

Critical Performance Analysis of Object Tracking Algorithm for Indoor Surveillance using modified GMM and Kalman Filtering

Navneet S. Ghedia^{1*}, Dr. C.H. Vithalani² and Dr. Ashish Kothari³

¹*Research scholar, Gujarat Technological University, Ahmedabad, Gujarat, India.*

²*Professor and Head of EC Dept., Government Engineering College, Rajkot, Gujarat, India.*

³*Associate Professor and Head of EC Dept., Atmiya Institute of Technology, Rajkot, Gujarat, India.*

Abstract

Our objective is to ensure high level of security in public places using static PTZ camera and robust detection and tracking algorithm for video sequences and also to generate the multi model background subtraction approach that can handle dynamic scenes. In this paper we are focussing to design the robust foreground and background detection technique using the statistical approach and implement it on the different indoor environments. Among all familiar background subtraction approach our proposed method used Gaussian mixture model and predictive filters to detect and track objection successive frames. To test the performance of our algorithm we will take the standard test sequences and own datasets. We will analyze our algorithm with the qualitative and quantitative approaches. so, our aim is to develop such an smart surveillance system that can not only analyze but also to interpret and act with reference to the object behaviour against illumination changes, clutter background, moving background, oclusions and complex silhouette.

Keywords: background model, gaussian mixture model, predictive filter, smart video surveillance system

I. INTRODUCTION

Visual monitoring system is the most research topic in today's era. Visual observation in computer vision helps to analyze object behaviours easily. For the indoor as well as outdoor video sequences the higher level real time smart visual surveillance system is required. A major part of the smart video surveillance system is characterized by perception, the robustness of a smart video surveillance system is not only to sense the environment but also to interpret and act accordingly in an intelligent way. Advancement made in perception leads to several application such as military and driving assistance. Now 'a days the biggest global challenge is to ensure the public places so the researchers are now focused more and more on object detection, object tracking , crowd analysis, Pedestrian detection and vehicle identification [20]. Object location, silhouette of object, object classification and the activities carried out by the object is a vital in scene analysis. The robust tracking depends on the robustness of the foreground detection. The role of the object tracking algorithm is to estimate the trajectory path which is compare against the ground truth, in a subsequent frames. The tracking task is challenging because of certain issues are associated with the objects such as, abrupt motion of the object, complex silhouette, illumination changes, occlusions etc,. The key role of the tracking algorithm is to track all objects against all constrains. Kalman filtering, particle filtering, template matching and silhouette are some of the well known approaches for object tracking [3]. Under the assumption that our detection model is linear and the system noise and posteriori distributions are Gaussian in nature, among all tracking approaches prediction filter approach gives better accuracy.

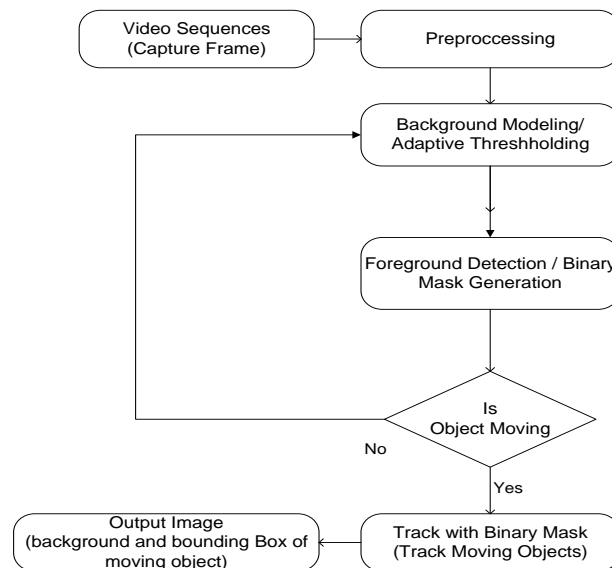


Figure 1 block diagram of object detection and tracking system

Figure 1 shows the important steps of the object detection and tracking system. Generally we are assuming that the static background is available in a video sequences but for the real time processing background frames are not available so for the initialization subsequent frames are required to generate the background. The next step is to preprocess the input frames, such as morphology, image resizing and the edge detection is required. Morphological approach will not only reduce noise in the moving object but also fill the gaps in it. The dilation approach will convert each background pixel into foreground if it touching with an foreground while erosion will convert each foreground pixel into background if it touching with the background. So in this way it will reduce the noise and fill the gaps. The next will be to estimate the background modeling which can ensure to handle any dynamic scene. Segmentation followed by mask generation will give us a robust detection. Object tracking is the final step of every surveillance system and on the basis of the performance analysis system can act intelligently.

II. RELATED WORK

Most researchers have adopted non adaptive approaches for the surveillance, such an approach fails the dynamic scenes and for the real time applications. Stauffer et.al. [1] proposed a very popular mixture of Gaussian approach. They have used multimodal GMM for the dynamic scenes their approach fails to detect object in sudden illumination and also it will take more computational cost. Bowden et.al. [6] proposed the system which overcome the demerits of [1] slower learning rate. Harville et.al. [11] Proposed a colour and depth based segmentation approach. The proposed algorithm is suffered with slower learning rate because of the traditional MoG approach. Harville et.al [12] proposed extended GMM approach for the detection and tracking purpose. Feedback in the algorithm will change the learning rate and it is applicable to handle the sudden illumination changes. Javed et.al. [13] proposed a gradient information of MoG to overcome the slow learning rate problem. Butler et.al. [15] proposed a cluster based background segmentation approach. Algorithm is suffered with the computational cost. Kim et.al. [5] proposed quantize based code book and traditional GMM approach for the moving object detection. The algorithm works well on slow as well as fast moving objects. Patwardhan et.al. [14] used a colour values and arrange them in a layer and proposed novel approach for the motion detection. Kim et.al. [2] Have used RGB background modelling for the real time moving object detection and used morphology for removing noise and blob labelling for real time moving object detection. They predict the velocity of the moving objects and detect it. Wang et.al. [7] have proposed that the suitability to background changes is not as satisfying as them especially to some phenomenon like sudden illumination changes. LI et.al. [9] Proposed GMM as a background modelling and update the learning parameters. They have proposed algorithm which works well on against illumination variations. Due to adaptive GMM and updating the parameters the convergence speed has shown very much improvement. Katharina et.al. [10] Proposed spatio-temporal adaptive GMM by using traditional GMM and spatial and

temporal dependency. Further the quality of the foreground detection can be improved by removing the shadows.

III. PROPOSED METHOD

Smart video surveillance system requires robust detection and leads to an accurate tracking. Our proposed approach estimates background which handles dynamic scenes, moving background, entering and leaving objects in a current frame, clutter background, complex object silhouette etc.,. Our proposed algorithm divides into three steps: Background Analysis, Object detection and a tracking. Our proposed algorithm works on traditional Gaussian mixture model and updates our model in every successive frame. The background model is continuously updated with learning rate. One of the beauty of GMM approach is if the moving object is remain stationary in a current frame it will be accommodated in a background without affecting the existing background model.

Background Model:

In a smart surveillance system background is usually a non static due to leaf's floating and twinkling of water surface. Generally the Gaussian approach and static or dynamic threshold will segment the foreground but this can be applicable to the scene where the background is almost constant. In practice the real time surveillance system must take challenges of the dynamic scenes. To handle such a moving background and other constraints we must have to adopt multivariate Gaussian mixture model. So, background modelling is generally used in certain applications to model the background for the motion detection.

Background Analysis:

The Mixture of Gaussian is required to do the segmentation of the foreground and background rather than the direct foreground object segmentation. Usually the processing is been carried out pixel by pixel or intensity based rather than region or contextual based. If each pixel is characterized from a static background and the illumination can be fixed or changed with respect to time single Gaussian would detect motion from the current frame.

$$F(X/\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

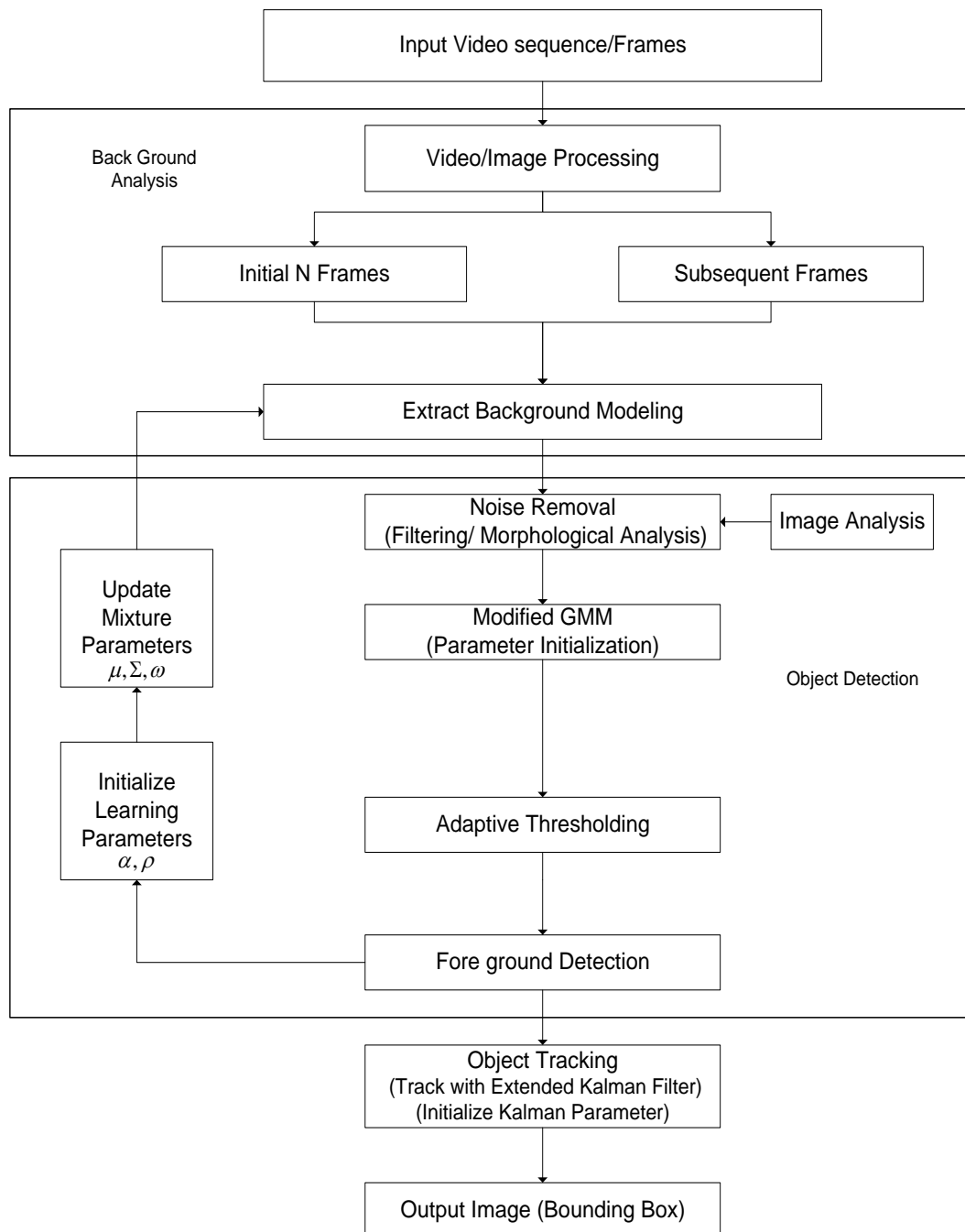


Figure 2 Proposed Method

The single Gaussian handles small and gradual variations in background but fails to handle large and sudden illumination variations. Generally multiple surfaces are available for a particular pixel in real time application. To handle such an situation instead of single Gaussian multiple Gaussian is required to estimate the background model. Every time mixture parameter is updated and it will maintain the model.

In a RGB color space, frame pixel can be characterized by intensities and its probability in the current frame is:

$$P(X_t) = \sum_{i=1}^k \omega_{i,t} \eta(X_t, \mu_{i,t}, \Sigma_{i,t})$$

Where,

$\omega_{i,t}$ = weighted associate to current frame Gaussian

k = no. of distributions.

$\mu_{i,t}$ & $\Sigma_{i,t}$ = mean and covariance matrix of the pixel intensities

η = the Gaussian probability density function,

$$\eta(X/\mu, \Sigma) = \frac{1}{\sqrt{2\pi|\Sigma|}} e^{\{-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)\}}$$

Each pixel is identified as a mixture of Gaussian and initializes the different mixture model parameters. The weight, the mean and the covariance matrix is initializing using an EM algorithm or Maximum A Posteriori (MAP) estimation [8].

$$\bar{\omega}_i = \frac{1}{T} \sum_{t=1}^T p_r(i | X_t, \lambda) \quad \text{---- Mixture weight}$$

$$\bar{\mu}_i = \frac{\sum_{t=1}^T p_r(i | X_t, \lambda) X_t}{\sum_{t=1}^T p_r(i | X_t, \lambda)} \quad \text{----mean}$$

$$\bar{\sigma}_i^2 = \frac{\sum_{t=1}^T p_r(i | X_t, \lambda) x_t^2}{\sum_{t=1}^T p_r(i | X_t, \lambda)} - \bar{\mu}_i^2 \quad \text{---Variance}$$

Identified the foreground and background pixel from the existing frame match has been found against the existing K Gaussian. The match has been found using the mahalanobis distance,

$$\sqrt{((X_{t+1} - \mu_{i,t})^T \Sigma_{i,t}^{-1} (X_{t+1} - \mu_{i,t})) \langle k \sigma_{i,t} \rangle}$$

This match will decide whether the pixel is a foreground or background pixel. Once the matched has been found our model parameters will be updated by means of learning parameters α and ρ .

$$\omega_{i,t+1} = (1 - \alpha)\omega_{i,t} + \alpha \quad \text{Updated weight}$$

$$\mu_{i,t+1} = (1 - \rho)\mu_{i,t} + \rho \cdot X_{t+1} \quad \text{Updated mean}$$

$$\sigma_{i,t+1}^2 = (1 - \rho)\sigma_{i,t}^2 + \rho(X_{t+1} - \mu_{i,t+1})(X_{t+1} - \mu_{i,t+1})^T \quad \text{Updated variance}$$

Object Tracking (Kalman Filtering):

In a video surveillance system generally, tracking can be carried out after the motion segmentation while in a smart video surveillance system tracking can be done along with the detection. Sometimes it is important to develop a robust foreground segmentation approach for getting better tracking accuracy. Image noise, complexity in silhouette and paths, dynamic background, object features etc. are increases the tracking complexities. Different kinds of recursive and non recursive tracking approaches are available. Among all approaches recursive Kalman approach will give somewhat more precise result.

Kalman filtering:

Object correct location can be easily and promptly calculated by means of statistical process. Kalman is the statistical approach which uses mathematical equations and successively inputs. Kalman filter is used to approximate the state of a linear system where the state is implicit to be distributed by a Gaussian [3]. The prediction state of the Kalman filter uses the state model to predict the new state of the variable.

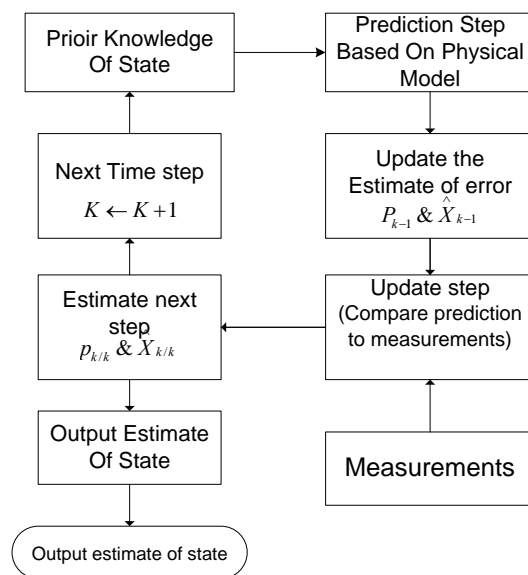


Figure 3 block diagram of Kalman filter

The Kalman filter estimates the entire process by means of feedback. The Kalman estimates the process state and obtain feedback in a form of measurement. Time update and Measurement update are the two states of the Kalman. The time update is liable for the analysing the current state and error covariance estimates the prior for the next time step. The measurement updates are liable for incorporating a new measurement into a prior estimate to get an improved posterior estimate [17].

Time update equation can also be considered as prediction equations.

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1} \quad \text{and} \quad \bar{P}_k = AP_{k-1}A^T + Q$$

The measurement equation is recognized as correction equations.

$$K_k = \bar{P}_k H^T (H\bar{P}_k H^T + R)^{-1}$$

$$\hat{x}_k = \hat{x}_k^- + K_k (z_k - H\hat{x}_k^-) \quad \text{and} \quad P_k = (I - K_k H)\bar{P}_k$$

IV. EXPERIMENT RESULT

The performance evaluation of our proposed algorithm has been experienced in indoor environments against light changes, partial occlusion and clutter background, abruptly entering or leaving moving objects. We have tested our proposed algorithm with the publicly available standard datasets ViSOR [19] CAVIAR [17] and APIDIS [18]. We have evaluated our proposed algorithm using some of the well known evaluation metrics like PR curves and a plot between FPPI and miss rate.

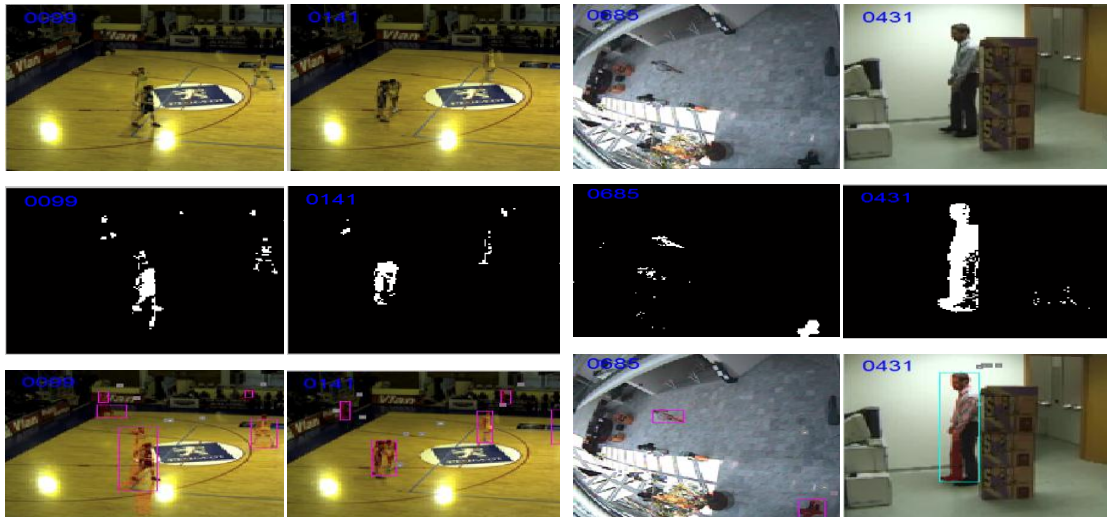


Figure 4 Indoor standard dataset sequences

(Row 1: Original sequence, Row 2: Foreground Mask, Row 3: Tracking Result)

Figure 5 Indoor standard dataset Sequences

(Row 1: Original sequence, Row 2: Foreground Mask, Row 3: Tracking Result)

Figure 4 is an standard indoor APIDIS [18] dataset. It is suffered with the similar appearance and with the reasonable level of occlusions. Our proposed approach can be able to track the moving objects accurately. Somewhere proposed algorithm also detects shadows because of the glossy surface areas Gaussian can predict it as a moving object.

Figure 5 is again from standard ViSOR [19] and CAVIAR [17] dataset for the Indoor surveillance. First sequence is taken it from a reasonable altitude and other is with high level of static occlusions. Our proposed algorithm can be able to detect and track the objects clearly with the high accuracy. Our proposed algorithm can handle static high level occlusions and intensity variations.

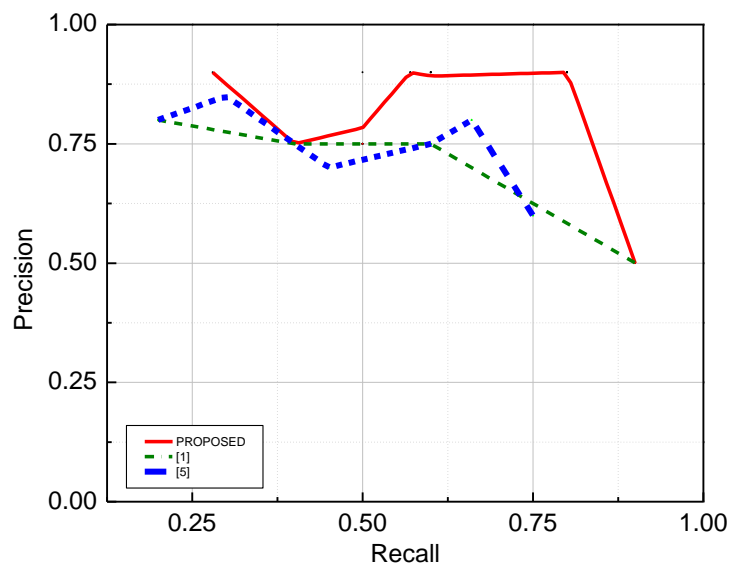


Figure 6 Precision-Recall curve

$$\text{Precision} = \frac{T_p}{T_p + F_p} \quad \text{and} \quad \text{Recall} = \frac{T_p}{T_p + F_n}$$

Figure 6 shows the comparative Precision – Recall analysis for the APIDIS sequences. Our proposed approach shows significant improvement in the Recall and Precision hence the false positives and false negatives are simultaneously reduces due to noise removal and post processing morphological approach. Our performance evaluation is compare with the other similar approaches.

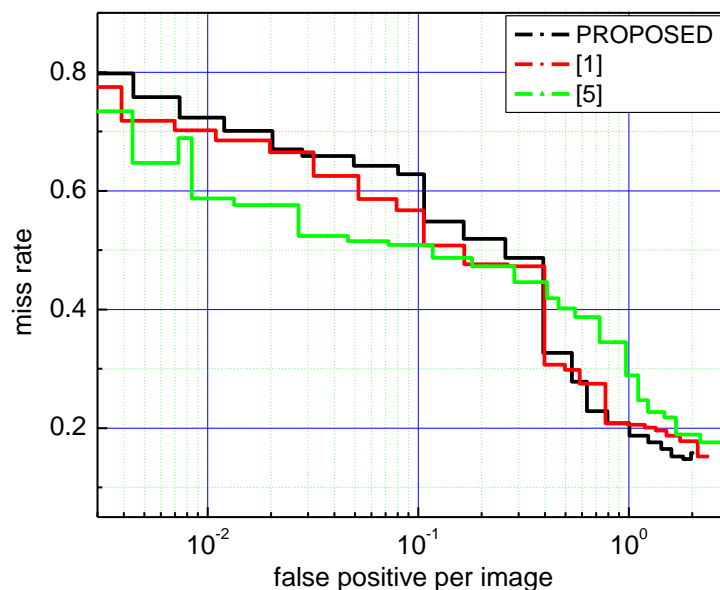


Figure 7 FPPI v/s miss rate

Figure 7 critically compare our proposed approach with GMM [1] and codbook [5] for the false negatives. Generally we compare it with reference to false positives per window or image. Comparative analysis indicates that between 0.1 FPPI to 1 FPPI our proposed algorithm shows significant improvements in miss rate.

V. CONCLUSION

For the indoor surveillance our proposed approach gives better Recall and Precision against the dynamic background, partial or high level of occlusion and sudden light variations. In contrast to some of the existing pixel based parametric approach, our algorithm is not constrained by illumination variation and occlusions. To demonstrate the performance evaluation of our proposed algorithms, we have taken some of the challenging standard indoor video datasets like APIDIS basketball, ViSOR and CAVIAR. In future work will concentrate on other features rather than only color to make it robust and can handle objects that are fully occluded and with similar appearance.

REFERENCES

- [1] Chris Stauffer and W. Eric L. Grimson, "Learning Patterns of Activity Using Real-Time Tracking," IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, pp 747-757 August 2000.
- [2] Jong Sun Kim; Dong Hae Yoem; Young Hoo Joo. "Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems"

- IEEE transaction on Consumer Electronics. Vol . 57, Issue 3, pp. 1165-1170, August 2011.
- [3] A. Yilmaz, O. Javed and M. Shah, "Object Tracking: A Survey", ACM Journal of Computing Surveys, Vol. 38(4), , pp. 1-45,2006.
 - [4] Thierry Bouwmans, Fida El Baf. Bertrand Vachon. "Background Modeling using Mixture of Gaussians for Foreground Detection – A Survey", Recent Patents on Computer Science, Bentham Science Publishers, 1 (3), pp. 219-237, 2008.
 - [5] K. Kim, T.H. Chalidabhongse, D. Harwood, and L. Davis. Real-time foreground-background segmentation using codebook model. Real-Time Imaging, 11:172–185, 2005.
 - [6] R. Bowden and P. Kaewtrakulpong, "An improved adaptive background mixture model for real-time tracking with shadow detection," in AVBS01, 2001.
 - [7] Wang, Yang. "Real-time moving vehicle detection with cast shadow removal in video based on conditional random field." IEEE Transactions on Circuits and Systems for Video Technology, 19. 3 pp. 437-441. (2009)
 - [8] Douglas Reynolds, "Gaussian Mixture Models", MIT Lincoln Laboratory, 244 Wood St., Lexington, MA 02140, USA.
 - [9] LI Xing-liang, WU Yu-bao, "Image Object Detection Algorithm Based on Improved Gaussian Mixture Model", Journal MULTimedia, Vol. 9, NO. 1, pp. 152-158 , January 2014.
 - [10] Katharina Quast, Matthias Obermann ,André Kaup, "Real-Time Moving Object Detection in video sequences using spatio-temporal adaptive Gaussian mixture models", International Conference on Computer Vision Theory and Applications, pp. 413-418,2010.
 - [11] M. Harville, "A framework for high-level feedback to adaptive, per-pixel, mixture-of-gaussian background models," in 7th European on computer vision, Vol. 3, (Copenhagen, Denmark), pp. 37-49,2002.
 - [12] M.Harville, G.G.Gordon and J.Woodfill. "Foreground segmentation using adaptive mixture models in color and depth, " in IEEE workshop on Detection and Recognition of Events in Video, pp. 3-11,2001
 - [13] O. Javed, K. Shafiq and M. Shah, "A hierarchical approach to robust background subtraction using color and gradient information," in IEEE workshop on Motion and Video Computing, pp.22-27, 2002.
 - [14] K.Patwardhan, G. Sapiro, and V. Morellas, "Robust foreground detection in video using pixel layers," IEEE trans. On Pattern Analysis and Machine Intelligence, Vol. 30, no.4, pp. 746-751, 2008.

- [15] D. Butler, S. Sridharan and V.M. Bove Jr, "Real time adaptive background segmentation" in IEEE workshop on Visual Motion, (Irvine, CA), 1989
- [16] Nima Seif Naraghi, "A Comparative Study of Background Estimation Algorithms", Eastern Mediterranean University pp. 16-43, September 2009
- [17] <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>.
- [18] <http://www.apidis.org/Dataset/>
- [19] R.Vezzani,R.Cucchiara, "Video Surveillance Online Repository (ViSOR): an integrated framework" in Multimedia Tools and Applications, vol. 50, n. 2, Kluwer Academic Press, pp. 359-380, 2010
- [20] A Novel Approach to Detect Foreground in Video Sequences based on Mixture of Gaussians Navneet S. Ghedia, Dr. C.H. Vithalani, Dr. Kiran Parmar, IJSRD - International Journal for Scientific Research & Development| Vol. 3, Issue 06, 2015.