4.1 Time analysis

Time taken to detect and track the multiple objects for all the frames (i.e. for full image stack) are computed and expressed in the Fig 9. The computation of time is calculated in milliseconds for a given football match image sequence.
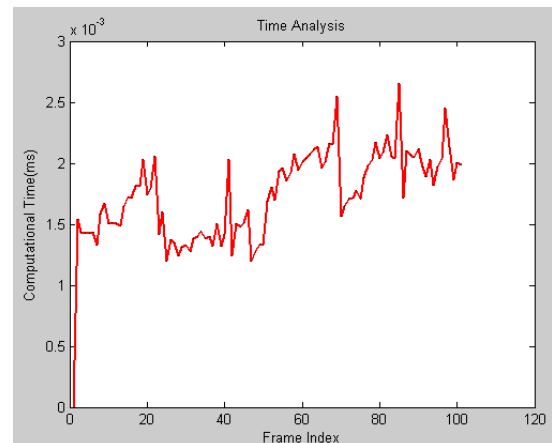

Fig 9: Time analysis for input video

Time computations are the major criteria for measuring the efficiency of the techniques. The algorithm that requires less time are considered to be the most successful and proficient method to analyze the video. With this concern the

proposed framework is analyzed based on the time it consumed to track the moving objects in the given input image stack.

The time analysis is also compared with other algorithms such as MIPF, MFMC, and GMOT. The time required for various algorithms are computed for the same football match video sequence and the results are shown in the table 1.

Table 1. Time analysis comparison

Technique	Time (ms)
MIPF	19.781
MFMC	29.366
GMOT	20.170
Proposed	1.77311

It is explicit from table 1 is that; the MIPF requires 19.781 milliseconds to track the moving object in the given video. Likewise, MFMC and GMOT need 29.366 and 20.170 milliseconds to track the objects in the video respectively. Whereas the proposed frame work requires 1.77311 milliseconds for tracking the objects in the same image sequence. The time analysis results also express that the proposed technique requires less time than the existing techniques. Therefore, from the time aspect (criteria) proposed system functions better and provide results that are efficient.

Table 2. GMOT-False Reduction

Frame Number	Objects	True Positive	False Positive
1	6	5	3
2	4	4	3
3	5	4	4
4	4	3	6
5	4	4	3
6	3	3	2
7	3	2	4
8	3	4	3
9	4	3	4
10	6	5	5
11	4	2	3
12	5	4	2

4.2 False reduction

There are various chances to track false objects due to information loss, occlusion, etc. The false reduction analysis helps to determine the accuracy of tracking of the moving object of an algorithm. The false reduction analysis results for GMOT algorithm is shown in table 2.
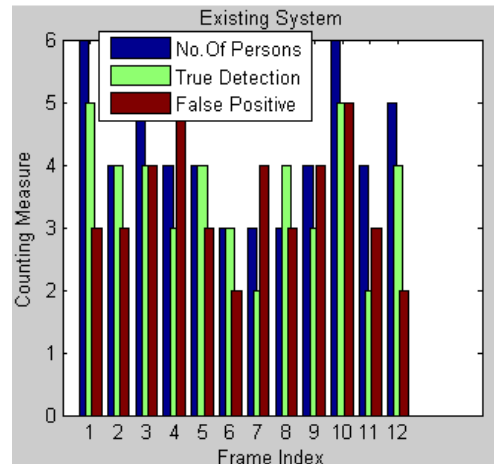

Fig 10: False Reduction – GMOT (Existing)

The table represents the same of twelve frames and their true positive and false positive rate for the objects present. Objects represent the total number of persons in the frame. True positive represents the number of objects that are correctly identified in a particular frame, where as the false positive denotes the number of objects that are recognized incorrectly in the same frame. GMOT algorithm has an average rate of 3.5 for false positive, i.e. GMOT recognizes 3.5 objects wrongly as an average. Likewise, it detects the correct object at an average rate of 3.6 in a single frame. Fig 10 express the false reduction for GMOT.

Table 3. Proposed Framework-False Reduction

Frame Number	Objects	True Positive	False Positive
1	9	6	0
2	10	7	0
3	9	6	0
4	10	7	0
5	9	7	0
6	10	7	0
7	12	9	1
8	12	9	2
9	12	9	2
10	12	9	2
11	12	12	3
12	12	12	3

Similarly, true positive and false positive values are computed for the proposed framework. The results are shown in the table 3. The computations are carried out for the same 12 frames of football match image sequence as taken for the

GOMT technique. The proposed technique has the average true and false positive values as 8.3 and 1.08 respectively. Fig 11 portrays the number of objects that are correctly and incorrectly identified by the proposed framework.
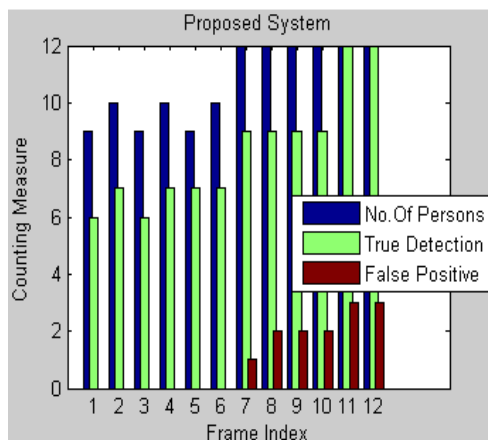

Fig 11: False Reduction – Proposed Framework

It is explicit from the Fig 10 and 11 that the proposed framework detects the correct objects efficiently than the existing technique. The number of objects identified by the proposed method is also higher than the existing technique.

The results for time analysis and false reduction comparison shows that the proposed technique tracks the moving objects by consuming very less time than the existing techniques namely GMOT, MFMC, and MIPF. Time analysis results implicitly explain the efficiency of the proposed framework. In addition to that, false reduction comparison express that the proposed system accurately determines the maximum number of moving objects than the existing system.

5. CONCLUSION

This paper concentrates on an adaptive framework to track the objects. This technique used local and global layer information for tracking multiple objects. Geometric deformations are carried out in local layer and the information regarding the deformation is constrained in the global layer. In order to perform the tracking process the information of deformation at the local layer should be passed to another layer. This is achieved through implementing particle filter and HMM model at local and global layer respectively. The proposed framework is experimented using an image sequence of football match containing more moving objects. The framework is tested based on the time and false reduction in order to determine the efficiency and accuracy in detecting the correct object. Result shows that the proposed technique accurately and proficiently determines the moving objects better than all the existing methods. This framework identifies only few of the objects that are moving in the video not all the objects. Therefore, determining all moving objects in an image stack is considered as future work of the authors.

6. REFERENCES

[1] BROIDA, T. AND CHELLAPPA, R. 1986. Estimation of object motion parameters from noisy images. IEEE Trans. Patt. Analy. Mach. Intell. 8, 1, 90–99.

[2] BAR-SHALOM, Y. AND FOREMAN, T. 1988. Tracking and Data Association. Academic Press Inc.

[3] SALARI, V. AND SETHI, I. K. 1990. Feature point correspondence in the presence of occlusion. IEEE Trans.Patt. Analy. Mach. Intell. 12, 1, 87–91.

[4] HUTTENLOCHER, D., NOH, J., AND RUCKLIDGE, W. 1993. Tracking nonrigid objects in complex scenes. In IEEE International Conference on Computer Vision (ICCV). 93–101.

[5] STREIT, R. L. AND LUGINBUHL, T. E. 1994. Maximum likelihood method for probabilistic multi-hypothesis tracking. In Proceedings of the International Society for Optical Engineering (SPIE.) vol. 2235. 394–405.

[6] SHI, J. AND TOMASI, C. 1994. Good features to track. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 593–600.

[7] RONFARD, R. 1994. Region based strategies for active contour models. Int. J. Comput. Vision 13, 2, 229–251.

[8] CASELLES, V., KIMMEL, R., AND SAPIRO, G. 1995. Geodesic active contours. In IEEE International Conference on Computer Vision (ICCV). 694–699.

[9] BLACK, M. AND JEPSON, A. 1998. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. Int. J. Comput. Vision 26, 1, 63–84.

[10] BLAKE, A. AND ISARD,M. 2000. Active Contours:The Application ofTechniques from Graphics,Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion. Springer.

[11] BERTALMIO, M., SAPIRO, G., AND RANDALL, G. 2000. Morphing active contours. IEEE Trans. Patt. Analy. Mach. Intell. 22, 7, 733–737.

[12] VEENMAN, C., REINDERS, M., AND BACKER, E. 2001. Resolving motion correspondence for densely moving points. IEEE Trans. Patt. Analy. Mach. Intell. 23, 1, 54–72.

[13] AVIDAN, S. 2001. Support vector tracking. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 184–191.

[14] COMANICIU, D. 2002. Bayesian kernel tracking. In Annual Conference of the German Society for Pattern Recognition. 438–445.

[15] KANG, J., COHEN, I., AND MEDIONI, G. 2004. Object reacquisition using geometric invariant appearance model. In International Conference on Pattern Recognition. 759-762.