DOI: 10.21917/ijivp.2017.0206

MULTIPLE TARGET TRACKING USING COST MINIMIZATION TECHNIQUES

Michael Kamaraj[1] and G. Balakrishnan[2]

[1]Department of Computer Applications, Pavendar Bharathidasan College of Engineering and Technology, India
[2]Department of Computer Science and Engineering, Indra Ganesan College of Engineering, India

Abstract

Many applications such as intelligent transportation, video surveillance, robotics of computer vision mainly depend on task of multiple target tracking. It consists of process of detection, classifications and tracking. In this novel approach of multi target tracking, cost terms are formulated to attain the global optimization which includes the entire representation of the issues such as target tracking, operational representation, collision handling and trajectory processing. Furthermore, two optimization strategies such as the gradient descent which is performed on multiple feature space to obtain local minima of a density function from the given sample of data and gradient ascent which is carried out to achieve a likelihood matching of the target and used to handle the partial evidence of the image, and also uncertainty of the various targets are minimized. . In this study, the proposed works are tested on different publicly available datasets using the metric evaluation and also compared with the various methods based on issues of target tracking. This study will also provide a better understanding of the problem, knowledge of the methods, and experimental evaluation skill for further research works.

Keywords:

Multiple Target Tracking, Surviellance, Cost Minimization, Optimization, Tracking Metrics

1. 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 MTT, 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 multi-target tracking, more number of targets is observed simultaneously. 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. The severe occlusion of highly crowded humans is responsible for the problems in detecting multiple humans as the complexity of varying background, diverse appearances, viewpoints and different postures related to the challenges of this well-known task. In general, any object which is not clearly observable due to any one of the three types of occlusions like inter object occlusion, self-occlusion and non-object occlusion is treated to be an occluded object. In this research work, the focus is on to improve the accuracy and the precision of the multiple target tracking of pedestrians, where it has shown significant progress. Nonetheless, many recent techniques only offer good performance with easy conditions where there are few numbers of targets. The performance of these algorithms tends to decrease when detecting multiple humans in highly crowded scenes where humans can be severely occluded. Hence, the data association, trajectory estimation of each target, and also the multiple possible identities complicate the trajectory prediction. The process of correctly matching the identity of the target for corresponding detection is known as data association. These are all the problems exist ahead of all the trackers.

In this work, the stated problems are concentrated, and an attempt is made to design the objective function such that it offers a more complete representation of the various aspects of the problem. The cost function is framed and defined in continuous space in order to guarantee global optimality. The cost depends on the observation of target locations and their movements in all the frames, the identity switches which reduce dynamically, mutual exclusion which improves the smoothness of the trajectories, trajectory persistence which helps in tracking of the missing evidence and standardization which prevents the arbitrarily growing of the targets. These terms are modelled in a continuous domain to express the situation close to the real. The conjugate gradient descent and trans-dimensional jump moves are applied to lower the cost and the gradient ascent is to find the best match between the target region and the various candidate regions to achieve the efficient and effective global optimization. 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. Furthermore, greedy search and sampling-based are used to analyse the optimization strategies and the study on the effect of all main parameters are experimented.

2. RELATED WORKS

In the recent years, the computer vision has performed a wide evaluation on the numerous tasks, such as object detection [1], pedestrian detection [2], 3D reconstruction [3] , optical flow [4],[5], etc. and the interest for performing the systematic evaluations in tracking approaches with common databases and metrics increases rapidly. The literature survey states that the multi object tracking has been an active area for the researchers. Initially, several algorithms based on recursive methods [6,7] were using recursive approach for tracking of the targets. In this method, Kaman 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 [8,9]. 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 [10] or probabilistic filtering. In this study, the main focus is on the new advancements in multiple target tracking. Some recent multiple target tracking formulations intend to find global optimal trajectories within a temporal frame area are discussed from [11-15]. In the previous paper, Michael et al. [16],[17] proposed Gaussian likelihood matching method to integrate the adaptive background detection, and combined data for the cost effectiveness and cost minimization in tracking of multiple targets. The complete and incomplete occlusions, scale changes and multifaceted backgrounds are highly invariant to mean shift clustering. This type of occlusions and ambiguous targets of appearances are handled using this clustering method. In addition, the optimization strategy is applied here. Leibe et al. [18] designed a quadratic binary algorithm by applying the heuristic technique that connects the task of detection of objects and estimate trajectory offer solution to the problem of local optimization. Jiang et al. [19] attempted to give the global optimized solution for multiple targets tracking using the integer linear program with certain linear constraint to emphasize the design to maintain the constant target between the frames, but still no guaranteed solution is obtained. Zhang et al. [20] introduced a network flow model in which graph was used for global multi target tracking. The movement of the target in the frame form edges between the respective detections and represent the likelihood of target. The min-cost flow algorithm presents an optimal set of path without taking occlusion into the consideration. Rodriguez et al. [21] employed an algorithm for tracking of people in crowded environment, there are few limitations in the implementation of the cost terms and the camera shooting points are not viable for certain surveillance applications. Xing et al. [22] created a tracklet with no occlusion, then an effort to overcome the problem of occlusion by linking the short tracklet to lengthy trajectories, but still the occlusion reasoning cannot be met. Wojek et al. [23] developed a 3D model to detect the whole body using the various parts of detectors, each target detections are weighted according to its respected visibility. This is somewhat closely related to the discussion of multi target detection. Breitenstein et al. [24] proposed an algorithm to raise the target detect rate. Accordingly, whenever there is another close target, the likelihood increases. Anton Milan et al. [25] developed a framework for global optimization of multi target tracking using a continuous energy minimization terms. In addition to the optimization of tracking multiple objects, the non-convex energy is minimized by descent gradient method and a set of discontinuous jump moves; however the resulting tracking performance is not able to achieve higher recall percentage.

According to Kratz and Nishino [26] studied the motion patterns in the crowded environment using the spatio-temporal method. The target likelihood is calculated by converting the distance of colour histogram into probability using the exponential function. Choi and Savarese [27] employed a mean shift tracker which utilizes colour histogram to detect the target sequentially. Qin and Shelton [28] 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. Collin et al. [29] adopted a scale space theory in order to determine the target objects while tracking. An explicit scale-space approach for tracking objects whose size varies over time. Further extending the flexibility of gradient ascent tracking J. Ning et al. [30] designed a framework in which the Mean Shift method was used at various scale of Gaussian kernels to decide the target object between the target model and the candidate model. Zhang et al. [31] developed an algorithm to simultaneously track the position, scale and orientation of the target, with the help of numerous ellipsoidal and unsymmetrical kernels. In addition to it, Yilmaz et al. [32] defined a set of function to track the object along the contour. De Villiers et al. [33] presented a framework of the Mean Shift method and the localization algorithm to track the target object using multiple features and Kaman filter. As a result, the algorithm uses features such as local binary pattern, edge, colour and background weighted colour features which handle the occlusion of the target in the complex scene. Zivkovic and Krose method [34] is based on expectation maximisation and is capable of following regions undergoing translation, rotation and changes in scale and aspect ratio. The algorithm of Nguyen et al. [35] can also recover the scale of the object by treating the pixel coordinates of the target as latent variables to be estimated again by an expectation maximisation algorithm. Regions undergoing affine transformation are handled by Guskov's tracker [36], which can also accommodate illumination changes. In the previous research, an effort to develop a tracker does not produce a comprehensive minimization solution to the tracking. Hence, in this proposed work, the novel technique for the occlusion finding and the cost minimization are proposed in multi target tracking using an optimization strategy of two gradient based algorithms.

The paper is organized as follows, section 3 & 4 discusses individually about multiple target cost minimization terms in detail; Section 5 covers the similarity model; Section 6 deals the optimization strategy followed in tracking of moving multiple objects and occlusion handling; Section 7 composed of cost minimization; Section 8 comprises the Dataset and metrics used for evaluation; Section 9 consists of experimentation and its evaluation and finally section 10 present the conclusion.

Table.1. Common Notations

Variable	Comments
Q	Global coordinates of all targets in all frames
Q_j^s	(P,Q) global coordinates of target j in frame s
S,T	Total number of frames and targets respectively
$S(j)$	Number of frames where target j is present
u_j, v_j	Starting and ending frame of trajectory j
$T(x)$,$S(x)$	Number of targets, respectively detections in frame x
G_g^x	(P,Q) global coordinates of detection g in frame x

3. MULTIPLE TARGET TRACKING

Tracking of Multiple targets simultaneously 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 cost 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 cost minimisation terms are formulated to develop an efficient and global optimal solution of multi-target tracking system.

3.1 PRIMARY NOTATION

The common notations used in this paper are introduced in Table.1 for a simple understanding. The global coordinate (*P*, *Q*) of *T* targets in all the frames *S* are defined in the vector of state *Q*. It is assumed that the plane $k = 0$, since all the target movements are taken from the common ground. Q_j^s is the target location in which *j* remains the target at the time *s*. Furthermore it includes the non-associated targets and occluded detection within the length of the trajectory are expressed as $x \in \{u_j...v_j\}$. The trajectory length of the target *j* is expressed as $S(j) = (v_j - u_j + 1)$, here u_j and v_j are the starting and ending frames. The targets in each frame differ from one to another. $T(x)$ is taken as number of targets in frame *x* and $G(x)$ denoted as number of detection in frame x. $\underline{q}_g^x$ is indicated as detected location *g* in frame *x*. The (*p*, *q*) is the image coordinate of target *j* in frame *x* is denoted as $\underline{q}_j^x$. *S* and *T* are the total number of frames and targets respectively.

4. COST MINIMIZATION FUNCTION

The cost 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 cost terms are developed in a closed form to obtain an efficient gradient optimization solution. Each cost term linearly combined to form a cost function.

$$W = w_{TT} + \alpha_1 w_{SM} + \alpha_2 w_{OR} + \alpha_3 w_{CH} + \alpha_4 w_{TP} + \alpha_5 w_{SD} \tag{1}$$

w_{TT} is the data term presents solution in proximity to the observation. w_{SM} is the similarity model used for unambiguous data association of the various targets. w_{OR}, w_{CH} and w_{TP} gives the possible movements and enforce physical constraints. w_{SD} is the standardization term maintains a simple solution and avoids over fitting. From an optimization perspective, it would certainly be beneficial to have a complex function where a global optimized solution can be achieved to minimize the cost which is independent of initial values. Hence, the aim is to discover the state Q^* that minimizes the high dimensional continuous cost from Eq.(1),

$$Q^* = \underset{Q \in R^d}{\arg} \; minW(Q) \tag{2}$$

Here, *d* is the dimension of the search space and it relies upon the length and the number of objects. The remainder of the section explains all function of cost terms separately in detail.

4.1 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.

$$w_{TT}(Q) = \sum_{j=1}^{T} \sum_{x=u_j}^{v^j} \left[O_j^x . \varepsilon - \sum_{g=1}^{G(x)} \omega_g^t \frac{r_g^2}{Q_j^x - G_g^{x2} + r_g^2} \right] \tag{3}$$

ω represents the quantity that weights detection, *r* is the scalar quantity of the object size, ε is the offset included for all existing target as penalty for false detection of image and it is excluded for occluded target which is not observed by the detector. Therefore, visibility fraction O_j^x is used to scale the unseen target.

4.2 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.

$$w_{OR}(Q) = \sum_{j=1}^{T} \sum_{x=u_j}^{v_j - 2} \left\| Q_j^x - 2Q_j^{x+1} + Q_j^{x+2} \right\|^2 \tag{4}$$

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.

4.3 COLLISION HANDLING

Collision avoidance represents multi target tracking aspect. The general for formulating the collision avoidance is to add high penalty when two targets move towards each other.

$$w_{CH}(Q) = \sum_{i=1}^{S} \sum_{j \neq i}^{T(x)} \frac{r}{\left\| Q_i^x - Q_j^x \right\|^2} \tag{5}$$

The scale factor *r* for people tracking and this value goes to infinity when they share one identical position. The problem of mutual exclusion handles two specific problems. Firstly, the overlap between the targets is checked at all times, even if both the targets are occluded. Secondly, the targets will separate the missed targets due to inaccurate observations considerably to avoid an unrealistic state. The flexible assignments of observation model and mutual exclusion are the two data terms responsible for the indirect data association, thereby producing a more satisfying and more realistic trajectories.

4.4 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,

$$w_{tp}\left(Q\right)=\sum_{\substack{j=1,\ldots S\\ x\in\left\{u_j,v_j\right\}}}\frac{1}{1+e(-w.b\left(Q_j^x\right)+1)} \tag{6}$$

where, $b(Q_j^{u_j})$ and $b(Q_j^{v_j})$ compute the distance of the initial and corresponding final recognized location of target j to nearby edge of the tracking area. *W* is the entry edge and it is set to w=1/r. The starting and ending frame of the trajectory is denoted as u and v respectively.

4.5 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,

$$w_{SD}\left(Q\right)=T+\sum_{j=1}^{T}\frac{1}{S(j)} \tag{7}$$

where, *S*(*j*) be the temporal length of trajectory *j* in the frame.

5. SIMILARITY MATCHING

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.

$$w_{sm}\left(X\right)=\overrightarrow{p_u^R}=C_h^R\sum_{j=1}^{N}\omega_i K\left(\frac{t-X_j^{x2}}{s}\right)\delta\left[h\left(X_j^x\right)-w\right]H\left(X_j^x,R\right) \tag{8}$$

$$H\left(Q_j^x,R\right)=\begin{cases}1, & \text{if the pixel at location } Q_j^x \text{ belongs to region } R\\ 0 & \text{otherwise}\end{cases} \tag{9}$$

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,

$$\omega_i=\sum_{R=1}^{N}\sum_{w=1}^{n}S^R C_s\sqrt{\frac{\bar{q}_w^R}{\bar{p}_w^R\left(t_0\right)}}\delta\left[h\left(Q_j^x\right)-w\right]\mathrm{H}\left(Q_j^x,\mathrm{R}\right) \tag{10}$$

S^R represent 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.

5.1 COLOUR 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.

Algorithm 1: Gradient descent procedure

Step 1: Choose initial Q^j to Detection *G*

Step 2: Repeat $Q^{j+1} \leftarrow Q^j - \eta\nabla W(Q^j)$ until the sequence $\|W(Q^j)$-$W(Q^{j-1})\| > \eta$ converge

Step 3: Repeat step 2 for every jump moves *x*

Step 4: Perform conjugate gradient descent $x \in$ {increase, decrease, inclusive, exclusive, integrate, differentiate} till step 3.

Step 5: Return $\arg\min_{Q^j} W\left(Q^j\right)$

5.2 OPTIMIZATION STRATEGY

The two gradient optimization strategies followed by the similarity model term considerably decreases the quantity of identity switches, reduces the fragmentation and increases the precision of tracking. An algorithm 2 is based on the gradient ascent to maximize a likelihood matching between the model and the current image region. An algorithm 1 is based on the gradient descent method for finding the nearest local minimum of a function.

Algorithm 2: Gradient Ascent procedure

Step 1: Choose initial Q^j randomly

Step 2: Repeat $Q^{j+1} \leftarrow Q^j - \eta\nabla W(Q^j)$ until $\|W(Q^j) - W(Q^{j-1})\| > \eta$ (small) converges

Step 3: Perform conjugate gradient ascent for every jump moves *x* and till step 2 converges.

Step 4: Returns $W(Q^j)$

6. COST REDUCTION

Each cost component is differentiable in closed form, making the entire formulation well suited for gradient-based minimization. A standard conjugate gradient descent is applied to minimize Eq.(1) locally. However, the given convex nature of the cost, a purely gradient-descent optimization would be very susceptible to initialization. Therefore, a set of jump moves are added. These non-local jumps in the cost landscape change the trajectory lengths and potentially the number of targets, thus allowing more flexible probing of the solution space to escape weak minima. Upon convergence of the gradient descent, one of six jump moves described below is executed in a greedy fashion.

Increasing and Decreasing: Each trajectory can be extended by linear extrapolation for an arbitrary number of space-time both forward and backward. Similarly, a track is shortened by discarding a fragment of a certain length from either ends. The growing is useful for ending new targets, while shrinking weeds out false positives that may have been introduced by noise or during intermediate optimization steps.

Integrating and Differentiating: the split yields the lower cost when two existing trajectories are merged into one. The individual cost components, the dynamics and the exclusion terms are particularly noted and this step is asserted which will not cause physical implausible situations with intersecting trajectories or unlikely motion patterns. A single track may also be split into two at a specific point of time. This move provides a method to bridge over regions with missing sensor responses and to reduce fragmentation of tracks and identity swaps.

Inclusive and Exclusive: These two moves operate on entire trajectories. Thereby, removing a false positive target from the current solution may decrease the overall cost because it results in a plausible explanation of the data. On the other hand, it is important to allow inserting of new tracks around active sensor locations that do not have a target nearby. This is done conservatively by adding a short tracklet of only three frames. Growing and merging with other existing trajectories at a later optimization step is noted.The cost term acquires the minimum values after minimization, to attain the optimal value of the tracked location. The trajectories thus obtained are very smooth.

7. DATASETS AND METRICS

This paper presents an overview of the video datasets used in the experimentation. The scope of the investigation to the tracking of pedestrians is limited to a single camera. The ground truth data and general datasets provide the modern tracking scenarios in the computation of metrics. The proposed tracker ascertain the reliability and the performance metrics used in comparison with the other tracking techniques provide the basis for a quantitative assessment, both in terms of robustness and accuracy.

CAVIAR dataset [37] includes, a collection of video sequences of people walking along a corridor in a shopping centre. The 26 sequences vary in length from 25 frames per second, and also exhibit significant variation in number of people visible in the scene at any one point of time. The videos represent a challenging tracking task for a number of reasons. Firstly, although the scenarios that do not feature occlusion are only considered, people walk in proximity to each other. Secondly, the lighting varies considerably over the area of the scene. A person’s appearance can thus alter dramatically as they enter shadow or as the colour of the light falling on them changes.

PETS 2009-S2L1, S2L2, S2L3, S2L1-2, S2L2-1 [40], and TUD-Stadtmitte dataset is an openly available video sequence proposed for the validation of the tracker. The videos are recorded from an altitude position of observation and identical to a surveillance arrangement. The sequence of S2L1 and S2L2 are having a frame rate of 7 fps and resolution of each frame is 768×576 is widely in the research studies of multiple target tracking. The targets are exposed to significant variation in the appearance due to silhouette and changes in illumination. It also consists of discontinuous motion, closeness of the target and a sight of occlusions.

7.1 METRICS

The proposed design is studied based on the results of individual weight of the cost function. 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 ground truth

MTT ↑ (Mostly Tracked Trajectories): Above 80% of the trajectory length are successfully covered by the tracker with respect to the ground truth 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.

(Note symbol ↑ indicates the higher is better and ↓ indicates lower is better)

8. EXPERIMENTAL STUDY

Some of the implementation details are presented before the experimentation. In this study, the rectangular boundary area for tracking on the ground is chosen for the distance computation; outside this area is not included for the calculations. The run time detection of the MATLAB 7.8 implementation takes one frame per second to get solution using mean shift occlusion handling and the real time performance is achieved by a simple inexpensive computations and speed up the optimization run. To speed up the convergence results, the number of iterations is limited to fifteen. During the experimentation the parameter values are included for accurate tracking and it is mostly relied on the implementation. In order to improve the result the parameters weight d_1, α_2, α_3, α_4 and α_5 and ε are set accordingly to the requirements.

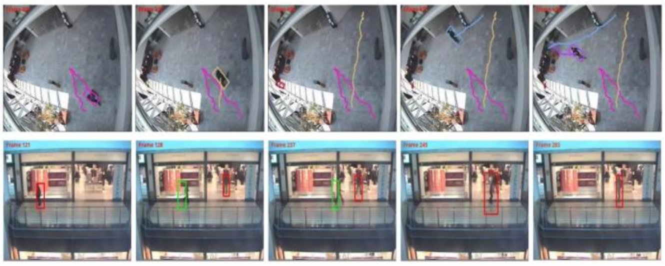

Fig.1. Tracking video sequence from CAVIAR dataset. Row 1 represents the tracking result of office lobby video. Row 2 represents the tracking result of shopping mall sequence

The CAVIAR dataset is composed of two set of video sequences, one was the entrance of the lobby and the other was recorded in a shopping centre. It also provides the ground truth indicating the bounding box around each visible person. The dataset is very challenging because of its heavy occlusions and poor image contrast from background. This dataset includes 26 video sequences of a walkway in a shopping centre taken by a single camera with frame size of 385 × 288 and frame rate of 25fps. The test was conducted on more than 20 videos (25587 frames total) and the training was carried out with the other six videos. The proposed method produces obvious improvements and fragments are greatly reduced on both data sets by over 60%, while keeping other scores competitive. Some visual results are shown in Fig.1.

Table.2. Evaluation of the video sequence of CAVIAR Dataset

Evaluation	AGT	MTT	PT	MLT	IS	FRGM	MTTA
Wu et al. [40]	140	106	25	9	17	35	77.7
Zhang et al. [20]	140	112	27	6	39	11	82.1
Proposed method	140	130	14	2	10	8	92.8

For almost overlapped persons, the trackers do not confuse with the identities and finds the correct associations. It is evaluated by its tracking performance, detection performance and speed. The comparison results shown in Table.2 are reported as the best tracking result on the dataset. The result shows good performance with fewer false alarms and trajectory fragments than the previous methods.

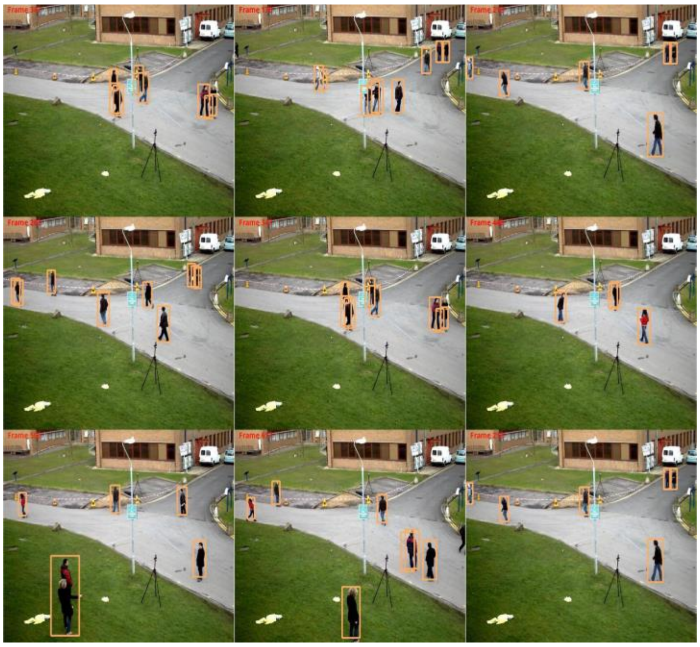

Fig.2. The Tracking of video sequence of PETS09-S3-MF1

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.

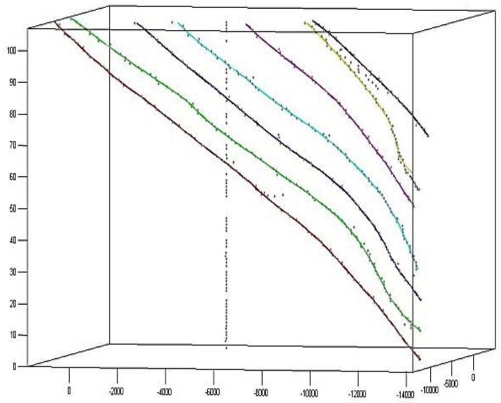

Fig.3. Trajectory of the PETS09-S3-MF1

Table.3. Evaluation of the video sequence of PETS09-S3-MF1

Evaluation	AGT	MTT	PTT	MLT	FP	FN	IS	FRGM	MTTA	MTTP	MTTAL
2D	7	7	0	0	1	9	1	0	97.9	75.4	98
3D	7	7	0	0	1	9	1	0	97.9	86.1	98

A detailed study of the video sequence PETS09-S3-MF1 is discussed in this section. It consists of 107 frame runs at 7fps. Each frame is of size 768×576. The frames are completely executed within the short duration of 0.21mins and the processing of each frame is done within 0.12sec. After 10 epochs the optimization is merged. The Fig.2 displays the sample output results of the proposed work. It is clear that the cost function yields a good accuracy in tracking multiple targets taking occlusion into consideration. The influence of each cost terms and the optimization of convergence rate are measured. From the Table.3 it is observed that the evaluation results of the proposed algorithm achieves 98.1% of mean object tracking accuracy and the tracking precision is 75.2% on 2D evaluation which is improved by 10.8% increase during the 3D evaluation of the dataset. The Fig.4 shows the trajectory moves of all the seven targets in the video sequence.

The experimentation of the proposed method is extended to the test data with S2L2 and S2L3 complex video scenarios with huge density of crowd. The Fig.5 shows the sample tracking results of four video sequences. The performance of the tracker is also tested with two more additional video sequences S1L1-2 and S1L2-1, which are exclusively designed for counting the number of pedestrians and estimation of the density. Finally, the TUD-Stadtmitte video sequences show a significant variation of people walking on a street.

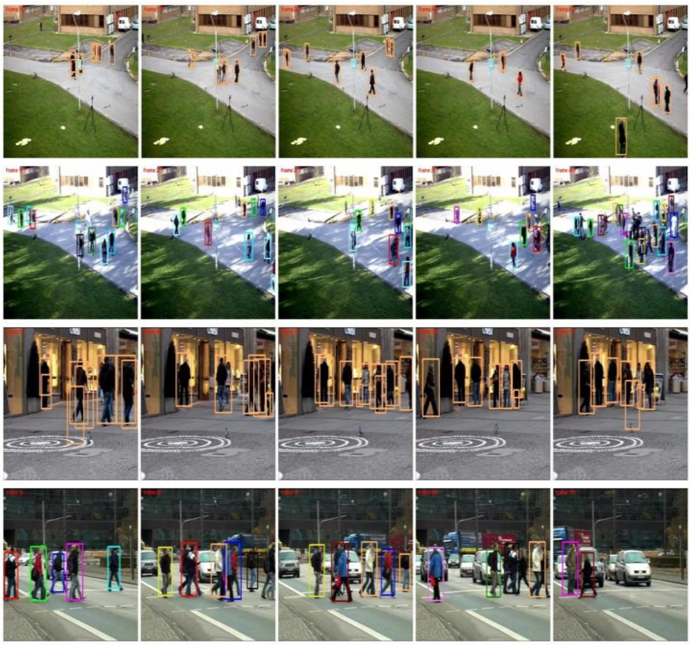

Fig.4. The tracking results of GEM-OT method: The rows 1,2,3,4 are the tracking video sequence of PETS2009 S2L2, PETS2009 S1L1-1, TUD-Stadtmitte and road crossing respectively

Table.4. Quantitative analysis of the proposed method

Sequence	**MTTA**	**MTTP**	**AGT**	**MTT**	**MLT**	**FP**	**FN**	**IS**	**FRGM**
PETS-S2L1	91.2	80.9	23	21	1	56	303	10	5
TUD-Stadtmitte	72.1	66.2	9	7	0	90	106	3	3
PETS-S3-MF1	96.8	82.9	7	7	0	4	12	0	0
PETS-S2L2	57.9	60.0	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-S1L2-1	31.6	50.1	43	7	18	230	2301	60	34
Mean	64.9	66.6	33.7	14.4	7.7	187	1126.6	32.7	21.4

The main objective of the cost function of multiple object tracking is to achieve an optimized solution. The proposed methods are tested on seven video sequences and its quantitative results are described in the Table.4 and Table.5 discusses the comparative results on specific video sequences. The mean performance of the low dense video set shows a considerably high performance results. On an average mean tracking, the accuracy is 64.9% which includes a highly crowded environment. The Table.5 illustrates the comparative results of the proposed method with the other strategic results. It is studied that the average mean performance of the video sequence containing few targets indicates the overall increase in the tracking accuracy. In the less dense group, the entire pedestrians are totally visible in most of the frame; therefore the mean shift occlusion computation cannot show its complete effectiveness. Conversely, for the complex environment such as PETS-S1L2 and PETS-S1L1-2 the mostly tracked target is 43% and 50% respectively that is more than 10% of increase in accuracy compared to the mean. It is observed that the trajectories of the targets are tracked above 90% in the underlying condition of the occlusion, keeping mostly lost fewer.

Table.5. Comparative results on the PETS2009-S2L1 video sequence

METHODS		**MTTA**	**MTTP**	**GT**	**MTT**	**MLT**	**FP**	**FN**	**IS**	**FRGM**
Proposed		91.2	80.9	23	21	1	56	303	10	5
EXISTING	Berclaz et al. [39]	80.3	72	23	7	1	-	-	13	22
	Henriques et al. [38]	84.8	78.7	23	-	-	-	-	10	-
	Milan et al. [25]	90.6	80.2	23	21	1	59	302	11	6

9. CONCLUSION

The performance of the algorithm is explained by using the methods on PETS2009, CAVIAR, and street crossing video sequence with the various perspectives of occlusions of targets, varying number of targets, appearance modelling, camera viewpoints etc. The proposed tracker is flexible by precisely localizing the target locations and the experimental evaluation on various complex datasets gives a competitive result compared to the other state-of-the-art methods. The energy minimization framework for multi-target tracking which includes two optimization strategies such as gradient ascent and gradient decent which is for likelihood matching of target occlusion and for the global optimization of multiple tracking targets respectively. The novel approach has produced comparatively consistent results on all the video sequences. It also considerably decreases the false positive counts, number of mismatches, and the false negative rate which is an essential factor in comparison of the MTT applications. An experimental result indicates that global data association is helpful, especially in reducing the trajectory fragments and improving the trajectory consistency. This framework is general and adaptable in the application of tracking of any class of targets. This state-of-art method is rapid and efficient to meet all the real time requirements such as complex target interaction with missing detections, crowded scenes and long-term occlusions.

REFERENCES

[1] M. Everingham et.al., “The Pascal Visual Object Classes Challenge”, *Proceedings of 1st Pascal Machine Learning Challenges Workshop*, pp. 117-176, 2012.

[2] P. Dollar, C. Wojek, B. Schiele and P. Perona, “Pedestrian Detection: A Benchmark”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 304-311, 2009.

[3] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski, “A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms”, *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 519-528, 2006.

[4] S. Baker, D. Scharstein, J.P. Lewis, S. Roth, M.J. Black, and R. Szeliski. "A Database and Evaluation Methodology for Optical Flow", *International Journal of Computer Vision*, Vol. 92, No. 1, pp. 1-31, 2011.

[5] Andreas Geiger, Philip Lenz and Raquel Urtasun, "Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3354-3361, 2012.

[6] J. Black, T. Ellis and P. Rosin, "Multi View Image Surveillance and Tracking", *Proceedings of Workshop on Motion and Video Computing*, pp.1-6, 2002.

[7] D. Reid. "An Algorithm for Tracking Multiple Targets", *IEEE Transactions on Automatic Control*, Vol. 24, No. 6, pp. 843-854, 1979.

[8] K. Okuma, A. Taleghani, O.D. Freitas, J.J. Little and D.G. Lowe, "A Boosted Particle Filter: Multitarget Detection and Tracking", *Proceedings of 8th European Conference on Computer Vision*, Vol. 1, pp. 28-39, 2004.

[9] J. Vermaak, A. Doucet and P. Perez, "Maintaining Multi-Modality through Mixture Tracking", *Proceedings of 9th Conference on Computer Vision*, pp. 1-7, 2003.

[10] T.E. Fortmann, Y. Bar-Shalom and M. Scheffe, "Multi-Target Tracking using Joint Probabilistic Data Association", *Proceedings of IEEE Conference on Decision and Control including the Symposium on Adaptive Processes*, pp. 807-812, 1980.

[11] Mykhaylo Andriluka, Stefan Roth and Bernt Schiele, "Monocular 3D Pose Estimation and Tracking by Detection", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 623-630, 2010.

[12] A. Andriyenko, S. Roth and K. Schindler, "An Analytical Formulation of Global Occlusion Reasoning for Multi-Target Tracking", *Proceedings of 11th International IEEE Workshop on Visual Surveillance*, pp. 1839-1846, 2011.

[13] Anton Andriyenko and Konrad Schindler, "Globally Optimal Multi-Target Tracking on a Hexagonal Lattice", *Proceedings of 11t European Conference on Computer Vision*, pp. 466-479, 2010.

[14] A. Andriyenko and K. Schindler, "Multi-Target Tracking by Continuous Energy Minimization", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1265-1272, 2011.

[15] B. Benfold and I. Reid, "Stable Multi-Target Tracking in Real-Time Surveillance Video", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2011.

[16] M. Kamaraj and Balakrishnan, "Surveillance of Human Tracking using Gaussian Beta-Likelihood Matching and Kalman Filter", *International Journal of Applied Engineering Research*, Vol. 10, No. 14, pp 34375-34382, 2015

[17] M. Kamaraj and Balakrishnan, "Optimization of Multi-Target Tracking and Occlusion Handling using Mean Shift Method", *International Journal of Advanced Research in Computer science and Software Engineering*, Vol. 5, No. 9, pp. 367-375, 2015.

[18] B. Leibe, K. Schindler and L. Van Gool, "Coupled Detection and Trajectory Estimation for Multi-Object Tracking", *Proceedings of 11t International Conference on Computer Vision*, pp. 2-8, 2007.

[19] Hao Jiang, Sidney Fels and James J. Little, "A Linear Programming Approach for Multiple Object Tracking", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2007.

[20] L. Zhang, Y. Li and R. Nevatia, "Global data association for multiobject tracking using network flows", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2008.

[21] M. Rodriguez, I. Laptev, J. Sivic and J.Y. Audibert, "Density-Aware Person Detection and Tracking in Crowds", *Proceedings of IEEE International Conference on Computer Vision*, pp. 2423-2430, 2011.

[22] Junliang Xing, Haizhou Ai and Shihong Lao, "Multi-Object Tracking through Occlusions by Local Tracklets Filtering and Global Tracklets Association with Detection Responses", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1200-1207, 2009.

[23] C. Wojek, S. Walk, S. Roth and B. Schiele, "Monocular 3D Scene Understanding with Explicit Occlusion Reasoning", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1993-2000, 2011.

[24] M.D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier and L. Van Gool, "Robust Tracking-by-Detection using a Detector Confidence Particle filter", *Proceedings of 12th International Conference on Computer Vision*, pp. 1515-1522, 2009.

[25] A. Milan, S.Roth and K.Schindler, "Continuous Energy Minimization for Multi-Target Tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 1, pp. 58-72, 2014.

[26] L. Kratz and K. Nishino, "Tracking with Local Spatio-Temporal Motion Patterns in Extremely Crowded Scenes", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 693-700, 2010.

[27] W. Choi and S. Savarese, "Multiple Target Tracking in World Coordinate with Single, Minimally Calibrated Camera", *Proceedings on 11th European Conference on Computer Vision*, pp. 553-567, 2010.

[28] Zhen Qin and Christian R. Shelton, "Improving Multi-Target Tracking via Social Grouping", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1972-1978, 2012.

[29] R.T. Collins, "Mean-Shift Blob Tracking through Scale Space", *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 234-240, 2003.

[30] J. Ning, L. Zhang, D. Zhang and C. Wu, "Scale and Orientation Adaptive Mean Shift Tracking", *IET Computer Vision*, Vol. 6, No. 1, pp. 52-61, 2012.

[31] Shou Zhang and Yaakov Bar-Shalom, "Robust Kernel-based Object Tracking with Multiple Kernel Centers", *Proceedings of 12th International Conference on Information Fusion*, pp. 1014-1021, 2009.

[32] Alper Yilmaz, "Object Tracking by Asymmetric Kernel Mean Shift with Automatic Scale and Orientation Selection", *Proceedings of* IEEE *Conference on Computer Vision and Pattern Recognition*, pp. 1-6, 2007.

[33] B.Z. De Villiers, W.A. Clarke and P.E. Robinson, “Mean shift Object Tracking with Occlusion Handling”, Available: http://www.prasa.org/proceedings/2012/prasa2012-36.pdf.

[34] Z. Zivkovic and B. Krose. “An EM-like Algorithm for Color-Histogram-based Object Tracking”, *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1-6, 2004.

[35] Q.A. Nguyen, A. Robles-Kelly and C. Shen, “Kernel-based Tracking from a Probabilistic Viewpoint”, *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1-8, 2007.

[36] I. Guskov, “Kernel-based Template Alignment”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 610-617, 2006.

[37] Cavira Test Case Scenario, Available: at: http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/

[38] J.F. Henriques, R. Caseiro and J. Batista, “Globally Optimal Solution to Multi-Object Tracking with Merged Measurements”, *Proceedings of IEEE Conference on Computer Vision*, pp. 2470-2477, 2011.

[39] J. Berclaz, F. Fleuret, E. Turetken and P. Fua, “Multiple Object Tracking using K-Shortest Paths Optimization”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 9, pp. 1806-1819, 2011.

[40] PETS-Performance Evaluation of Tracking and Surveillance, Available at: http://www.cvg.reading.ac.uk/slides/pets.html

[41] B. Wu and R. Nevatia, “Tracking of Multiple, Partially Occluded Humans based on Static Body Part Detection”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2006.