International Journal of Computational Intelligence Research
ISSN 0973-1873 Volume 13, Number 5 (2017), pp. 851-864

http://www.ripublication.com

# A Novel Approach for Monocular 3D Object Tracking in Cluttered Environment

**Navneet S. Ghedia**
*Research scholar, Gujarat Technological University, Gujarat, India.*

**Dr. C.H. Vithalani**
Professor and Head of EC Dept., Government Engineering College, Rajkot, India.

**Dr. Ashish Kothari**
*Associate Professor and Head of EC Dept., Atmiya Institute of Technology and Science, Rajkot, Gujarat, India.*

**Abstract**

Machine perception is an essential feature for an autonomous system. For the computer vision researcher perception of scene is an important aspect. Smart surveillance system can be able to sense the environments and understand it in a smartly. Location and behavior of objects in a space is helpful in detection and tracking of it in dynamic scenes. Detection and Tracking of objects is really a difficult task if it is to be estimated in 3D. This paper presents novel and robust approach for 3D object detection and tracking using monocular scene. Our statistical approach takes geometric information of the 3D scene. So our proposed algorithm is capable to track rigid objects in 3D using Monocular camera and it can also handle non static background and partial occlusions. The performance evaluation will shows the significant amount of improvements, robustness and the efficiency of our proposed algorithm.

**Keywords:** Computer vision, Tracking, Monocular scene, Geometric information, MAP-EM.

## I. INTRODUCTION

In the area of computer vision and machine perception the static camera and the gray or color images are the only sources and tracking objects from monocular scene is an critical research due to its numerous application like surveillance, autonomous driving assistance, Sport analysis, Augmented Reality, etc, [11]. Non static or dynamic, clutter background and partial occlusions and the estimation of 3D coordinates in 2D image plane will add the navigational complexity. Our proposed algorithm shows novel probabilistic 3D object detection and tracking approach. Our method is capable to strongly track a different number of targets in 3D coordinate in dynamic and complex scenes and in presence of large amount of variations in visual appearance. In spite of monocular video sequence and other constraints, proposed algorithm gives reasonable dynamics and gives unlikely false positives. Generally we can reduce false positives and on the same way we can decrease false negatives or we can increase the recall by means of adopting the multi view instead of monocular view. The use of multiview is essential for the precise tracking purpose against the multiple and fully occlusions. Multiview tracking is used to reduce the unseen

Areas and give 3D information regarding the objects and the scene.

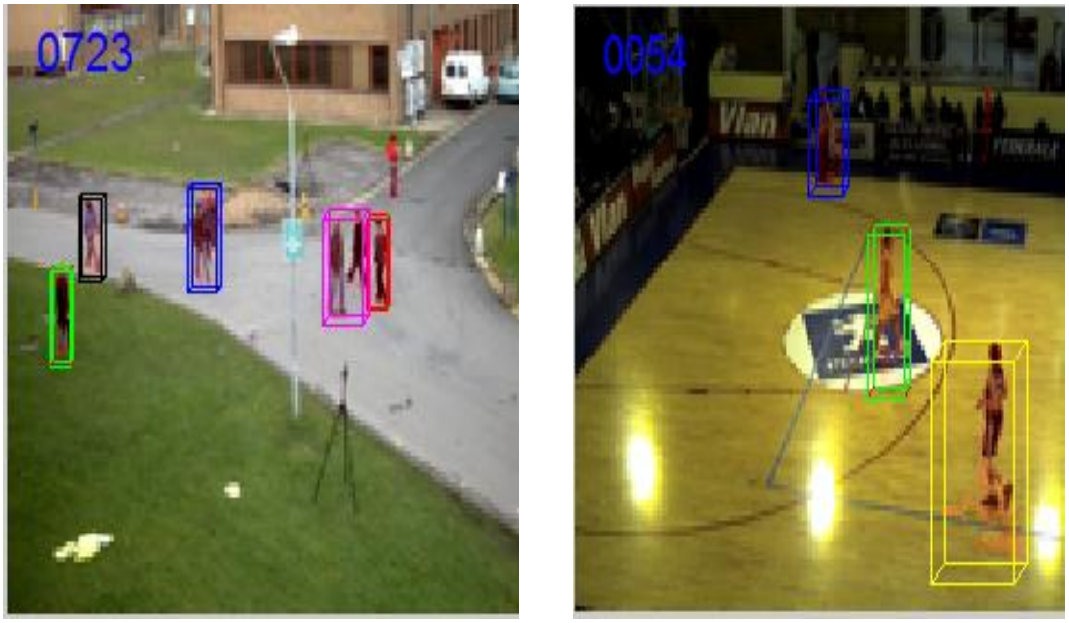


**Figure.1** 3D object detection in monocular scenes

Our Paper presents offline tracking of multiple objects in monocular video scenes using geometric information of 3d coordinates in 2d image plane. We are also using modified Gaussian mixture model as a background subtraction and color model as a feature type. The mixture model gives robustness to handle dynamic scenes.

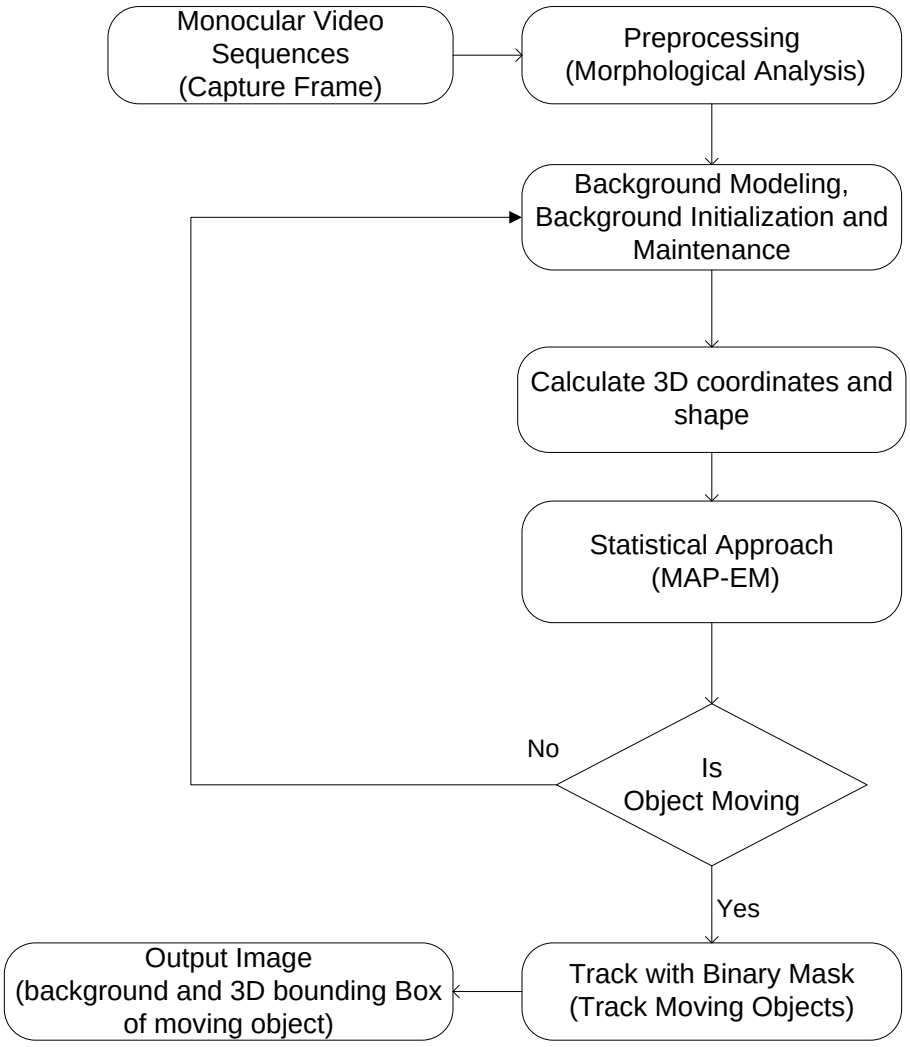


**Figure 2.** Block diagram of 3D object detection and tracking

Figure 2 shows general approach of the 3D moving object detection and tracking. Object position, object shape, and the actions carried out by the object is an essential in scene investigation. The robust 3D tracking depends on the robustness of the foreground voxel detection. 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 [3].

Generally, in a video sequences static background is available in a video. For the typical applications like sports events, autonomous driving assistance having a dynamic background, background frames are not available so for the Initialization subsequent frames are required to generate the background. In the next step preprocessing such as morphology, image resizing and the edge detection is required. Gaussian mixture model is required for the background subtraction and model parameters is initialized by PDF's and maintain the model parameters by MAP-EM approach. For the 3D object detection and tracking calculate 3D shape and

coordinates in 2D image plane. Finally the segmented voxel is required to track using the recursive approach. Among all the tracking approaches Kalman will give better tracking accuracy.

## II. RELATED WORK

We present background of the proposed algorithm and recent advances of 3D object tracking using monocular video sequence. There is plenty of related work on single static camera moving object detection and tracking algorithms are available, roughly most of are suffered from the tracking difficulties in partial and fully occlusions. Zhao et.al, [4] proposed a monocular 3D tracking method. They have used 3D shape model of object in 2D image plane to help in segmentation and handling the fully occlusions. Complex object silhouettes can be tracked in 3D using kalman filtering. Appearance and geometric information are the two wide approaches for the 3D foreground detection information. In appearance approach, visual hulls are created by a silhouette method and every model are trained on their shapes of every moving object voxel categorization in to objects [5]. The categorized objects can be tracked using kalman filtering [6] or mean shift [7]. For non similar or discriminative appearance such algorithms provides excellent tracking accuracy and recall while it cannot performed well against the similar object appearance. the level set method [9], human body model fitting [8], are the different approaches available to handle similar object appearance and these approaches can able to handle dynamic scene as well as occlusions. Hoiem et al. [10] and Ess et al. [11] presented a unifying approach to locate object in 3D space using segmentation and object detections. Kelly et al. [12] proposed a 3D situation model using the voxel characteristic. Humans were modeled as a set of these voxels to decide the camera-handoff difficulty. it is simple to design a trajectory with then excellent quality of the filtering by dynamic programming which trajectories are expected one after another [13]. For the detection and tracking in 3D space using monocular video sequence, extract foreground voxels and its location is evaluated by analyzing or considering detection responses from more than one – multiview in a common 3D space [3]. Peter Carret.at.[14] proposed a novel approach for the 3D object detection using geometric primitives because in occupancy map it produced fixed location specific patterns. Mean shift is responsible for determining objects and its positions.In comparison to 3D, 2D algorithm gives low robustness of the foreground detection information in crowded scenes due to the absence of altitude. Our objective is to develop a statistical probabilistic scene model that can deduce the location of moving objects in 3D coordinates in monocular video sequences.

## III PROPOSED METHOD

Multiple object detection in 3d space under the dynamic background is required vigorous motion segmentation and precise tracking. To track 3D objects in monocular sequences our proposed algorithm works on geometry based method. In our proposed approach object appearances are discriminative. The object silhouette can be detected using the traditional modified GMM approach. To represent 3D object shape, semi parametric PDFs and for the object parameter MAP-EM are used.

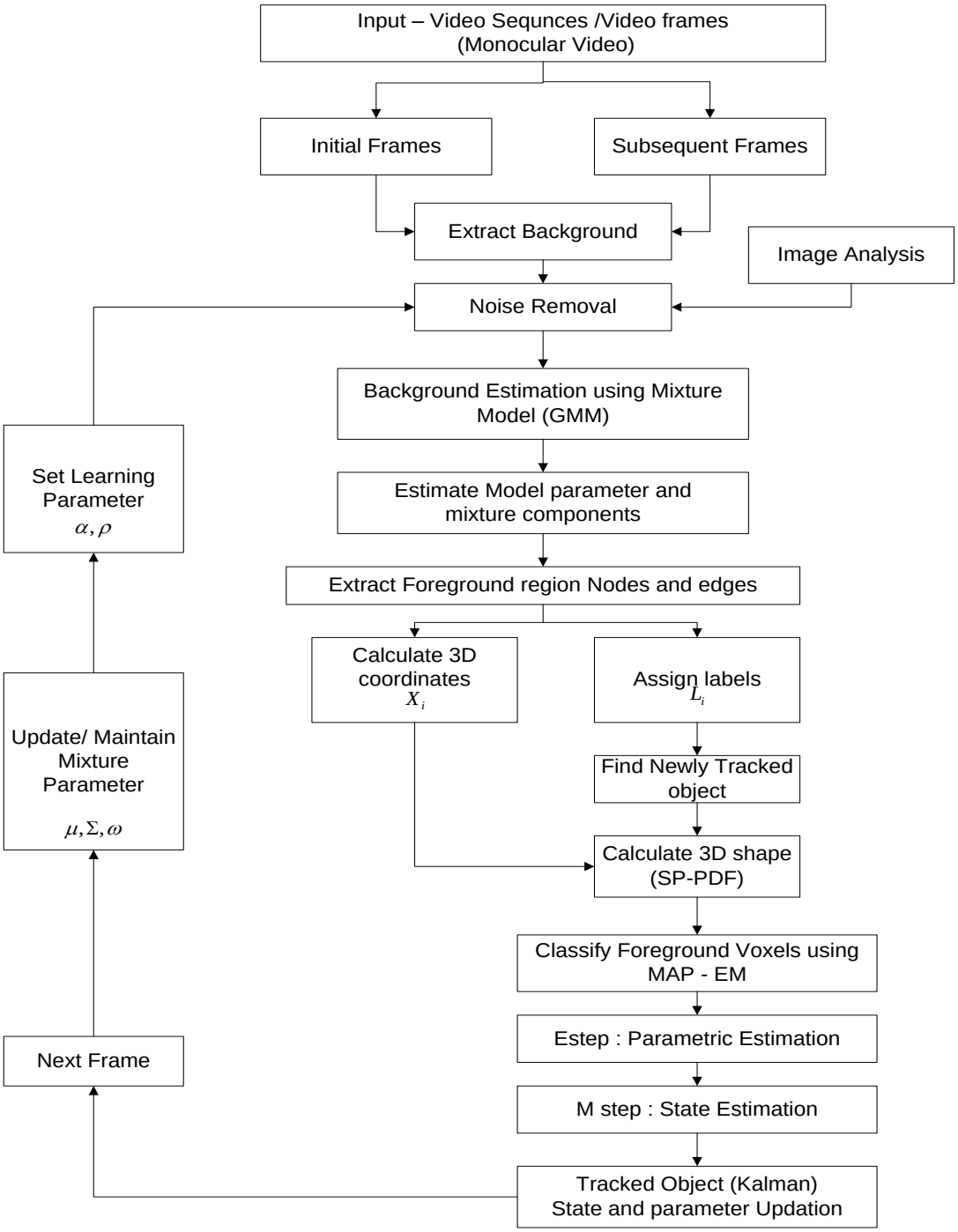


**Figure 3.** Proposed Algorithm

Our proposed algorithm based on modified GMM approach. The background model can be estimated and initialize the parameters using MLE estimation and maintain the background model parameters using EM algorithm. Once foreground voxel are estimated calculate 3D coordinates and 3D shape using semi parametric PDFs.. The

foreground voxels can be classified using MAP-EM algorithm. Finally the segmented voxels can be tracked using the 3D kalman filter.

**Foreground Voxel Classification:**

We classify foreground voxels of objects as the assumption of the mutual distribution $p(x,l)$.

Foreground voxels of objects at time t, $A=\{1,2,...,n\}$.

Where, $x$ =3D coordinates of moving object voxel, $l \in A$). n= number of objects.

The joint and conditional distributions can be shown as, $p(x,l)=p(l)p(x|l)$

Where,

$p(l)$ =prior, is used as a mixing coefficient in EM frame work.

$p(x|l)$ =likelihood function.

Prior for the EM algorithm can be expressed as, $p(l)=prior=\pi l$ $(\sum_{l\in A}\pi l=1 \wedge 0\le \pi l\le 1, \forall l\in A)$

Let, $\psi_l^t$ = moving object 3D position $l$ at $t$.

$p(x|l)$ is the probability of $x$ given $\psi_l^t$ is gives, $p(x|l):p(x|\psi_l^t)$

Where, $p(x|\psi_l^t)$ = semi parametric PDF of 3D shape of a object.

For the MAP estimation of the $\psi_l^t$ in MAP-EM is assume the positions of object at $t$ are close to $(t-1)$, the prior of $\psi_l^t$ can be defined as a multivariate Gaussian distributions, $p(\psi_l^t):\eta(\psi_l^t|\psi_l^{t-1},\Sigma)$

where $\Sigma$ =covariance

**MAP-EM:**

Let $\chi$ represents the 3D coordinates for the moving object voxels. For the said dataset $\chi$, calculate the Maximum a posteriori for mixture model.

Model parameters can be initialized as, $\Theta=\{\pi l,\psi_l^t\}$

Calculate $\Theta^{t+1}$ from the current approximation $\Theta^t$ using the repetitive computation and E M steps.

**E step:**

We use $\Theta^t$ to find the posterior distribution.

$p(l|x,\Theta^t)$ $(\forall x\in\chi)$ Of the latent variables

**M step:**

We estimate $\Theta^{t+1}$ by maximizing the sum of the expectation of the log likelihood $Q(\Theta|\Theta^t)$ and the logarithm of the prior $p(\Theta)$ with $\Theta$ as follows,

$$\Theta^{t+1}=\arg\max_{\Theta} R(\Theta|\Theta^t)$$

Where, $R(\Theta|\Theta^t)=Q(\Theta|\Theta^t)+\ln p(\Theta)$

$Q(\Theta|\Theta^t)$ is known as Q-function.

Where, $\ln p(\Theta) \sum_{l \in A} \ln p(\psi_l^t)$

We can maximize it by,

$$\Theta^{t+1} = \arg\max_{\Theta} Q(\Theta|\Theta^t) + \ln p(\Theta)$$

Where, $Q(\Theta|\Theta^t) = \int_{l(x)} \log p(l|\Theta) p(l|x, \Theta^t) dx$

According to Bayes' theorem,

$$p(l|x,\Theta) = \frac{p(l)\,p(x|l,\Theta)}{p(x)} \qquad \text{and} \quad \ln p(x|l,\Theta) = \sum_{x \in X} \{\ln p(l) + \ln p(x|l)\}$$

Model parameters are updated as,

Prior can be updated by.

$$\pi l^t \rightarrow \pi l_u^t$$

The 3D position of object can be updated by,

$$\psi_l^t \rightarrow \psi_{lu}^t$$

Object Tracking:

Multiple object tracking requires high precision. Proposed algorithm can handle dynamic environment and occlusions. For the robust tracking approach object motion information and object feature examination are required. Generally for the predictive and static environment tracking can be done along with the detection. While in presence of noise or dynamic background tracking can be done after detection or motion segmentation. Fast moving and non-rigid object may increase the tracking complexity. We can just track moving objects by applying certain constraints on object motion, appearance and silhouette. Prior knowledge and number of objects information make it easy to track moving objects. Various object tracking approaches are available to meet the different challenges like object representation, feature, motion, appearance, silhouette and environment [3].

Kalman filtering:

Tracking and data prediction can be achieved using the kalman filtering. It can gives an analytical approach for the visual motion estimation and tracking. It is the numerical approach which uses successive inputs and statistical equations to minimize the measurement and process noise.

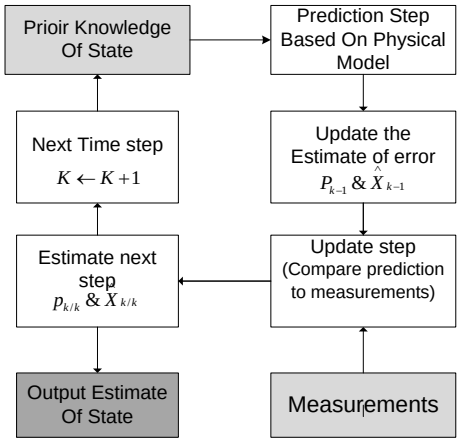


**Figure 4.** Block diagram of Kalman filter

The Kalman filter is used to estimate past, present and future state under unpredicted state. The kalman filtering recursively estimates the process state. It attains feedback as filter measurements. All the filter equations are generally categorized into time and measurement updates. By using time equations, kalman filter estimates next time steps with the help of current state and the error state. The complete recursive approach is beautifully managed by measurement equation. The rime and measurement equations can also be considered as prediction and correction steps for the kalman filtering [19].

| | |
|---|---|
| Time Update Equations | $\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1}$<br>$\bar{P}_k = AP_{k-1}A^T + Q$ |
| Measurement Update Equations | $K_k = \bar{P}_k H^T (H\bar{P}_k H^T + R)^{-1}$<br>$\hat{x}_k = \hat{x}_k^- + K_k\left(z_k - H\hat{x}_k^-\right)$<br>$P_k = (I - K_k H)\bar{P}_k$ |

## IV. EXPERIMENT RESULT

To evaluate our proposed approach, we rely the standard CLEAR multiple-object tracking (MOT) performance metric. The multi object tracking accuracy is calculated with the help of different frame errors like error ratios, false negatives (FNs), false positives (FPs), and identity switches (IDSs). The robustness of the system can be verified from the MOTA values. The MODA score utilize missed counts and false alarms counts [15]. We are evaluating our algorithm on couple of standard challenging video datasets. (APIDIS [16] , ICG- Lab-6 [17] and PETS 2009 [18 ]). Our proposed algorithm is considered for 3D object detection and tracking for monocular sequences. In monocular sequences the ground truth is available in 2D and we are going to track object in 3D without the object altitude.

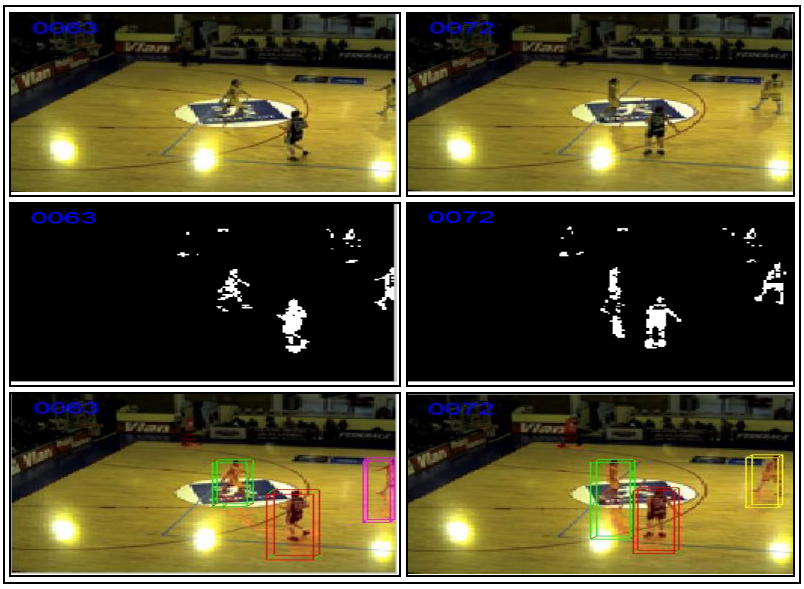


**Figure 5:** 3D tracked model APIDIS

Figure 5 shows the typical tracking result. The sequence is suffered with the clutter background, occlusions and similar appearance. Our proposed approached will attain precise and exact tracking results, even with the constraints like shadowing effects, heavy reflections and the random movement of the moving objects. Our proposed algorithm fails to detect and track objects which are fully occluded.

In addition to this, fig. 6 shows an additional tracking result of PETS data set validate our proposed approach. In both the sequences we have generated 3D detection maps using one of the multiviews available from the datasets. Under the various constraints and challenges our proposed approach shows significant improvements in tracking and detection result.

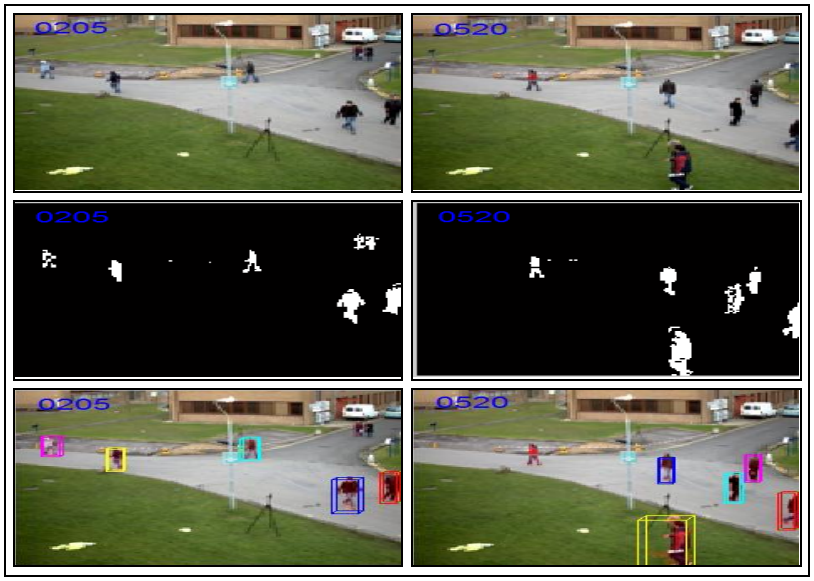


Figure 6 3D tracked Model PETS 2009

**Table 1:** Multiple Object detection accuracy at a tolerance or distance threshold of 0.5m

| Sequence | [3 ] | [ 2] | [14] | Proposed |
|---|---|---|---|---|
| APIDIS | 0.685 | 0.5453 | - | 0.8697 |
| PETS 2009 | 0.719 | 0.787 | 0.6789 | 0.738 |
| LEAF 2 | 0.7630 | 0.8437 | - | 0.8329 |
| MUCH | 0.7503 | 0.7867 | - | 0.773 |
| PETS 2006 | 0.618 | 0.446 | 0.472 | 0.639 |

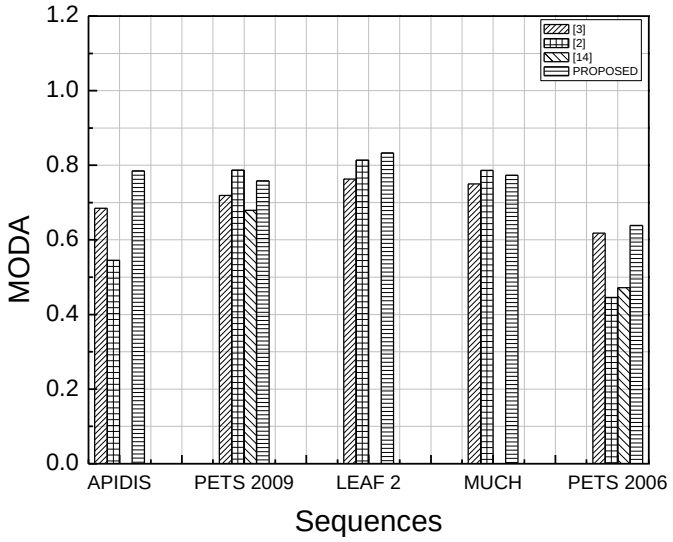


**Figure 7.** Multiple Object Detection Accuracy for the different Datasets.

Table 1 exhibits comparative analysis of the standard datasets. For the monocular 3D object detection proposed algorithm shows significant improvement in terms of the recall for the APIDIS sequences. While for the other sequences it sill exhibits comparative improvements for atypical tolerance of 0.5m. Our proposed algorithm reduces the amount of false positives and ultimately it leads to better detection results. Fig. 7 indicates the comparative analysis of the Multiple Object detection Accuracy.

| MODA | MOTA |
|---|---|
| $MODA = 1 - \frac{F_n + F_p}{T_p + F_n}$ | $MOTA = 1 - \frac{\sum_t m_t + fp_t + mme_t}{\sum_t g_t}$ |

**Table 2:** Comparative Evaluation.

| Sequence | Method | MOTA | TP | FP | FN | IDS |
|---|---|---|---|---|---|---|
| APIDIS | Vol [3] | 0.675 | 656 | 88 | 172 | 9 |
| | POM [2] | 0.49 | 607 | 156 | 220 | 46 |
| | Proposed | 0.799 | 719 | 47 | 108 | 11 |
| LEAF 2 | Vol [3] | 0.727 | 856 | 115 | 117 | 34 |
| | POM [2] | 0.819 | 913 | 87 | 66 | 24 |
| | Proposed | 0.818 | 837 | 71 | 92 | 15 |
| MUCH | Vol [3] | 0.736 | 694 | 99 | 99 | 11 |
| | POM [2] | 0.754 | 770 | 139 | 32 | 26 |
| | Proposed | 0.752 | 719 | 102 | 79 | 17 |

APIDIS [16], ICG- Lab-6 [17] Multiple Object Tracking Accuracy MOTA, the total number of true positives (TP), false positives (FP), false negatives (misses, FN), and identity switches (IDS).

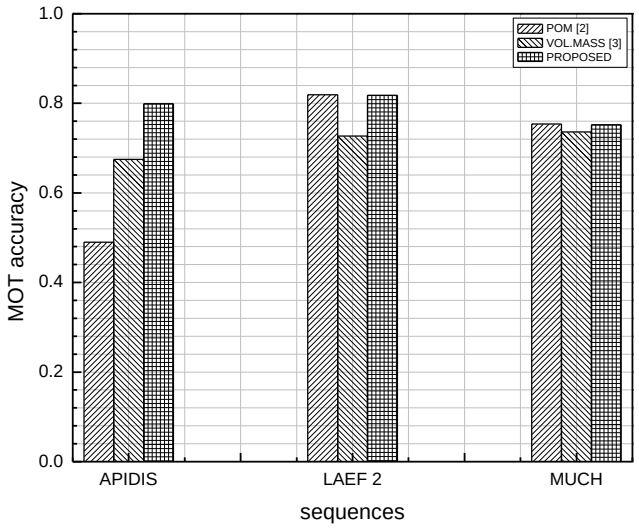


**Figure 8.** Multiple Object Tracking Accuracy for the standard datasets.

Table 2 shows the comparative analysis of the MOTA and errors like False Positives, Negatives and mismatch error and true positives. Higher Multiple Object Tracking Accuracy exhibits robustness of the algorithm, generally it is calculated from the ratio of false positives, negatives and mismatch or identity switches. Our proposed method shows significant improvement in APIDIS sequences. While in case of the other sequences it shows comparative improvement. Our proposed algorithm gives better

precision at the cost of lower recall as false negatives increases but overall it will increases the tracking accuracy compares to other approaches.

Figure 9 shows the comparative evaluation for the standard video sequence PETS 2009. Analyze the evaluation at a constant FPPI rate of 0.1, detctor achieves missrate of 0.48. Dynamic and clutter backgorund introduces false detection. We can get lower missrate or improvement in issrate at the cost of false positives. Our proposed algorithm shows cosiderable improvement in missrate and to the detction and tracking accuracy.

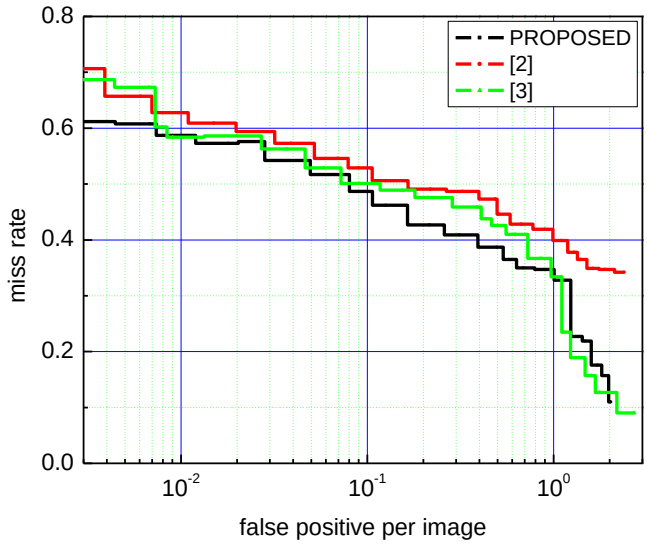


**Figure 9.** Comparative anaysis for PETS 2009

## V. CONCLUSION

We have presented an algorithm for tracking multiple moving objects in monocular 3D scene geometry. There are numerous possible approaches are available, one is to use object appearance to differentiate the objects. Such approach cannot distinguish objects with similar appearance. Our proposed method used geometric information regarding 3D scene structure. We have proposed a probabilistically MAP-EM model to classify foreground voxels. In performance evaluation metrics, our proposed method shows significant improvements in detection and tracking accuracies. Our approach also shows the comparative improvements in false detection. For the future aspects we can plan to expand our proposed model that can handle extensive occlusions, shadowing effects and heavy reflections.

**REFERENCES**

[1] Taiki Sekii," Robust, Real-Time 3D Tracking of Multiple Objects with Similar Appearances", In CVPR 2016.

[2] J. Berclaz, F. Fleuret, E. T¨uretken, and P. Fua. Multiple object tracking using k-shortest paths optimization. PAMI,33(9):1806–1819, 2011.

[3] H. Possegger, S. Sternig, T. Mauthner, P. M. Roth, and

H. Bischof. Robust real-time tracking of multiple objects by volumetric mass densities. In CVPR, 2013.

[4] Zhao, T. and Nevatia, T. 2004. Tracking Multiple Humans in Complex Situations, IEEE PAMI, 2004.

[5] L. Guan, J. S. Franco, and M. Pollefeys. Probabilistic multiview dynamic scene reconstruction and occlusion reasoning from silhouette cues. IJCV, 90(3):283–303, 2010.

[6] M. C. Liem and D. M. Gavrila. Joint multi-person detection and tracking from overlapping cameras. CVIU, 128:36–50,2014..

[7] A. Tyagi, M. Keck, J. Davis, and G. Potamianos. Kernel based 3D tracking. In IEEE International Workshop on Visual Surveillance, 2007..

[8] X. Luo, B. Berendsen, R. T. Tan, and R. C. Veltkamp. Human pose estimation for multiple persons based on volume reconstruction. In ICPR, 2010

[9] Y. Iwashita, R. Kurazume, K. Hara, and T. Hasegawa. Robust motion capture system against target occlusion using fast level set method. In ICRA, 2006.

[10] Hoiem, D., Efros, A.A., Hebert, M.: Putting objects in perspective. IJCV 80 (2008)

[11] Ess, A., Leibe, B., Schindler, K., Van Gool, L.: Robust multi-person tracking from a mobile platform. PAMI 31 (2009)

[12] P. Kelly, A. Katkere, D. Kuramura, S. Moezzi, S. Chatterjee, and R. Jain, "An Architecture for Multiple Perspective Interactive Video," Proc. Third ACM Int'l Conf. Multimedia, 1995.

[13] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multi-Camera People Tracking With a Probabilistic Occupancy Map," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 2, pp. 267–282, February 2008.

[14] P. Carr, Y. Sheikh and I. Matthews, "Monocular Object Detection Using 3D Geometric Primitives", ECCV - 2012

[15] K. Bernardin and R. Stiefelhagen. Evaluating multiple object tracking performance: The CLEAR MOT metrics. In EURASIP JIVP, 2008.

[16] 1http://www.apidis.org/Dataset/

[17] http://lrs.icg.tugraz.at/download#lab6

[18] PETS 2009 Dataset

[19] Greg Welch and Gary Bishop. An introduction to the kalman filter, 1995 & 2006.

[20] PETS 2006 Dataset