Comparative analysis of multiple target tracking methods

Michael Kamaraj Devadoss [a,1,*], Balakrishnan Ganesan [b,2]

[a] Research & Development Centre, Bharathiar University, Coimbatore, India
[b] IndraGanesan College of Engineering, Tiruchirappalli, India
[1] dmkraaj@gmail.com*; [2] balakrishnan.g@gmail.com
* corresponding author

ARTICLE INFO

ABSTRACT

Many applications such as intelligent transportation, video surveillance, robotics of computer vision mainly depend on task of multiple object tracking. It includes the process of detection, classifications and tracking. The main focus of the study is to develop an efficient and effective multiple target tracking methods to solve the issues of illumination changes, occlusions and affinity matching. Accordingly, the various multiple target tracking methods are tested and evaluated using the metrics on publicly available datasets from which it is obvious that the outcome of the global energy minimization and optimization techniques is comparatively better than any other existing techniques in all aspects. This comparative study work will also help in better understanding of the problem, knowledge of the methods and experimental evaluation skill for further research works.

I. Introduction

In the field of computer vision, the multi-target tracking plays a vital task in detecting and tracking of targets and at the same time their identities are preserved. Accurate tracking of the targets is the key for several applications such as video surveillance, motion and pattern analysis, pedestrian tracking, etc. The main focus of this work has concentrated on the tracking of the human in the video sequences. In the recent years, there are many developments in the tracking of multi-target, however, in real-time situations the accuracy and precision are still a challenging task for the state-of-art algorithms. In MTT (Multiple Target Tracking), the first step is the target detection process which comprises of the segmentation, foreground and background extraction. The estimation of the trajectories can be performed on the later stage. In the single target tracking, the target is tracked within the specified area and possible trajectory can be obtained by joining the locations in which target has been moved from time to time. Similarly, in the multiple target tracking, more number of targets is observed simultaneously. The difficult process of correctly matching the identity of the target for corresponding detection is known as data association. The multi-target tracking also faces many challenges such as change in illumination, scale variations, out-of-plane rotation, severe occlusions, similar appearance of targets and multiple target interaction. In this work, various methods have been proposed to solve these issues. The solutions derived from the system are discussed on different aspect of the MTT system. The comparative studies of the various datasets are applied on the different methods are studied. The performances of the methods are evaluated quantitatively using the MTT metrics, and experimental comparisons among the state-of-the-art methods are proposed.

II. Related Work

In the literature survey, it is understood that for more than a decade the multiple target tracking is the active area of research in computer vision. Now it is time to elaborately discuss the existing works based on multiple target tracking. Initially, several algorithms based on recursive methods [1] [2] were using recursive approach for tracking of the targets. In this method, Kalman filter approach

has been used, in which the present state is updated based on the previous frame information. In sequential Monte Carlo sampling method, the distribution consists of weighted particles which are used to specify the current and hidden state information [3-4]. This can handle the non linear and different mode of occurrence. This process will work well for less number of targets with small sample size. Practically, when the number of target increases, the reliable representation of target is difficult because it requires a large number of samples, to handle the data association. However, this can be done with the help of Markov chain Monte Carlo method [5] or probabilistic filtering. During the last few years, non-recursive methods are used progressively due to its popularity. These approaches are used to estimate the trajectory within a time window. The computation of the trajectory can be controlled by allowing the steps to pass through any one of the following, one is the locations on a regular discrete grid and the other is non-maxima suppression for target detection. Thereby the solution space is limited to a finite state.

Michael et al. [6] has designed an algorithm to detect the number of active blob in the video sequence, and also it discusses the speed efficiency of the application. However, it does not insist on the tracking accuracy and precisions. Leibe et al. [7] introduces a heuristic approach to solve the local optimality of the target, the prior task of target detection and trajectory estimation are carried out by quadratic binary technique. Michael et al. [8, 9] introduces a Gaussian mixture model which is based on likelihood matching methods in order to track the multiple targets in the video sequences. This includes the active background extraction, segmentation process and likelihood matching to distinguish the affinity between the targets. Furthermore, the cost optimization is performed by the assignment problem. In addition, the Kalman filter is applied to precisely track the target. Jiang et al. [10] designed a framework which integrates the process of tracking multi-targets as an integer linear program with few constraints imposed on layouts. Furthermore, the LP relaxation is applied to attain the global optimal solutions, but it is unsuccessful many times. Whenever a target is passed through the occlusion, a special node is created to handle the problem of occlusions. There are some serious limitations in this approach, such as the number of target has to be specified in advance and the number of occluded targets has to be defined to avoid collision. The exact localization of occluded target is important to achieve the high accuracy of target tracking. Michael et al. [11, 12] has proposed an optimization of the multi-target tracking and occlusion handling technique using mean shift method. This focuses on the cost minimization of the target localization, the similarity matching between the candidate target and the actual targets by the colour feature. Furthermore, it discusses the optimization strategy and the energy minimization procedures.

According to Berclaz et al. [13] the tracking region is part into fragments of disjoint cells and created a concept of virtual location, which can generate fresh trajectories and take up the existing trajectory at some point of locations. The resulting solution of integer linear program is again being fed into the K-shortest path or LP-relaxation algorithm to speed the computation. The framework also includes the concept of similarity matching, thereby lowering the count of identity switches between the targets. Rodriguez et al. [14] employed a head-based tracking in a highly crowded region. The binary energy minimization function is used to point out the exact count of the targets with a certain constraint terms. Here the camera is set up at a high viewpoint which can be used in surveillance however this is not viable for other applications such as intelligent transportations, entertainment, etc. Xing et al. [15] designed a framework in which the gap occurs in the trajectory due to occlusion. The long trajectory can be built by connecting the tracklets along with the short tracklet generated without occlusions. In the crowded environment, many serious problems may occur as there will be a large variation of dynamic targets and repeated occlusions. This type of target tracking is rarely processed due to its implicit difficulties. However, Kratz and Nishino [16] worked spatio-temporal method to study the motion patterns in the crowded environment. The target likelihood is calculated by converting the distance of colour histogram into probability using the exponential function. Choi and Savarese [17] employed a mean shift tracker which utilizes colour histogram to detect the target sequentially.

According to Qin and Shelton [18] the appearance model is initialized as colour histograms, then the mean weight of all the detection responses are developed into trajectories. Bhattacharyya coefficient is used to calculate the likelihood of the two targets as colour histograms. The similarity of appearance is measured as Bhattacharyya distance between the tracklets and the mean HSV colour histogram. Zhang et al. [19] studied the appearance model uses RGB colour histogram to calculate the link affinity between detection responses. The similarity between the same target and

different targets are obtained by the Gaussian distributions. However, the colour histogram cannot represent spatial information. Henriques et al. [20] employed a covariance matrix descriptor in order to represent target appearance model. The likelihood matching is performed by linking Gaussian distributions of detection response. Breitenstein et al. [21] represented the 2D image speed and the positions using motion model with constant velocity. Here the previous states of object are considered to reduce the noise obtained as the mean and variance from the Gaussian. Andriyenko and Schindler [22] and Milan et al. [23] also utilized a constant velocity model in which the energy terms such as trajectory persistence, mutual exclusion, fragmentation, and handling of occlusion are sum to obtain a cost minimization of multiple object tracking. According to Bae et al.[24] the online multi-object tracking problems are solved by associating the tracklet confidence and the fragmentation of tracklets are linked up without iterations. However, the issue of the association of tracklet in complex scenario remains unsolved. Wu et al. [25] employed a four body part detector for tracking human in the inter-object and scene occlusions. de Villiers et al. [27] designed a mean shift tracker for handling of objects in occlusions, but still there exist a few issues which remains unsolved.

III. Multiple Tracking Methods

MTT play an important role in computing the similarity between the appearances of the target. It is important to note that single object tracking is mainly focus on the discrimination of the target from the active backgrounds. In the real time, it is not easy to discriminate the multiple targets, hence the MTT need some additional information of appearance to the multiple targets from the background.

A. Gaussian Mixture Model Based Beta-likelihood Matching and Kalman Filter (GMM- BLM-KF)

The GMM-BLM-KF [8] is a parametric probability density model consists of Gaussian components is developed with the basic concept of the adaptive background model [26]. The multimodal density of targets is obtained by combining these component functions. The colour feature is very commonly used features in tracking. These colour components can be utilized in the real time applications in tracking the colour based MTT and segmentation. In order to develop a sturdy model, a component mixture model related to the colour of background with respect to the foreground model is generated. The pixel classification is accomplished by Bayes theorem. Gaussian mixture model is the adaptable and effective for the online mode of target tracking and also it is suitable for slow illumination changing conditions.

A Gaussian mixture model is defined as the weighted sum of M Gaussian components which is given as in (1).

$$\rho\left(x|\lambda\right)=\sum_{i=1}^{M}\omega_{i}g(x\,|\,\mu_{i},\sum\nolimits_{i}) \tag{1}$$

Here x represents D-dimensional data, ω_i is the mixture weights varies from 1 to M. And $g(x\,|\,\mu_{i},\sum\nolimits_{i})$ i= 1,......, M ,represents the Gaussian component densities. Each Gaussian component function is derived as

$$g(x\,|\,\mu_{i},\sum\nolimits_{i})=\frac{1}{2\pi^{D/2}\left|\sum\nolimits_{i}\right|^{\frac{1}{2}}}exp\left\{\frac{-1}{2}\left(x-\mu_{i}\right)^{'}\sum\nolimits_{i}^{-1}(x-\mu_{i})\right\} \tag{2}$$

μ_{i} represents the mean vector and $\sum\nolimits_{i}$ represents the covariance matrix. $\sum\nolimits_{i}^{-1}\omega_{i}=1$ is the constraint of the Gaussian mixture components.

The parameter of the Gaussian mixture components are fulfilled by mixture weight, mean vectors and covariance matrix of the component densities. This is represented as

$$\lambda = \left\{ \omega_i, \mu_i, \sum_i \right\} \qquad i=1,...,\ M \tag{3}$$

The covariance matrix $\sum_i$ can be represented as diagonal. The covariance of matrix, number of components, and the type of parameter are decided by the quantity of the data available for calculating the GMMs parameter.

On the whole, the feature density is formed by the Gaussian components action, and the correlation between feature elements of vector can be obtained by combining the diagonal covariance matrices.

1) Background Extraction

The Gaussians of the mixture model of background and foreground extraction is determined based on the changes in the pixel distributions. The Gaussian distribution includes the least variance and most supporting evidence. This can be dealt as follows, firstly, when the target is visible and persist till the end, the accumulation of supportive evidence and the variance are low for the background distributions. Secondly, when a new target occludes the background target, a new distribution is created or the variance of the existing distribution will increase. Thirdly, a moving target will have higher variance in distribution. The fig. 1 shows the tracking results of the method.

Here, a traditional technique of Gaussian is required to model the best background. Firstly, the Gaussians are ordered according to the values of ω/σ . Secondly, re-estimate the parameters and sort the distribution, so that most possible distributions remain on the top. Next, the B distribution is chosen accordingly,

$$B = argmin_b (\sum_{k=1}^{b} \omega_k > T) \tag{4}$$

T represents the measure of the minimum portion of the background data. For single modal distribution, the T is assumed to be small. The T value will be higher for multi-modal distribution due to the frequent changes in the background.
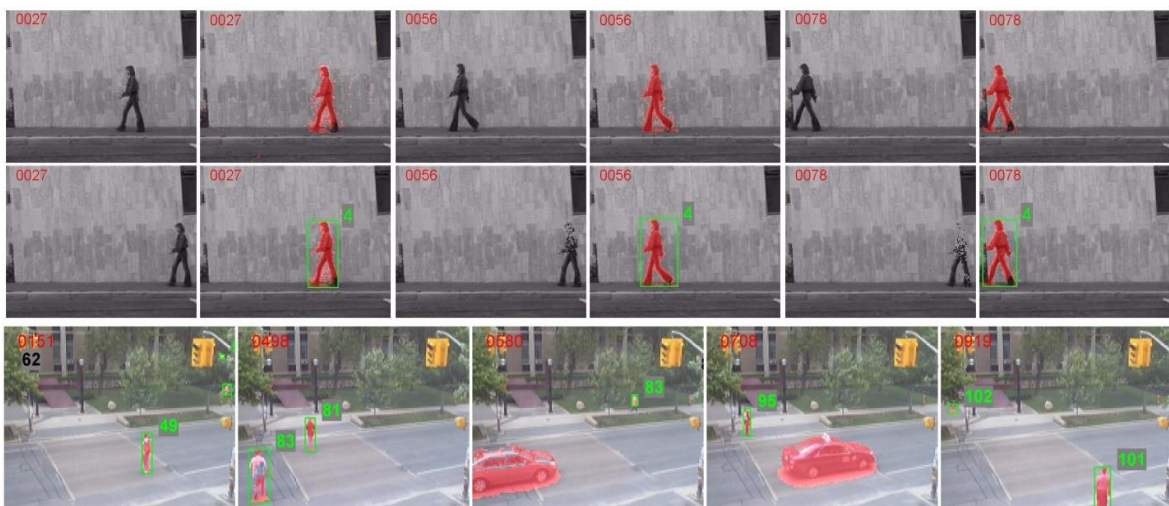

Fig. 1. The tracking results of GMM-BLM-KF method: Row 1 & 2 represents the tracking video sequences of daria walks and Row 3 represents the street crossing sequences.

2) Data Association and Likelihood Matching

In multi-target tracking, data association is a very important and the fundamental task. It is the process of associating uncertain measurements to known tracks. The probability distribution or the multiple targets of the state vector are generated. A weight is associated with each state. The target

area of interest can be obtained as the weighted sum of distributions. Furthermore, likelihood of a target is measured using a Gaussian density function which is given as

$$L_{x_1,\dots,x_n}(\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \times \dots \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}} \tag{5}$$

3) Kalman Filtering

Kalman Filtering is applied to the Gaussian distributed output to estimate the state of linear system. The efficiency of the computation is improved, and also the optimal solution is obtained. The target tracked from one frame to another frame in the video sequence, allows predicting the next instant of the target based upon the previous trajectories. Kalman filter for a Gaussian system are modelled to handle the target changes in the consecutive frames as

$$X_{t+1} = \Phi_t X_t + S_t \tag{6}$$

X represents the state vector $[x,y,u,v,\omega,\Delta t]^T$, the state transition $\Phi = \begin{vmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{vmatrix}$

S_t denotes random vector modelling the uncertainty of the model.

B. Global Energy Minimization and Optimization Technique(GEM-OT)

This recent work [11] discusses the global energy minimization and optimization techniques exclusively for efficient tracking where the tracking of multiple targets still remains a challenging task in the field of computer vision. This task cannot be fulfilled unless the target is tracked accurately in many applications such as video surveillance, pattern matching, intelligent system, and robots etc. The energy terms is determined by the motion of targets in each frame, locations, missing evidence of image, and limitations such as target motion smoothness and mutual target exclusions. The energy minimisation terms are formulated to develop an efficient and global optimal solution of multi-target tracking system.

1) Energy Function

The energy minimization technique is one of the most important tasks for tracking of multiple targets. The general objective of the methods is to provide a possible solution with a low cost. In order to accurately express the multi-target tracking, the energy terms are developed in a closed form to obtain an efficient gradient optimization solution. Each energy term linearly combined to form an energy function

$$W = w_{tt} + k_1 w_{sm} + k_2 w_{or} + k_3 w_{ch} + k_4 w_{tp} + k_5 w_{sd} \tag{7}$$

The terms w_{tt} represents the targets observed, w_{sm} denotes the affinity of the target appearance, w_{or}, w_{ch}, w_{tp} are the data terms includes the physical limitation to promote a smooth target trajectory, w_{sd} represents the regularized solution. $k_1,\,,k_5$ are the parameters weight which is set according to the type of video dataset used, and it highly depends on the details of implementation. The state X* is used to minimize the energy.

$$X^* = arg \min_{X \in R^d} W(X) \tag{8}$$

Here d represents the search space, depending on the length and the target count, its values varies accordingly. X is the world coordinates of all targets in all frames.

2) Target Tracking

A popular method of tracking by detection is applied to track the pedestrian in video sequences. The SVM detector based on sliding window is used to locate the pedestrian. The histogram of gradient is included in the detector to extract the feature of the pedestrian. The Non-maxima suppression detects the peaks and transform into image evidence, which is considered as a global coordinate system for tracking. The trajectories of the target which has been kept closer to the observations are the main objective of the data term.

$$w_{tt}(X) = \sum_{j=1}^{T} \sum_{x=u_j}^{v^j} \left[O_j^x . \varepsilon - \sum_{g=1}^{G(x)} \omega_g^t \frac{r_g^2}{X_j^x - G_g^{x2} + r_g^2} \right] \tag{9}$$

The energy increases proportionately with difference of the calculated target location X_j^x and the detected location G_g^x. The visibility fraction O_j^x is used to scale the unseen target, r_g is the scalar. T represents the total number of targets, ε is the offset. x is the image coordinates of target i in frame t. ω_g^t is the weight for each detection.

3) Operational representation

The relative difference in movement of the target and the slow frame rate can be handled by introducing a constant velocity model which will reduce the gap between consecutive velocity vectors.

$$w_{or}(X) = \sum_{j=1}^{T} \sum_{x=u_j}^{v_j-2} \| X_j^x - 2X_j^{x+1} + X_j^{x+2} \|^2 \tag{10}$$

This model helps in reducing the switching of identities and supports the straight path. This model also smoothen the most of the misaligned detections. The smooth target trajectory produced is known as the intelligent smoothing.

4) Collision Handling

Another important issue while tracking the multiple targets is the collision. This model can handle the collision by applying a penalty when one or more target comes close to one another.

$$w_{ch}(X) = \sum_{i=1}^{S} \sum_{j \neq i}^{T(x)} \frac{r}{\| X_i^x - X_j^x \|^2} \tag{11}$$

This model can handle difficult problems in the collision avoidance of the targets. The continuous terms of operational model and collision avoidance model will be helpful to achieve the data association indirectly. This will also provide an interpretation of data, in addition to the plausible trajectory and pleasing visualization.

5) Trajectory Processing

Fragmentation and sudden termination of tracking will occur when the target evidence is missed within tracking area. Hence, it is advisable to start and end the trajectories, once the target reaches the frame border. The tracking which do not abide the rule are penalized. The sigmoid is utilized in the middle of the border region.

$$w_{tp}(X) = \sum_{\substack{j=1,\dots S \\ x \in \{u_j, v_j\}}} \frac{1}{1 + e(-w.b(X_j^x) + 1)} \tag{12}$$

W is the entry edge and it is set to w=1/r. S denotes the total number of frames. The starting and ending frame of the trajectory is denoted as u and v respectively.

6) Standardization

Lastly, to fit the data accurately, standardization is introduced to stop the arbitrary growing of the number of targets. It is handled by penalizing the number of previous targets. This model reduces the un-liked short tracks from the scene by comprising the trajectory length and the standardization term, thereby attaining a better performance.

$$w_{sd}(X) = T + \sum_{j=1}^{T} \frac{1}{S(j)} \tag{13}$$

C. Mean Shift Target Tracking

The Mean shift tracking [12] plays an important role in tracking of targets in an occlusion situation. This can handle different types of occlusions, scale variations and complex backgrounds. A similarity matching can be represented by obtaining the local minima between the actual target and reference target. It also uses various features in order to determine the target scale.

1) Similarity Model

The similarity measure between the target and the reference target is determined as a function. The RGB colour features with 16 bins per channel are used, in which large area is divided into number of sub-areas, each area has its own histogram. The similarity measure is achieved by applying the Bhattacharyya distance between sub-areas in order to obtain the relative closeness of the targets.

$$w_{sm}(X) = \overrightarrow{p_u^R} = C_h^R \sum_{j=1}^{N} \omega_i K\left(\frac{t - X_j^{x\,2}}{s}\right) \delta\left[h(X_j^x) - w\right] H(X_j^x, R) \tag{14}$$

$$H(Q_j^x, R) = \begin{cases} 1, \; if\ the\ pixel\ at\ location\ Q_j^x\ belongs\ to\ region\ R \\ \qquad\qquad 0, \; otherwise \end{cases} \tag{15}$$

The binning function is represented as H. h is the histogram associated with the pixel location. δ is the Kronecker delta function. w is the probability of the feature in the target given w=1....m. K is the kernel and weights ω_i are given by

$$\omega_i = \sum_{R=1}^{N} \sum_{w=1}^{n} S^R C_s \sqrt{\frac{\overline{q_w^R}}{\overline{p_w^R}(t_0)}} \delta\left[h(Q_j^x) - w\right] H(Q_j^x, R) \tag{16}$$

S^R represents the size of sub region R in pixels. C_s is the normalization constant and s is the kernel bandwidth, n is the normalized pixel location in the candidate region. The reference target and the candidate target are represented as q and p respectively.

2) Color features

The mean shift algorithm applies the colour histogram feature to obtain solution for the occlusions, scale variation, etc. There are few difficulties which occur when the target and the background colours are similar. According to this method, a three dimensional colour histogram will distinguish the target affinity.

3) Energy Reduction

The gradient based optimization techniques is chosen to minimize the differentiable energy components in a closed form. A six set of jump moves such as growing, shrinking, adding, removing, merging, splitting are iteratively processed for each greedy parameter selection and the conjugate gradient descent are performed finally to attain the independent optimization of individual targets upon convergence.

IV. Experimental Studies

The performances of the MTT algorithms are measured using the common metrics and the datasets. It is also compared with the other methods in order to validate the proposed algorithms.

A. PETS2009

The state-of-the-art multiple target tracking methods are evaluating using the publicly available datasets. Firstly, PETS 2009 consists of more video sequences with multiple views and its length varies from 90 to 800 frames. The videos are captured using a high calibrated camera with the pixel resolution of 768 X 576 with 7 fps. In each frame the number of target individual varies from 7 to 42. In this work, seven video sequences are chosen for testing. The first view of six video sequence (S1l1-1, s1l2-1, s2l2, s2l3) are captured in a crowded environment which is specifically used for the event recognition or density estimation and the remaining two video sequences (s2l1 and s3Mf1) are captured in a medium crowd. In addition to the above sequence TUD-Stadtmitte is also employed. This video is captured at a low view point in a very busy pedestrian street with 179 frames. This difficult video is applied to estimate the precision of the tracker.

B. CAVIAR

The video sequences are classified into two set, firstly, the scenario is captured using a single camera at the lobby and second at the corridor of a shopping mall. The groundtruth is obtained as a bounding box on all observed person. It also includes some complex occlusion and bad contrast background image. The fig.2 is the tracking results of CAVIAR dataset. The resolution of the captured video is 385 X 288 pixels and 25 frames per second.
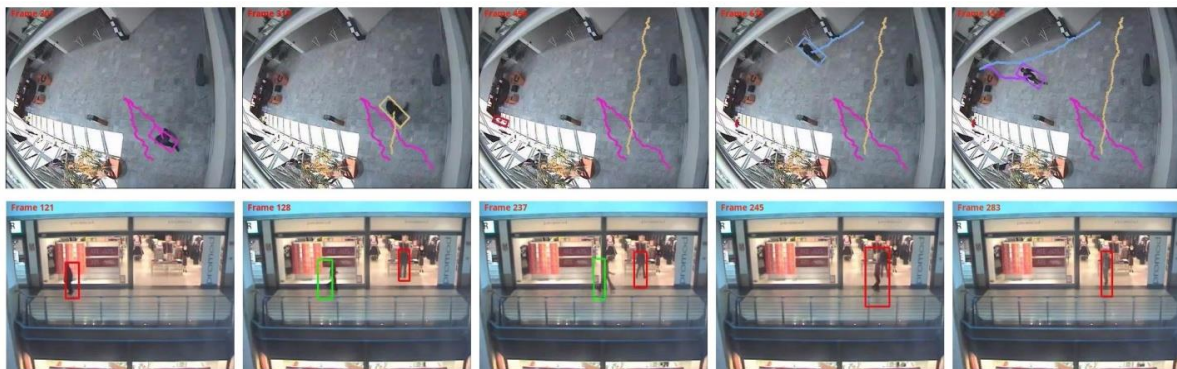

Fig. 2. The tracking video sequence from CAVIAR dataset: Row 1 represents the tracking result of office lobby video and Row 2 represents the tracking result of shopping mall sequence.

V. Metrics

A fair and standard metrics are used for the evaluation of the multiple target tracking approaches. In the earlier MTT, the tracking-by-detection strategy is employed to determine the performance of detection and tracking. Here, metric for MTT is employed and discussed in the components of metrics for the clear understanding.

- **MTTA** ↑ (Mean Target Tracking Accuracy): This metric measures the accuracy of the target tracking algorithm. This measure comprises of the three elements, such as false negative rate, mismatch and false positive to determine the accuracy of the method.
- **MTTP**↑ (Mean Target Tracking Precision): This metric determines how precisely the target is tracked with respect to the groundtruth

- **MTT↑** (Mostly Tracked Trajectories): Above 80% of the trajectory length are successfully covered by the tracker with respect to the groundtruth which are considered as mostly tracked.
- **MLT↓** (Mostly Lost Trajectory): Less than 20% of the trajectory length is covered by the tracker compared to the real scenario which is called as Mostly Lost.
- **IDS↓** (Identity switches): It defines the number of times the trajectory switches their identity.
- **FRGM↓** (Fragmentation): It defines the total number of times the trajectory is interrupted.
- **FP↓ and FN↓** (False Positive and False Negative): It defines the total number of false positive and false negative respectively.

VI. Evaluation

The performance evaluation and comparative study of multi-target tracking is important and a challenging task. The different video sequences of the datasets are tested using the different methods in order to enable the work of the fair comparison. The video sequence from the CAVIAR, PETS2009 and TUD-Stadtmitte datasets are applied to influence training. The task of the computation and the measurement are performed using the CLEAR MOT metrics. In the proposed works of GMM-beta likelihood and Global optimization techniques depends on the parameters. The values of the parameters are set according to the requirements of the video sequences used, in order to achieve a maximum MTTA and MTTP scores.

Table 1 The results obtained by applying the S2L1, S2L2 and the St. George street crossing video sequences on the proposed GMM methods. Apart from the pedestrian, St. George street crossing video comprises of a varying backgrounds, swaying of the leaves, illumination changes and vehicle crossing. When the number of events is very large, the Gaussian distribution is applied to describe the physical events. The mean accuracy of the target tracked is above 90% and the precision is 70.2% which is also better. Initially, it takes little iteration of the distributions to distinguish the active Gaussian background and foregrounds. Once, the reference background is obtained, the target has been detected and tracked accurately. This type of the method can be used for the long run of the applications. The cost optimization is achieved by implementing the Munkres algorithm. This algorithm models an assignment problem as an N X M cost matrix, where each element represents the cost of assigning i-th frame to the j-th process and it figures out the least cost solution, choosing a single item from each row and column in the matrix, such that no row and no column are used more than once. It runs in $O(n^3)$ times rather than $O(n!)$. It is also used to maximize the likelihood matching of the targets.

Table 1. Result of GMM-BLM-KF on PETS2009

SEQUENCE	MTTA	MTTP	GT	MTT	MLT	FP	FN	IS	FRGM
PETS-S2L1	89.7	69.9	23	20	1	65	227	7	12
PETS-S2L2	50.1	51.96	74	37	9	704	3040	114	79
STREET CROSSING	90.3	70.2	17	15	1	59	224	6	11

Table 2 gives the results obtained from global energy minimization and optimization method. The metrics is applied on all video sequences individually. The performance of the tracker is varied upon the number of the targets encountered in the frame. The video sequences such as PETS-S2L1, TUD-Stadtmitte comprises of less than 10 targets in a frame. In these scenario, the MTTA are over 90% and 70% respectively, shows a better performance, because all the target pedestrian are visible all the time, and contains less occlusions. However, MTTA is reduced to 58%, because PETS-S2L2, S2L3 are the challenging datasets consisting of more than 40 targets which appears in the same frame with severe occlusions. The performance are measured based on the number of target, in this case, six dataset is divided into two groups, the group one consist of the video sequence with targets less than 10 pedestrians with average accuracy is 81.95% and the second group is above 40 targets with the mean accuracy of 48.52%. The accuracy of the tracking result increases by 8% in the case of most difficult scenes (PETS-S2L2), and the number of target mostly tracked rises by 37 %(

approx), only less than 10% of the target trajectories are missed due to the occlusion, which is considered as better results in MTT.

Table 2. Results of GEM-OT on PETS2009 Video sequence

SEQUENCE	MTTA	MTTP	GT	MTT	MLT	FP	FN	IS	FRGM
PETS-S2L1	**91.2**	**80.9**	**23**	**21**	**1**	**56**	**303**	**10**	**5**
PETS-S2L2	**57.9**	**60**	**74**	**32**	**8**	**620**	**2678**	**100**	**70**
PETS-S1L3	46.2	65.1	44	9	17	163	1576	35	26
PETS-S1L1-2	58.4	61.3	36	18	10	146	910	21	12
PETS-S1L1-1	31.6	50.1	43	7	18	230	2301	60	34
TUD-STADTMITTE	72.7	67.1	9	7	0	89	104	3	3

Table 3, 4 and 5 discusses the comparative results on the PETS2009-S2L1, PETS2009-S2L2, and CAVIAR video sequence with other methods. The proposed global optimization result has shown a better performance over the other multiple target tracking methods in terms of accuracy, target identity switching, fragmentation, etc. Here, the MTTA of the S2L1, S2L2 are 2% higher than the other methods and it is more than 5% in CAVIAR dataset thereby reducing the false positive rate, number of the mismatches and false negative rate. The mostly tracked trajectory is over 90% successfully covered by the tracker with respect to the groundtruth which is 10% higher than the assumed limit. The mostly lost trajectory is 10% to 12 % which is lower than the expected which is a good indication for the best tracking. The number of identity switching between the targets is 4% approximately and thereby the rate of fragmentation is reduced considerably.

Table 3. Comparative results on the PETS2009S2L1 video sequence

	Methods	MTTA	MTTP	GT	MTT	MLT	FP	FN	IS	FRGM
PROPOSED	GMM-BLM-KF	89.7	69.9	23	20	1	65	227	7	12
	GEM-OT	**91.2**	**80.9**	**23**	**21**	**1**	**56**	**303**	**10**	**5**
EXISTING	Berclaz et al.[13]	80.3	72	23	7	1	-	-	13	22
	Henriques et al.[20]	84.8	78.7	23	-	-	-	-	10	-
	Andriyenko and Schindler et al.[22]	86.3	78.7	23	78	4	-	-	38	21
	Milan et al.[23]	90.6	80.2	23	21	1	59	302	11	6

Table 4. Comparative results on the PETS2009S2L2 video sequence

	METHODS	MTTA	MTTP	GT	MTT	MLT	FP	FN	IS	FRGM
PROPOSED	GMM-BLM-KF	50.1	51.96	74	37	9	704	3040	114	79
	GEM-OT	**57.9**	**60**	**74**	**32**	**8**	**620**	**2678**	**100**	**70**
EXISTING	Berclaz et al.[13]	24.2	60.9	74	1	5	-	-	13	38
	Breitenstein et al.[21]	50	51.3	74	-	-	-	-	-	-
	Andriyenko and Schindler et al.[22]	60.2	63	74	-	-	-	-	104	-
	Milan et al.[23]	56.9	59.4	74	28	12	622	2881	99	73

Table 5. Comparative results on the CAVIAR video sequence

	METHODS	MTTA	MTTP	GT	MTT	MLT	IS	FRGM
	GEM-OT (PROPOSED)	**92.8**	**89.7**	**140**	**130**	**2**	**10**	**8**
EXISTING	Zhang et al.[19]	87.2	76.3		85	1	6	32
EXISTING	Bao and Yoon et al.[24]	86.5	87.2		90	1	9	8
EXISTING	Wu et al. [25]	77.7	72.6		106	9	17	35

VII. Conclusion

The performance of the all the tracking methods are discussed on the various perspectives of issues such as occlusions of targets, varying number of targets, appearance modelling, camera viewpoints and affinity matching. The results of the various methods are discussed and it is clearly understood that the proposed Global Energy Minimization and Optimization Technique is ahead of all the trackers, because it precisely localizing the target locations, considerably decreases the false positive counts, number of mismatches, and false negative rate which is an essential factor in the comparison of the MTT applications. The experimental evaluation on various complex datasets gives a better result compared to the other state-of-the-art methods.

References

[1] J. Black, T. Ellis, and P. Rosin. Multi view image surveillance and tracking. In Motion & Video Computing Workshop, Dec. 2002.

[2] D. Reid. An algorithm for tracking multiple targets. IEEE Transactions on Automatic Control, 24(6):843–854, Dec. 1979.

[3] K. Okuma, A. Taleghani, O. D. Freitas, J. J. Little, and D. G. Lowe. A boosted particle filter: Multitarget detection and tracking. In ECCV 2004, volume 1, pages 28–39.

[4] J. Vermaak, A. Doucet, and P. Perez. Maintaining multi-modality through mixture tracking. In ICCV 2003.

[5] T. E. Fortmann, Y. Bar-Shalom, and M. Scheffe. Multi-target tracking using joint probabilistic data association. In IEEE Conf. on Decision and Control, volume 19, pages 807–812, Dec. 1980.

[6] Michael Kamaraj and Balakrishnan. An improved motion detection and tracking of active blob for video surveillance, 4th ICCCNT 2013, ISBN: 978-1-4799-3927-5, Page 569-575.

[7] B. Leibe, K. Schindler, and L. Van Gool. Coupled detection and trajectory estimation for multi-object tracking. In ICCV 2007.

[8] Michael Kamaraj and Balakrishnan. Surveillance of human tracking using Gaussian beta-likelihood matching and kalman filter, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 14 (2015) pp 34375-34382 © Research India Publications .Available at: http://www.ripublication.com.

[9] Michael Kamaraj and Balakrishnan. Human Motion tracking using Gaussian Mixture model and Beta-likelihood matching, IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661, p-ISSN: 2278-8727 PP 40-48.

[10] H. Jiang, S. Fels, and J. J. Little. A linear programming approach for multiple object tracking. In CVPR 2007.

[11] Michael Kamaraj and Balakrishnan. Global energy minimization and optimization of multi-target tracking. 2nd ICEEET16, March 2016.

[12] Michael Kamaraj and Balakrishnan. Optimization of multi-target and Occlusion handling using mean shift method, International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X , Volume 5, Issue 9, September 2015.

[13] J. Berclaz, F. Fleuret, and P. Fua. Multiple object tracking using flow linear programming. In winter-PETS, Dec. 2009.

[14] M. Rodriguez, I. Laptev, J. Sivic, and J.-Y. Audibert. Density-aware person detection and tracking in crowds. In ICCV 2011.

[15] J. Xing, H. Ai, and S. Lao. Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses. In CVPR 2009.

[16] L. Kratz and K. Nishino. Tracking with local spatio-temporal motion patterns in extremely crowded scenes. In CVPR 2010.

[17] Choi W, Savarese S, Multiple target tracking in world coordinate with single, minimally calibrated camera. In:Proc. Eur. Conf. Comput. Vis., 2010, pp 553-567.

[18] Qin Z, Shelton CR, Improving multi-target tracking via social grouping. In: Proc. IEEE Int. Conf. Comput.Vis. Pattern Recognit. 2012, pp 1972-1978.

[19] Zhang L, Li Y, Nevatia R, Global data association for multi-object tracking using network flows. In:Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. 2008, pp 1-8.

[20] Henriques JF, Caseiro R, Batista J, Globally optimal solution to multi-object tracking with merged measurements. In: Proc. IEEE Int. Conf. Comput. Vis. 2011, pp 2470-2477.

[21] Breitenstein MD, Reichlin F, Leibe B, Koller-Meier E, Van Gool L, Robust tracking-by-detection using a detector confidence particle filter. In: Proc. IEEE Int. Conf. Comput. Vis. 2009,pp 1515-1522.

[22] Andriyenko A, Schindler K, Multi-target tracking by continuous energy minimization. In: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognition 2011, pp 1265-1272.

[23] Milan A, Roth S, Schindler K, Continuous energy minimization for multi-target tracking. IEEE Trans. Pattern Anal Mach Intel. 2014, 36(1):58-72.

[24] Bae SH, Yoon KJ, Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning. In: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. 2014, pp 1218-1225.

[25] B. Wu and R. Nevatia, Tracking of multiple, partially occluded humans based on static body part detection, In *CVPR*, 2006.

[26] Stauffer C., and Grimson W., "Adaptive Background Mixture Models for Real-Time Tracking" in Proceedings of IEEE Computer Vision and Pattern Recognition, Vol. 2, 1999, pp.246-252.

[27] B.Z. de Villiers, W.A.Clarke, P.E. Robinson, "Mean shift Object Tracking with Occlusion Handling", Available at: www.prasa.org/proceedings/2012/prasa2012-36.pdf.