Computer Systems Science & Engineering
DOI: 10.32604/csse.2023.026742
Article

Tech Science Press

Moving Multi-Object Detection and Tracking Using MRNN and PS-KM Models

V. Premanand* **and Dhananjay Kumar**

Department of Information Technology, Anna University, MIT Campus, Chennai, 600 044, India
*Corresponding Author: V. Premanand. Email: premanandv@mitindia.edu
Received: 03 January 2022; Accepted: 22 February 2022

Abstract: On grounds of the advent of real-time applications, like autonomous driving, visual surveillance, and sports analysis, there is an augmenting focus of attention towards Multiple-Object Tracking (MOT). The tracking-by-detection paradigm, a commonly utilized approach, connects the existing recognition hypotheses to the formerly assessed object trajectories by comparing the similarities of the appearance or the motion between them. For an efficient detection and tracking of the numerous objects in a complex environment, a Pearson Similarity-centred Kuhn-Munkres (PS-KM) algorithm was proposed in the present study. In this light, the input videos were, initially, gathered from the MOT dataset and converted into frames. The background subtraction occurred which filtered the inappropriate data concerning the frames after the frame conversion stage. Then, the extraction of features from the frames was executed. Afterwards, the higher dimensional features were transformed into lower-dimensional features, and feature reduction process was performed with the aid of Information Gain-centred Singular Value Decomposition (IG-SVD). Next, using the Modified Recurrent Neural Network (MRNN) method, classification was executed which identified the categories of the objects additionally. The PS-KM algorithm identified that the recognized objects were tracked. Finally, the experimental outcomes exhibited that numerous targets were precisely tracked by the proposed system with 97% accuracy with a low false positive rate (FPR) of 2.3%. It was also proved that the present techniques viz. RNN, CNN, and KNN, were effective with regard to the existing models.

Keywords: Multi-object detection; object tracking; feature extraction; morlet wavelet mutation (MWM); ant lion optimization (ALO); background subtraction

1 Introduction

As far as the MOT is concerned, it is deemed as the basic issue in computer vision. It is vital for various applications like house-care, home-care, surveillance and security, video communication, traffic monitoring, robot vision and animation, and computer vision [1]. Its main goal is to parallelly determine the trajectories of numerous objects by localising the identical targets across multiple frames [2]. This capability of classifying the dynamic objects and the static objects is referred as the base. Also, it maintains the tracking of the dynamic objects [3]. In order to capture the combinatorial complexity on a frame-by-frame basis, the two

conventional methods viz. Multiple Hypothesis Tracking (MHT) and Joint Probabilistic Data Association Filter (JPDAF) were employed with the aim of establishing sophisticated models [4]. Tracking could be made strenuous by disparate viewing angles, diverse brightness or occlusion levels, along with an unfixed number of targets [5]. Selecting the objects of interest, tracking of detected objects as of frame to frame, along with analysis of object tracks are the primary steps involved in tracking [6].

It is evidential that the MOT has been a benefactor of many applications in object detection (OD) [7]. However, detection-centred tracking methods gained more attention with a considerable progress in OD researches [8]. The data association (DA) process is the key factor of this strategy which is generally considered as '3' separate parts: feature extraction (FE) for candidate representation, affinity metric for assessing the linking probability betwixt candidates or tracklets, and the association algorithm for finding the optimum association [9]. By this means, tracking of failures could be recovered by the technique of determining the object hypothesis similar to the detection techniques [10]. In that, the offline and the online methods are their categorizations [11]. Several trajectories were formed using the offline tracking methods by optimizing the detections in a complete video or a huge sliding window [12].

Formerly, the tracklets were produced by connecting individual detections in several frames, and then, they were iteratively linked to create long trajectories of objects in the whole sequence or in a time-sliding window with a temporary delay [13]. As a setback, owing to the poor adaptiveness to the dynamic variations, there are high chances for the occurrence of failure of tracking to the off-line trackers. Therefore, online approaches are utilized for fulfilling the runtime specification necessity of arbitrary tracking that links existent tracks with detecting frame by frame, possessing higher practical utility [14,15]. Recognizing every tracklet (i.e., linked detections in short frames) at frame is considered to be a vital process as the trajectories are constructed based on both the techniques by utilizing global or local associations [16]. Since the track fragments, along with the identity switches, can be recovered by matching the tracks with the equivalent detections, the tracking accuracy can also be enhanced [17].

The DA issue is addressed by tracking-by detection approach via optimizing the detection assignments over a bigger temporal window [18]. Nevertheless, these models do not possess the ability to study the long-term information of the history and alter the estimation as the templates are restricted in a fixed time window [19]. Besides, the OD might be kept unreachable to the temporal information of a specific object by separating the detection exclusively based on the tracking [20]. And, for the computer vision methods, creating an effective and a robust system to estimate the human vision mechanisms has become quite essential. Currently, promising improvements have been made to the deep learning algorithms [21]. A Deep Learning (DL) MRNN model and a PS-KM algorithm, which is considered as a moving multi-OD and tracking mechanisms were used in the present study. The objects are efficiently tracked and identified via the proposed framework, and also, the targets were differentiated with regard to the background. By identifying and tracking the objects with greater accuracy and lower false alarm rates, the proposed framework was executed efficiently.

This paper has been systematized into 5 sections. The Section 1 introduces the MOT as the basic issue in computer vision and the Section 2 mentions the various related literature surveyed to substantiate the proposed framework. Then, the Section 3 proffers the efficacies of the present work, and the results, along with the discussions, are given in Section 4. Finally, the Section 5 elucidates the conclusions that were drawn and expounds the possibilities for any related works in the future.

2 Literature Review

A multiple modal tracking by merging a distance-centred tracker with an appearance-based tracking technique [22] is regarded as the suitable method for the trajectory creation of numerous objects with complicated random motion structure. To introduce the structural context information in the tracking-by-

detection framework, a proximity estimation scheme was employed. For incorporating the target's spatial-temporal information, choosing important features, and establishing the statistical correlation betwixt a target's prior model and its current observation, the numerous-instance framework was specified. The tracking performance was considerably enhanced by the present approach by decreasing the total fragmented trajectories, along with the ID switches. It is not appropriate to track the objects captured through the multi-camera, because, a target-specific appearance model was incorporated in the framework.

A robust MOT algorithm focuses on the amalgamation of independent features like colour histogram (CH) model, sparse appearance model, optical flow histogram, and spatial model [23]. The integration of feature descriptors into a DA method was done where in a way that every target was matched with every candidate under local geometric limitations, and target states that manage the target's data like occlusion, birth, and death. A hierarchical DA process was presented by this method where every target was split into occluded along with unoccluded targets for handling the occlusion issue. Competitive results were attained by the proffered MOT framework and it had the capability of managing numerous difficult issues. The CH illustration was reliant upon the object's color, neglecting its shape along with texture, which was considered as a disadvantage.

A tracking technique for detecting and tracking the objects in motion or in long-lasting occlusion by merging information as of enlarged structural along with temporal domains [24]. Concentrating the meta-measurements of object affinity, the detections were initially gathered as a small tracklets in this approach. Grounded upon a motion pattern, the association task for tracklets-to-tracks were resolved by structural information. Moreover, for recovering objects, restrictions of the temporal domain were presented that were disappearing because of failed detection or long-term occlusion. By gathering the heterogeneous domain information, enhanced top-notch performance was exhibited by this technique on standard benchmarks with comparatively low processing time. However, this method did not deem the optical flow, pose information, and deep features as the shortcomings.

A robust MOT method is grounded on an affinity model for DA, along with a trajectory assessment strategy for managing the detector's defect [25]. In this model, for extracting appearance, a CNN was model was designed. For extracting motion cues to encode the target's dynamics, a long short-term memory network (LSTM) was employed. An end-to-end deep metric learning was executed by a triplet loss function which was merged with both the cues. Thus, to distinguish the targets, fused features were produced. The LSTM network's hidden state was taken as the input by an RNN-centred Bayesian filtering module, and it executed recursive prediction, along with updating the openly assessing target state. This method had executed efficiently in identifying the objects in the occluded environment as revealed by the experimental outcomes. A more fragile and efficient optimization strategy would be useful for further improving the tracking performance as a simple linear program algorithm was employed for the association.

The Markov decision process for integrating the discriminative correlation filters (DCFB) tracking method into the MOT framework was proposed [26]. The ‘2’ DCFB trackers were used with disparate update frequencies. For predicting the target's location, an updated strategy was utilized. For extracting robust features to tackle the problems of occlusion and scale change, the part-centred method was applied. The results exhibited that top-notch performance was attained by this method and surpassed the state-of-the-art algorithms in road scenarios. Nevertheless, the background was studied by the filter, which might generate drift with failure.

Multi-Object Detection along with Tracking (MODT) aimed at real-time video surveillance systems was presented by Elhoseny [27]. An optimum Kalman filtering (KF) was employed in this technique for tracking the moving objects in Video Frames (VF). Depending on the total frames, the conversion of video clips into morphological operations was done using the region-growing model. Applying the probability-centred grasshopper algorithm, KF was employed for parameter optimization after differentiating the objects. The

chosen object's tracking was done in every frame via a similarity measure with the help of optimum parameters. The experiments exhibited that maximum detection, along with tracking accuracies of 76.23%, and 86.78% correspondingly, was attained by the MODT. The minimum error rate and similarity were obtained when contrasted to the prevailing techniques. The detection rate was relatively low, which was the major drawback of the submitted work.

3 Proposed Methodology

MOT is closely associated with video Object Detection (OD) and target re-identification. It is a potential tool with an arduous part in the intelligent transportation systems with computer vision applications. Recently, the most top-notch method in OD, along with tracking, are based on deep neural networks with the representation power brought by deep learning. Nevertheless, ameliorating the performance of MOT in complicated scenes still remains as an open issue. Exploiting a DL model along with PS-KM, moving multi-OD and tracking mechanisms are proposed in this paper. An object's occlusion, shadows, along with camera jitter could be managed by the proposed work. From the publicly accessible dataset such as MOT15, MOT17, and MOT20 the input data (videos) are initially gathered and converted into VF. Next, utilizing Entropy-like Divergence-based Kernel K-Means Algorithm (EDK^{2}MA), the background subtraction of the frame is performed. Then, from the background-subtracted frames, the features are extracted. Next, for transforming the higher dimensional features into lower-dimensional features, the features are decreased utilizing IG-SVD. After that, for OD and classification, MRNN is utilized. Therefore, the detection of disparate classes of objects is contained by the classifier's output which will further be offered to the object tracking mechanism. Utilizing the PS-KM algorithm, the identified objects are then tracked. Fig. 1 indicates the proposed system's architecture.

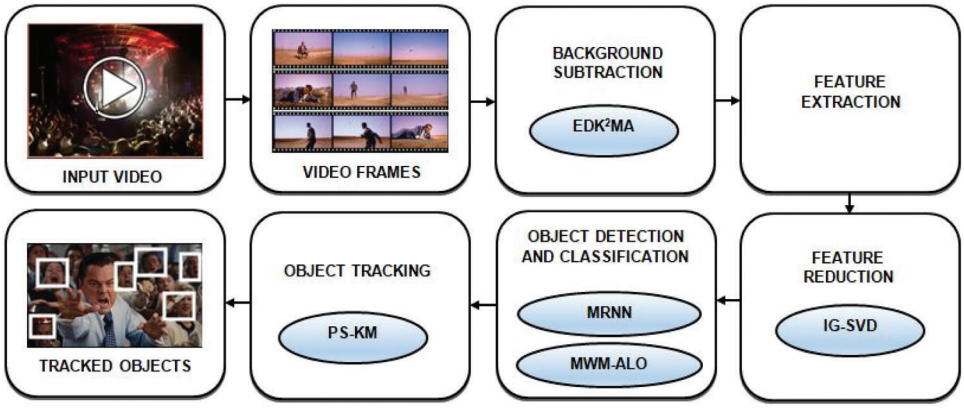


Figure 1: Proposed system architecture

3.1 Background Subtraction

From the MOT dataset, the input videos (V_i) are initially gathered. Next, the amassed data is converted into VF (V_f) which are presented by,

$$V_f = \{V_1,\ V_2,\ \ldots,\ V_N\} \tag{1}$$

wherein, N signifies the number of VF. After that, utilizing an EDK^{2}MA, the background subtraction of the VF is executed. The irrelevant information is filtered by the EDK^{2}MA via partitioning the frames into k clusters and the clusters are grouped based on identical features and also, unrelated features are eliminated. The Gaussian Kernel in cluster formation was utilized by the general K-Means Algorithm

(KMA) that was specified by the Euclidean distance (ED). For complex non-convex data, ED was challenging to attain the adequate results and it was receptive to outliers or noise. By merging Jenson-Shannon/Bregman divergence with convex function, along with its mercer kernel function called entropy-like divergence-based kernel (EDK), entropy-like divergence was induced into conventional KMA for overcoming this problem. Thus, the clustering process's segmentation accuracy will be enhanced and stronger anti-noise robustness is also possessed by the algorithm. This amalgamation of EDK with KMA is called EDK^{2}MA. The steps incorporated in EDK^{2}MA are as follows.

1. Let the input be the number of VF(V_f), and the cluster centroids C_i is initialized wherein, $i = 1, 2, \ldots, k$ signifies the number of cluster centroids.
2. The number of clusters k with initial cluster centroids C_i is selected.
3. Utilizing the EDK function, assign every data point (V_f) to their closest centroid C_i, which is specified by,

$$E(V_f,\ C_i) = \sum_{f=1,i=1}^{N,k} \exp\left(-\frac{d_e(V_f,\ C_i)}{2\sigma_e^2}\right) \tag{2}$$

$$d_e(V_f,\ C_i) = \sum_{f,i=1}^{N,k} \frac{V_f \ln V_f + C_i \ln C_i}{2} - \frac{V_f + C_i}{2} \ln \frac{V_f + C_i}{2} \tag{3}$$

where, σ_e^2 implies the scale parameter of the EDK function, $d_e(V_f, C_i)$ signifies the entropy-like divergence.

4. By computing the cluster centroids based on their group associates, the steps are repeated by rearranging the data points focus on the new cluster centroids. Until every cluster's centroid comes together, the process is stopped. Hence, V_c signifies the clustered VF where $c = 1, 2, \ldots, k$ which signifies the number of clusters created.

3.2 Feature Extraction

From the clustered VF's foreground images, the extraction of '3' disparate sorts of features was executed in this phase. From the foreground objects, textural features, shape features, along with Haar features, were the features that were extracted. For textural FE, the GLCM features were employed which comprised energy, homogeneity, entropy, contrast, and angular second-order moment. Eccentricity and solidity, along with circularity features, were utilized for shape FE. Lastly, the object's edges and corners were extracted by the Haar features in the foreground image. The extracted features (E_f) are represented as follows.

$$E_f = \{E_1,\ E_2,\ \ldots,\ E_N\} \tag{4}$$

where, N signifies the number of extracted features as of the input VF. Next, for feature reduction, these features are inputted to the IG-SVD.

3.3 Feature Reduction Using IG-SVD

Here, the extracted features (E_f) are offered into the IG-SVD algorithm, where, the higher-dimensional features are transformed into lower-dimensional features. An entropy-centred feature assessment method is named as Information Gain (IG) which looks at every feature individually, assesses its information gain, and estimates the significant class label. A score of 1–0 was attained by each extracted feature which indicated the most relevant to the least relevant. Thus, the evaluation of IG for every extracted feature was done, and also, the features with higher information was chosen and input to SVD. The decrease of entropy was archived by a joint feature set E_f is the IG for a joint set of features, where, $\langle f = 1, 2, \ldots, N \rangle$. It is expressed as,

$$\delta(E_f) = H(D) - \sum_f \frac{D_f}{D} H(D_f) \tag{5}$$

where, $H(D)$ implies the input dataset's entropy, $H(D_f)$ indicates the $f-th$ subset's entropy produced by partitioning D on every feature in the joint feature set E_f. Next, the IG-centred feature's result was offered to SVD. One among the effective data analysis tools in linear algebra is called SVD which decreases the image features into smaller invertible accompanying the square matrices. It also possesses good energy compaction properties. Reducing the data's dimensions and also clearly separating the sample classes, if possible is the SVD's major objective within the analysis.

3.4 Object Detection and Classification

The objects were detected and categorized utilizing MRNN after feature reduction. A sort of artificial neural network (ANN) which utilized the sequential data, and also, the preceding output for the next output's prediction known as RNN. A memory was possessed by these networks that record the information observed by them. Hence, the neural network's generalization performance along with training stability is heavily reliant upon the activation function (AF)'s choice. Sigmoid functions could suffer severely from gradient diffusion issues regardless of their popularity with extensive acceptance in RNNs. It happened mostly on account of the saturation issue of an AF, wherein the changes on the neurons' outputs could be hardly perceived for any inputs near or in a saturation region. Thus, an AF was incorporated in the RNN, which is the bounding approach for overcoming this disadvantage of traditional RNN. A new AF was specified for supporting the thought of possessing a bounded output range understanding the significance of bounding the AF's output for superior training stability. This amalgamation of a new AF in the common RNN is called MRNN. The MRNN's process is detailed as follows.

Step 1: Let, the RNN's input be the reduced features (R_f), where, $f = 1, 2, \ldots, n$ explains the number of features. At every time step (τ), the correlation amongst the neural network's input, and also, the output could be denoted below,

$$H_\tau = \chi(w_{iH} R_f + w_{HH} H_{\tau-1} + B_H) \tag{6}$$

$$O_\tau = \delta(w_{Ho} H_\tau + B_o) \tag{7}$$

wherein, $\tau = 1, 2, \ldots, m$, O_τ illustrates the network's output at a time step τ, H_τ indicates the hidden layer (HL), B_H *and* B_o implies the bias of HL along with output layer, w_{iH} *and* w_{io} implies the weight matrix amongst the input-HL and hidden-output layer correspondingly, w_{HH} signifies the recurrent weight matrix betwixt the last hidden state $\tau - 1$ and the current hidden state τ, χ *and* δ indicates the bounded ReLU AF of the HL along with output layer.

Step 2: The bounded ReLU AF of the HL along with output layer χ, δ is exhibited in Eq. (8)

$$\chi_{R_f} = \delta_{R_f} = \min(\max(0,\ R_f),\ Z) = \begin{cases} 0 & R_f \le 0 \\ R_f & 0 < R_f \le Z \\ Z & R_f > Z \end{cases} \tag{8}$$

where, Z signifies the maximum output value the function could produce. Utilizing Morlet Wavelet Mutation-centered Ant Lion Optimization (MWM-ALO), the weight values within the RNN are optimized.

Weight optimization using MWM-ALO

The optimization algorithm used for optimizing the RNN's weight values is named Ant Lion Optimization (ALO). The antlion's hunting behavior was imitated by this algorithm in which a trap was a dug for gathering insects, mainly ants. A cone-shaped dump was dug in the sands by the antlion and they

hide under the trap and wait for catching their prey. The general ALO algorithm is susceptible to stagnate in a local optimum. It needs a different exploration with a suitable blending of exploitation. Thus, the Morlet Wavelet Mutation (MWM) mechanism was integrated into conventional ALO for overcoming such limitations to understand the mutation space's dynamic adjustment. It enhances the algorithm's capability of escaping as of local optimization and also ameliorates the algorithm's convergence speed and also accuracy. This MWM combined ALO is the so-called MWM-ALO. The algorithm's mathematical model is communicated further,

Step 1: The hunting technique begins with the Random Walk (RW) of ants along with ant lions. Utilizing the MWM, the random walking of ants $w(m)$ within the search space when looking for food is signified as,

$$w(m) = [0,\ \Pi(2p(m_1) - 1),\ \ \Pi(2p(m_2) - 1),\ \ldots,\ \Pi(2p(m_j) - 1)] \tag{9}$$

where, $\prod(y)$ signifies the MWM $\prod(y) = e^{-\frac{y^2}{2}}.\ \cos(5y)$, j indicates the number of iterations, m denotes the step of RW, $p(m)$ implies the stochastic function, and is estimated by

$$p(m) = \begin{cases} 1 & if\ x > 0.5 \\ 0 & if\ x \leq 0.5 \end{cases} \tag{10}$$

where, x implies an arbitrary number produced with uniform distribution in the interval [0, 1].

Step 2: At every step of optimization, the updation of RWs of ants is done as a boundary is possessed by every search space. Therefore, the RWs are normalized utilizing the following min-max normalization for keeping them inside the search space,

$$w_i(m) = \frac{(w_i(m) - x_{\min}(i)) \times (z_m(i) - y_m(i))}{x_{\max}(i) - x_{\min}(i)} + y_m(i) \tag{11}$$

where, $x_{\min}(i)$ implies the minimum of RW in i^{th} variable, $x_{\max}(i)$ indicates the maximum of RW in i^{th} variable, $y_m(i)$ denotes the minimum of i^{th} variable at $m-th$ iteration, $z_m(i)$ signifies the maximum of i^{th} variable at $m-th$ iteration.

Step 3: The ant lion's traps affect the ants' RW which could be formulated in Eq. (12).

$$y_m(i) = AL_m(k) + y_m \tag{12}$$

$$z_m(i) = AL_m(k) + z_m \tag{13}$$

where, y, z implies the range of values that illustrates the minimum and maximum of every variable, $AL_m(k)$ indicates the position of $k-th$ ant lion at $m-th$ iteration.

Step 4: Next, for modelling the hunting ability of antlions, the roulette wheel was utilized. Ants were presumed to be trapped in only one chosen antlion. The antlion was selected via the roulette wheel operator as per its fitness value throughout the optimization. The higher probability of catching ants was signified by the antlion's higher fitness values.

Step 5: The sand was thrown outwards by the antlion when the ants started to fall into the trap, and it slides the ant towards them. Therefore, the ants' RW reduced in radius which is mathematically signified as:

$$y_m = \frac{y_m}{10^c \frac{m}{M}} \tag{14}$$

$$z_m = \frac{z_m}{10^c \frac{m}{M}} \tag{15}$$

where, M implies the maximum iteration, m signifies the existing iteration, c implies the constant defined based upon the existing iteration for adjusting the ant's speed.

Step 6: Lastly, the ant was caught by the antlion and consumed the body when the ant turned fitter when contrasted to the antlion. For augmenting its possibility aimed at a new hunt, the antlion's position was updated to the hunted ant's position. It could be represented in Eq. (16).

$$AL_m(k) = Ant_m(i) \quad if\ f(Ant_m(i)) > f(AL_m(k)) \tag{16}$$

where, $f(°)$ indicates the fitness value, $Ant_m(i)$ signifies the position of i^{th} ant at $m-th$ iteration.

Step 7: A feature of an evolutionary algorithm that permits the best solution attained to be kept throughout the optimization method is called elitism. The best antlion attained at each step is saved along with deemed as elite in the ALO. The elite simulation is presented as follows.

$$Ant_m(i) = \frac{E_{mA} + E_{mE}}{2} \tag{17}$$

where, E_{mA} explains the RW around the chosen ant lion at $m-th$ iteration, E_{mE} implies the RW around the elite at $m-th$ iteration. The optimization of RNN's weight values is performed in this way. It is utilized for OD and classification. Thus, the detected disparate classes of objects are encompassed by the MRNN classifier's output which will further be offered to the object tracking mechanism.

3.5 Object Tracking Using PS-KM

From the DL model, the detected objects were treated as the target that is primarily in the first frame. Utilizing the PS-KM algorithm, the target was tracked in the upcoming frame. For tracking, the objects recognized by the DL neural network in the next frames were utilized. The target along with its new position is signified by the maximum correlation output value. The detected objects were tracked by the Kuhn-Munkres (KM) algorithm via extracting the appearance, shape, along with the motion of the identified objects as of the categorization results, which were jointly optimized for detection and also association. Different similarity measures namely velocity, scale, along with appearance, were incorporated in the creation of the similarity matrix betwixt trajectories and targets. The Pearson Similarity (PS) metric was the similarity model used in the present experiment. For evaluating the target's motion, Kalman filtering was applied. In the OD process, the appearance features along with shape features were extracted concurrently. For object tracking, the PS-centred KM algorithm is known as PS-KM. The steps implied in the PS-KM algorithm is detailed as follows.

Step 1: Let, the identified classes of objects be, O_j which denotes the MRNN features, the j^{th} trajectory as $T(j)$ along with the j^{th} detection as $D(j)$. Utilizing the below Eq. (18), the Pearson Similarity (PS) metric (σ_{sim}) concerning appearance is computed by,

$$\sigma_{sim}(D(j),\ T(j)) = \frac{\sum_{j=1}^{m} (O_{T(j)} - \hat{O}_{T(j)}) \circ (O_{D(j)} - \hat{O}_{D(j)})}{\sqrt{\sum_{j=1}^{m} [(O_{T(j)} - \hat{O}_{T(j)})^2 \circ (O_{D(j)} - \hat{O}_{D(j)})^2]}} \tag{18}$$

where, $\hat{O}_{T(j)},\ \hat{O}_{D(j)}$ signifies the object's mean values, m denotes the number of detected objects.

Step 2: The motion similarity metric (σ_{motion}) is specified as,

$$\sigma_{motion}(D(j),\ T(j)) = \sum_{j=1}^{m} \left(-W_1 \left[\frac{(P_{T(j)} - P_{D(j)}) \circ (Q_{T(j)} - Q_{D(j)})}{\sqrt{(W_{D(j)} - W_{t(j)})^2 \circ (H_{D(j)} - H_{T(j)})^2}} \right] \right) \tag{19}$$

wherein, W and H indicates the width along with the height of the bounding box (BB), P and Q illustrates the horizontal along with vertical coordinates of the BB center correspondingly.

Step 3: Utilizing the below Eq. (20), the shape similarity metric (σ_{shape}) is assessed.

$$\sigma_{shape}(D(j),\ T(j)) = \sum_{j=1}^{m} \left(-W_2 \left[\frac{(H_{T(j)} - H_{D(j)}) \circ (W_{T(j)} - W_{D(j)})}{\sqrt{(H_{T(j)} - H_{D(j)})^2 \circ (W_{T(j)} - W_{D(j)})^2}} \right] \right) \tag{20}$$

Step 4: The overall similarity metric (σ) amongst trajectories along with target is estimated as,

$$\sigma(D(j),\ T(j)) = \sigma_{sim}(D(j),\ T(j)) * \sigma_{motion}(D(j),\ T(j)) * \sigma_{shape}(D(j),\ T(j)) \tag{21}$$

Therefore, from the series of images, the trajectories are produced. In a complex environment, numerous objects could be identified along with tracked in this way. Algorithm 1 exhibits the pseudocode of the proposed PS-KM algorithm.

Algorithm 1: Pseudocode of MWM-ALO

Weight optimization using MWM-ALO

Input: Weight w

Output: Optimized weight W_o

Begin

Initialize *Ant* and *AL*

Compute the fitness value

Find the best *AL* and assume it as elite

While the end criterion is not satisfied

  For every *Ant*

    Select *AL* using Roulette wheel

    Update y_m and z_m

    Generate $w(m)$ and normalize it using $w_i(m)$

    Update $Ant_m(i) = \frac{E_{mA}+E_{mE}}{2}$

  End for

Determine $f(Ant_m(i))$

Replace $AL_m(k) \leftarrow Ant_m(i)$; *if* $f(Ant_m(i)) > f(AL_m(k))$

Update elite if *AL* is fitter than elite

End while

Return elite

End

---

4 Results and Discussions

For authenticating the method's effectiveness, the results of the present study is specified. Based on their performance metrics, the performance analysis of the proposed EDK^{2}MA for background extraction, MRNN for OD and classification, MWM-ALO for weight optimization, and PC-KM for object tracking were contrasted. From the MOT dataset, the input VF was taken. Fig. 2b exhibits the tracked objects.

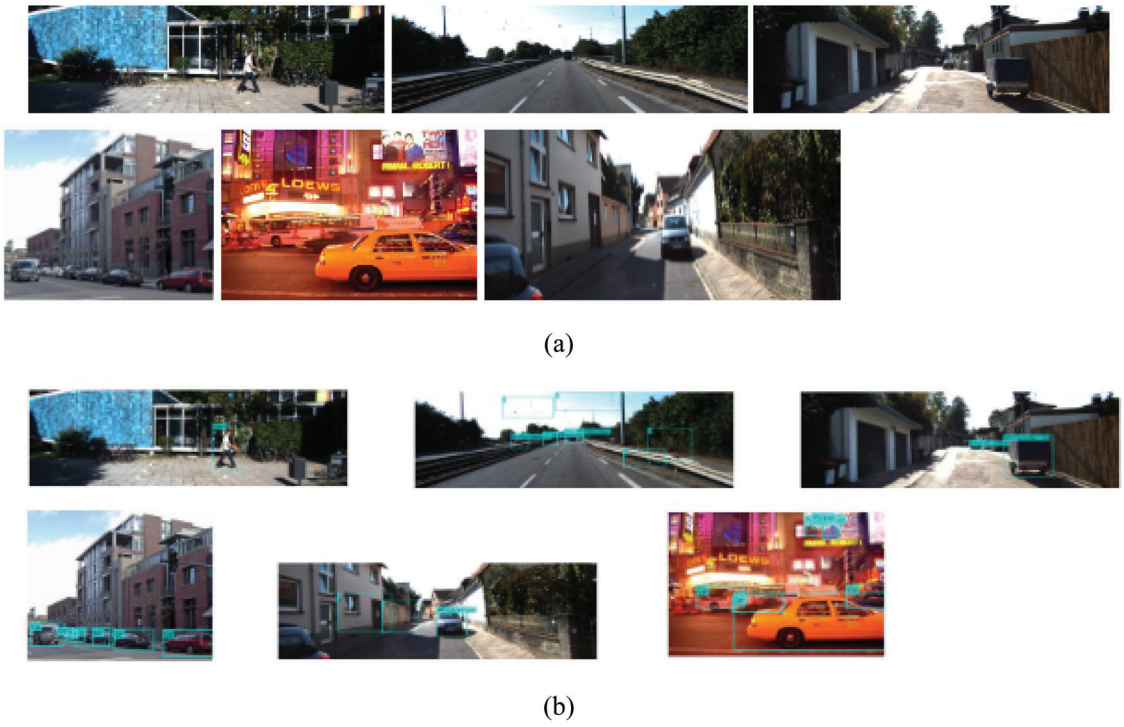


Figure 2: Input video frames and tracked objects from the dataset (a) Input frames (b) Tracked objects

4.1 Performance Analysis of EDK^{2}MA

For the background subtraction, the proposed EDK^{2}MA was authenticated by contrasting it with the existent KMA, K-Medoids, and Fuzzy C-Means (FCM) focused on precision, recall, f-measure, and accuracy. Tab. 1 depicts the performance assessment.

Table 1: Performance evaluation of proposed EDK^{2}MA with existing methods based on precision, recall, F-measure and accuracy

Evaluation metrics	Proposed EDK^{2}MA	KMA	K-Medoids	FCM
Precision	97.3245	96.7902	94.1028	93.1019
Recall	96.4061	95.1254	94.6902	93.5345
F-measure	96.7609	94.9876	93.5678	92.8643
Accuracy	98.1267	97.5609	95.2390	94.1985

In terms of the statistical measures, Tab. 1 exhibits the proposed technique with the existent method's performance. Higher precision of 97.3245% was attained by the proposed EDK^{2}MA. Thus, it is apparent

that when contrasted to the existent techniques, the proposed one executes better. It is complex to attain adequate results for complex non-convex data since the existent methods were receptive to outliers or noise. The aforementioned drawback was overcome by the adoption of Entropy-like Divergence by merging Jenson-Shannon or Bregman divergence with convex function in KMA and superior results were also attained.

4.2 Performance Assessment of MRNN

Grounded on the accuracy, FPR, False Negative Rate (FNR), and precision, the proposed MRNN's performance was contrasted with the existent classifiers. RNN, CNN, and KNN were the techniques utilized for comparison. Fig. 3 demonstrates the performance comparison concerning FPR, FNR, precision, and accuracy.

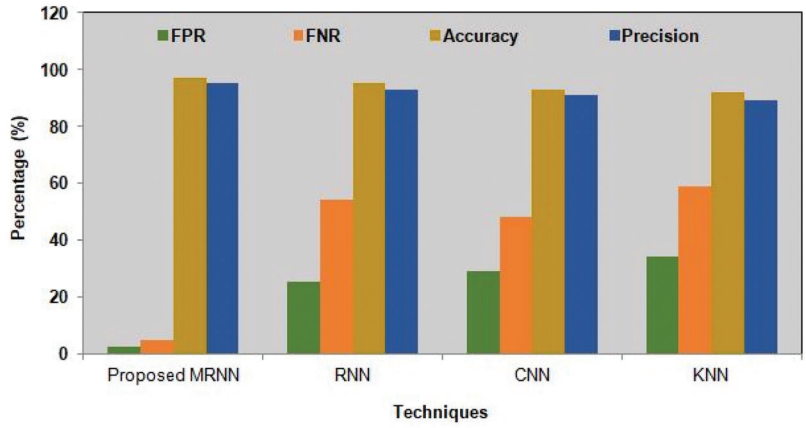


Figure 3: Performance comparison of proposed MRNN with the existing classifiers based on FPR, FNR, precision and accuracy

Concerning FPR, FNR, precision and accuracy, Fig. 3 exemplifies the performance comparison of the proposed MRNN classifier. It is obvious from Fig. 3 that lower FPR and FNR of 2.3%, 4.5%, correspondingly are attained by the proposed work. Hence, higher FPR and FNR values were found to be possessed by the prevailing methods when contrasted to the proposed one. Thus, higher accuracy of 97%, and also, higher precision of 95% were attained by the proposed work. It was on account of the conventional RNN's improvement by utilizing a new activation function rather than sigmoid functions. The vanishing gradient issue is possessed by other existent classifiers that affect the detection performance. Thus, when contrasted to the prevailing classifiers, the proposed MRNN classifier for OD and classification was well executed.

4.3 Performance Analysis of PS-KM

As per their performance metrics, like Mostly Tracked Object (MTO), Mostly Loss Targets (MLT), Number of track fragments (FRAG), ID Switch (IDS), Run time speed, MOT Accuracy (MOTA), and MOT Precision (MOTP), the performance of the proposed PS-KM was compared with the conventional methods namely Markov Decision Process (MDP), Relative Motion Network Object Tracking (RMOT), and Gaussian Mixture Probability Hypothesis Density Filter (GMPHD). Fig. 4 depicts the PS-KM with the existent methods' performance comparison concerning MTO, MLT.

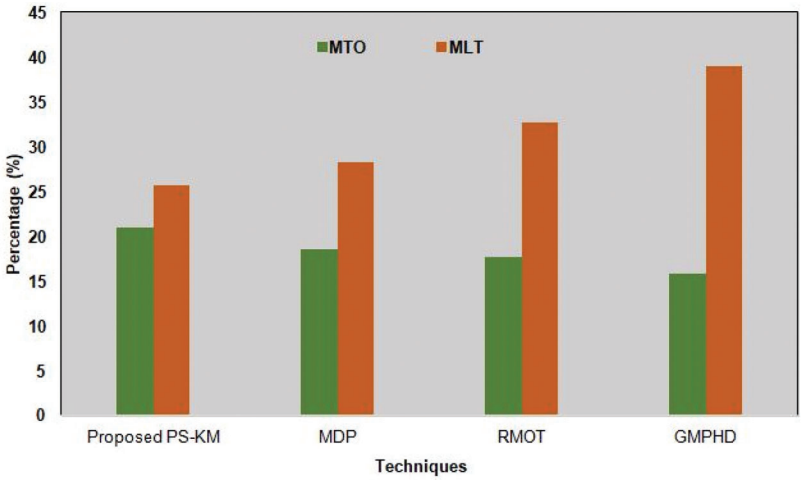
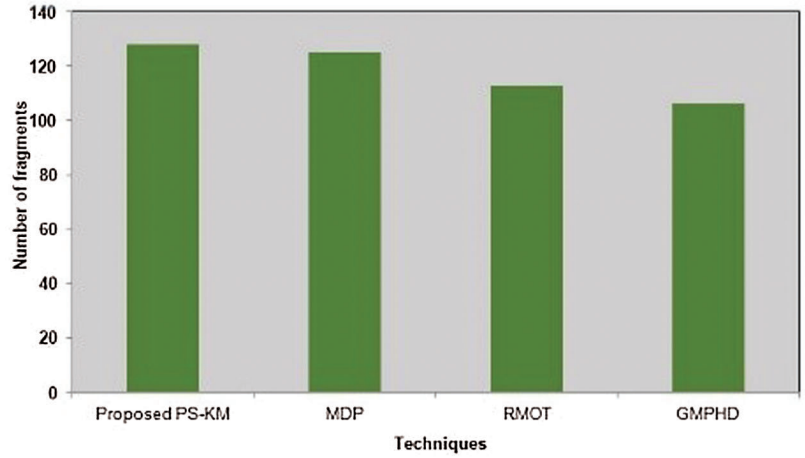


Figure 4: Performance assessment of proposed PS-KM based on MTO and MLT

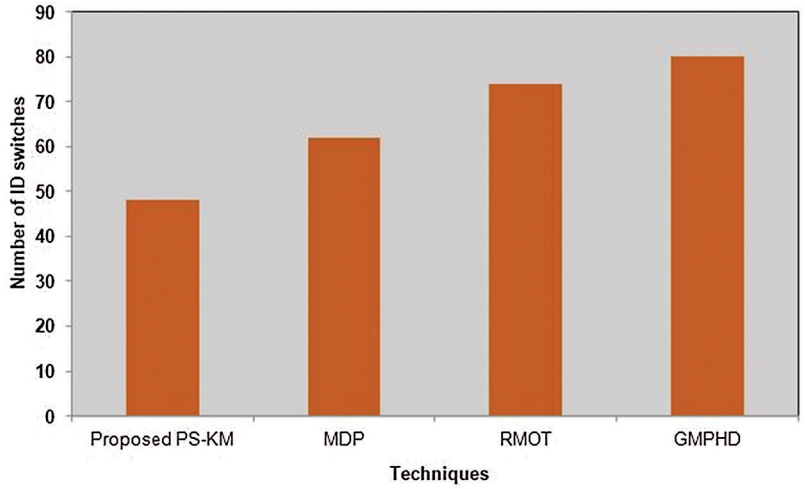
It is apparent from Fig. 4 that a higher MTO of 21% and a lower MLT of 25.6% was attained by the proposed model which has several methods failed in tracking the target for an entire sequence of frames. Thus, less MTO and high MLT were depicted by the existing techniques. From this, it is evident that the target was more accurately tracked via the proposed work with minimum loss when contrasted to the prevailing methods. Fig. 5 represents the performance assessment of the proposed PS-KM concerning IDS and FRAG.

The performance analysis of the proposed PS-KM with the conventional methods is exemplified in Fig. 5. Hence, when weighted against the proposed PS-KM, the values were found to be higher. In contrast, the FRAG aimed at the proposed PS-KM was 128, whilst the value obtained in the present study was found to be lower than the existent methods. Thus, the objects can be better tracked by the proposed model when analogized to the prevailing techniques. Based on RTS, MOTA, and MOTP, Tab. 2 presents the performance assessment of the proposed PS-KM.

The performance measure of the proposed PS-KM algorithm was compared with the existent MDP, RMOT, and GMPHD methods. When weighted against the existing techniques, from the Tab. 2, it can be comprehended that the proposed method had superior MOTA and MOTP. For complex templates, the existent object tracking methods were not found to be appropriate. The detection rate was reduced when the object left the frame or occlusion was present in the object. Thus, the RTS of 9.6 s was depicted by the proposed PS-KM for every frame. For MOTA and MOTP, 86.34% and 84.32% were possessed by the proposed work, which is extremely higher when contrasted to the conventional methods. It is apparent from this evaluation that the objects were well-identified along with tracked by the proposed technique when weighted against other prevailing techniques.



(a)



(b)

Figure 5: Performance evaluation of PS-KM by means of (a) FRAG and (b) IDS

Table 2: Performance estimation of proposed PS-KM with the existing methods based on RTS, MOTA and MOTP

Performance metrics	Proposed PS-KM	MDP	RMOT	GMPHD
RTS (s)	9.6	8.8	8.3	7.4
MOTA (%)	86.34	82.12	78.98	72.15
MOTP (%)	84.32	79.24	73.54	69.78

5 Conclusion

For a perception system, the visual tracking of numerous objects is a vital component in autonomous driving vehicles. The persistent problems for object tracking are managing motion noise, along with ambiguities betwixt long-ranged objects. There are numerous challenges even though a few existent approaches were examined in this area. Thus, an effective technique for MOT in complex environment is

proposed employing MRNN, along with PS-KM algorithms, in the present study. For examining the proposed work's efficiency, the proposed work's performance was contrasted with the existent techniques. MOTA, MOTP, RTS, IDS, Accuracy, Precision, Recall, FRAG, MTO, MLT, FPR, FNR, and F-measure were the evaluation metrics used for the analysis. Higher MOTA (86.34%), MOTP (84.32%), RTS (9.6 s), FRAG (128), MTO (21%) along with lower MLT (25.65%), IDS (48) was achieved by the proposed work. It is apparent that efficient outcomes were attained by the proposed model by deeming these metrics and obtained efficient detection along with tracking of objects in a complex environment. For tracking occluded objects in a complicated environment, the proposed work will be prolonged by utilizing enriched tracking methods in the upcoming future.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study

References

[1] N. T. L. Anh, F. M. Khan, F. Negin and F. Bremond, "Multi-object tracking using multi-channel part appearance representation," in *2017 14th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance (AVSS)*, Lecce, Italy, pp. 1–6, 2017.

[2] Z. Zhou, W. Luo, Q. Wang, J. Xing and W. Hu, "Distractor-aware discrimination learning for online multiple object tracking," *Pattern Recognition*, vol. 107, pp. 1–9, 2020.

[3] A. Kampker, M. Sefati, A. S. A. Rachman, K. Kreiskother and P. Campoy, "Towards multi-object detection and tracking in urban scenario under uncertainties," in *Proc. of the 4th Int. Conf. on Vehicle Technology and Intelligent Transport Systems*, Funchal, Madeira-Portugal, pp. 156–167, 2018.

[4] X. Wan, J. Wang, Z. Kong, Q. Zhao and S. Deng, "Multi-object tracking using online metric learning with long short-term memory," in *2018 25th IEEE Int. Conf. on Image Processing (ICIP)*, Athens, Greece, pp. 788–792, 2018.

[5] B. Jiang and C. Lee, "Online layered multiple object tracking using residual-residual networks," in *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conf. (APSIPA ASC)*, Lanzhou, China, pp. 383–390, 2019.

[6] A. R. Shravya, K. M. Monika, V. Malagi and R. Krishnan, "A comprehensive survey on multi object tracking under occlusion in aerial image sequences," in *2019 1st Int. Conf. on Advanced Technologies in Intelligent Control, Environment, Computing & Communication Engineering (ICATIECE)*, Banglore, India, pp. 225–230, 2019.

[7] L. Chen, H. Ai, Z. Zhuang and C. Shang, "Real-time multiple people tracking with deeply learned candidate selection and person re-identification," in *2018 IEEE Int. Conf. on Multimedia and Expo (ICME)*, San Diego, CA, USA, pp. 1–6, 2018.

[8] J. Xing, H. Ai and S. Lao, "Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses," in *2009 IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, FL, pp. 1200–1207, 2009.

[9] Q. Zhao, L. Wang and Z. Zhou, "Multiple object tracking based on the deep neural networks and correlation filter," in *2019 IEEE Int. Conf. on Robotics and Biomimetics (ROBIO)*, Dali, China, pp. 2989–2994, 2019.

[10] S. Bae and K. Yoon, "Confidence-based data association and discriminative deep appearance learning for robust online multi-object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp. 595–610, 2018.

[11] H. Zhou, W. Ouyang, J. Cheng, X. Wang and H. Li, "Deep continuous conditional random fields with asymmetric inter-object constraints for online multi-object tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 4, pp. 1011–1022, 2019.

[12] J. Ju, D. Kim, B. Ku, D. K. Han and H. Ko, "Online multi-object tracking with efficient track drift and fragmentation handling," *Journal of the Optical Society of America A*, vol. 34, no. 2, pp. 280–293, 2017.

[13] J. H. Yoon, C. R. Lee, M. H. Yang and K. J. Yoon, "Structural constraint data association for online multi-object tracking," *International Journal of Computer Vision*, vol. 127, no. 1, pp. 1–21, 2019.

[14] P. Zhang, T. Zhuo, W. Huang, K. Chen and M. Kankanhalli, "Online object tracking based on CNN with spatial-temporal saliency guided sampling," *Neurocomputing*, vol. 257, pp. 115–127, 2017.

[15] Z. Lin, H. Zheng, B. Ke and L. Chen, "Online multi-object tracking based on hierarchical association and sparse representation," in *2017 IEEE Int. Conf. on Image Processing (ICIP)*, Beijing, China, pp. 655–659, 2017.

[16] S. Lee, M. Kim and S. Bae, "Learning discriminative appearance models for online multi-object tracking with appearance discriminability measures," *IEEE Access*, vol. 6, pp. 67316–67328, 2018.

[17] S. Bae, "Online multi-object tracking with visual and radar features," *IEEE Access*, vol. 8, pp. 90324–90339, 2020.

[18] Z. Wang, B. Xu and F. Huang, "Real-time multiple object tracking in particle filtering framework using codebook model and adaptive labelling," in *Proc. of the 2015 Conf. on Research in Adaptive and Convergent Systems*, Prague, pp. 141–145, 2015.

[19] K. Fang, Y. Xiang, X. Li and S. Savarese, "Recurrent autoregressive networks for online multi-OBJECT tracking," in *2018 IEEE Winter Conf. on Applications of Computer Vision (WACV)*, Lake Tahoe, NV, USA, pp. 466–475, 2018.

[20] M. Li, X. He, Z. Wei, J. Wang, Z. Mu *et al.,* "Enhanced multiple-object tracking using delay processing and binary-channel verification," *Applied Sciences*, vol. 9, no. 22, pp. 4771, 2019.

[21] S. Chen, Y. Xu, X. Zhou and F. Li, "Deep learning for multiple object tracking a survey," *IET Computer Vision*, vol. 13, no. 4, pp. 1–14, 2019. https://doi.org/10.1049/iet-cvi.2018.5598.

[22] T. Badal, N. Nain and M. Ahmed, "Online multi-object tracking multiple instance based target appearance model," *Multimedia Tools and Applications*, vol. 77, no. 19, pp. 25199–25221, 2018.

[23] D. Riahi and G. A. Bilodeau, "Online multi-object tracking by detection based on generative appearance models," *Computer Vision and Image Understanding*, vol. 152, pp. 88–102, 2016.

[24] W. Tian, M. Lauer and L. Chen, "Online multi-object tracking using joint domain information in traffic scenarios," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 1, pp. 374–384, 2019.

[25] J. Xiang, G. Zhang and J. Hou, "Online multi-object tracking based on feature representation and Bayesian filtering within a deep learning architecture," *IEEE Access*, vol. 7, pp. 27923–27935, 2019.

[26] C. Wu, H. Sun, H. Wang, K. Fu, G. Xu *et al.,* "Online multi-object tracking via combining discriminative correlation filters with making decision," *IEEE Access*, vol. 6, pp. 43499–43512, 2018.

[27] M. Elhoseny, "Multi-object detection and tracking (MODT) machine learning model for real-time video surveillance systems," *Circuits Systems and Signal Processing*, vol. 39, no. 3, pp. 611–630, 2020.