Design of an effective multiple objects tracking framework for dynamic video scenes

Sunil Kumar Karanam[1], Narasimha Murthy Pokale Kavya[2]
[1]Department of Computer Science & Engineering, BMS College of Engineering, Affiliated to Visvesvaraya Technological University, Belagavi, India
[2]Department of Computer Science & Engineering, RNSIT Institute of Technology, Affiliated to Visvesvaraya Technological University, Belagavi, India

Article Info

Article history:

Received Jan 30, 2024
Revised Feb 24, 2024
Accepted Mar 21, 2024

Keywords:

Cost evaluation
Dynamic scene
Internet of things
Mobile object tracking video surveillance
Object detection accuracy
Public safety
Security

ABSTRACT

Nowadays, the applications corresponding to video surveillance systems are getting popular due to their wide range of deployment in various places such as schools, roads, and airports. Despite the continuous evolution and increasing deployment of object-tracking features in video surveillance applications, the loopholes still need to be solved due to the limited functionalities of video-tracking systems. The existing video surveillance systems pose high processing overhead due to the larger size of video files. However, the traditional literature report quite sophisticated schemes which might successfully retain higher object detection accuracy from the video scenes but needs more effectiveness regarding computational complexity under limited computing resources. The study thereby identifies the scope of enhancement in traditional object-tracking functions. Further, it introduces a novel, cost-effective tracking model based on Gaussian mixture model (GMM) and Kalman filter (KF) that can accurately identify numerous mobile objects from a dynamic video scene and ensures computing efficiency. The study's outcome shows that the proposed strategic modelling offers better tracking performance for dynamic objects with cost-effective computation compared to the popular baseline approaches.

Corresponding Author:

Karanam Sunil Kumar
Department of Computer Science and Engineering, BMS College of Engineering
Bull Temple Rd, Basavanagudi, Bengaluru, Karnataka 560019, India
Email: sunilkaranamresearch2020@gmail.com

1. INTRODUCTION

The growth of the global surveillance market has made dynamic object detection and tracking from video scenes popular in recent years. The advancement of computer vision technology and image processing makes this market size grow faster. The prime reason behind its rapid development is urbanization construction and the wide range of deployment of surveillance systems over large buildings, public places, parks, roads, and airports. Monitoring and surveillance systems play a crucial role in various aspects, viz., traffic movement management, automotive safety, activity-based recognition for cyber-security applications, and sports analysis [1], [2]. Here arise the requirements of reliable and accurate multiple-object tracking (MOT) so that the purpose of public safety concerns can be fulfilled under interconnected smart cities. The prime motive of single or multiple object tracking (MOT) is to consistently localize and identify several objects in a video sequence which facilitates video analysis applications of video surveillance systems. Most conventional works on MOT follow the idea of a tracking-by-detection framework due to its simplicity and effectiveness in fulfilling tracking requirements. Traditional MOT tracking consists of two stages of operations [3]–[8].

In the first stage of operations, the framework employs an object detector to detect objects of interest in the current video frame, whereas, in the second stage of operations, the detected objects are associated with the tracks from the previous frames to construct the trajectories further. Here the system associates the detected objects between frames using features that could be either location or appearance [9]–[11]. The recent progress in tracking-by-detection strategy has evolved towards solving the ambiguities associated with object detection. It can also handle the constraints that result in object detection failures. However, object detection is also closely studied with motion estimation, which is capable of identifying an object's mobility between two consecutive frames [12].

The segmentation plays a significant role in developing applications or techniques for tracking the video or the frame sequences in the video. There are studies which have also been worked in this direction where a significant study is being conducted by the authors [13], where the objective function for optimizing the accuracy of the segmentation uses two parameters: i) entropy and ii) clustering indices. Further, the validation of the method has experimented with traditional segmentation techniques that include: i) statistical region merging, ii) watershed and K-mean. Although they have tested this method on four different datasets, all these datasets are heterogeneous images, not video sequences. Minhas *et al.* [14] propose a novel concept of building a semantic segmentation network from skin features of high significance that fine-tunes the object boundaries information at different scales. The method is being tested and validated on many human activity databases. Cheng *et al.* [15] introduces a framework namely ViTrack which targets to efficiently implement multi-video tracking systems on edge to facilitates the video surveillance requirements. The problem formulation in the study addresses the core research challenges in three prime areas of video tracking in surveillance systems such as i) compressed sensing (CS) [16]–[18], ii) object recognition, and iii) object tracking. Xing *et al.* [19] explored the evolution of intelligent transportation systems where vehicular movement tracking is an important concern for traffic surveillance. The authors mostly emphasized on designing a real-time tracking system of vehicular movement considering complex form of scenes from captured video feeds. The authors introduce the tracking model namely NoisyOTNet which realises the problem of object tracking on complex video scenes as reinforcement learning with parameter space problem. The study explores traditional vehicle tracking methods such as correlation filter-based method [20]–[22], deep learning-based methods [23], [24] for vehicle tracking purposes. It finds that correlation-based methods and deep learning-based methods adopt static learning approach unlike reinforcement learning [25], [26].

Abdelali *et al.* [27] also addresses the problem of vehicular traffic surveillance and road violations and further attempts to design an approach to tackle this issue. In this regard the study introduces a fully automated methodical approach namely multiple hypothesis detection and tracking (MHDT) to deal with the multi-object tracking in videos. The research method jointly integrates Kalman filter [28] and data association-based tracking using YOLO detection [29] to robustly track vehicular objects in the complex video scenes.

Once the vehicle objects are detected then the system employs Kalman filter based tracking model. This applies a temporal correlation-based theory to track vehicles among one frame to another. The design of Kalman filter [28] is constructed in such a way where for each time instance of t, it provides the first prediction $\acute{y}t$. Here yt correspond to the state.

$$\acute{y}t = T \times yt - 1 \tag{1}$$

The Kalman filter also estimates the state prediction steps considering a covariance estimation calculation. The study also analyses various related works and observed that most of the studies and their incorporated algorithms consider convolutional neural network (CNN) as classifiers and it yields better accuracy which lies between 93% to 97%. The computational complexity is evaluated with respect to the estimation of bounding box coordinates (b) which states that the overall computational cost of the model stands as $O(\mathrm{b}^3 + \mathrm{b}^2 + \mathrm{b})$.

It has been observed that the variation factor in illumination causes significant challenges in video surveillance systems towards multiple object detection and tracking in the presence of motion factors. Even though various schemes being evolved and studied for several decades for different tasks, due to illumination variation factors, there remain constraints of deformation of mobile objects, pause motion blur, occlusions (full/partial) and camera view angle. These crucial aspects are yet unsolved problems associated with mobile object detection and tracking from dynamic video scenes. Also, the challenges with the traditional tracking systems are lack of effectiveness in localizing the object of interest properly in the presence of dynamic transition of background, lack of handling the presence of variation in aspect ratios, variation of intra-class objects, appropriate contextual information and presence of complex background [30], [31]. Apart from this, the most significant challenge arises with higher accuracy of multiple object detection and tracking while balancing considerable cost-effective computational performance, which is less likely explored in the existing systems of MOT models.

After reviewing the existing studies on MOT, the identified research problems outline the fact that even though there exist various form of work on MOT but the majority of the tracking models accomplish higher accuracy of detection and tracking at the cost of computational complexity, which is the similar case with the existing machine learning (ML) based approaches as well. Secondly, most studies do not consider contextual connectivity factors of an object with its background, which remains a challenge in the existing works. The appropriate inclusion of feature engineering is also missing in the existing ML-based MOT techniques for tracking dynamic mobile objects in the complex video scenes, where contextual scene information also plays a crucial role.

The study's problem statement is "To design a cost-effective and highly accurate MOT framework to perform object detection and tracking from complex video scenes considering contextual information is a highly challenging task". This proposed study addresses this problem, and a novel computational contextual framework is introduced for effective MOT. The novelty of this framework is that it can identify numerous mobile objects from the dynamic scenes and also reduces the cost of computational effort with a simplified tracking module. The contribution of the proposed system is it applies cost-effective modelling of assigning object detection in the current frame to existing tracks with an optimal estimator. It also explores the scope of improvement in mobile object detection considering the method of Gaussian mixture model (GMM) and improves the tracking performance using Kalman filter-based approach. Here the strategy also explores the association among the detected mobile objects from one frame to the next and overcomes the association problem. Here the inclusion of the Kalman filter method predicts the state variables effectively, which enhances the tracking performance with cost-effective trajectory formulation for the mobile objects even in the presence of complex and dynamic scenes. It has to be noted that the identification of mobile objects in the proposed study considers the contextual aspect of the object, which is also referred to as the line of movement (LoM). Another novelty of the proposed approach is implied design execution which makes the entire system computationally efficient when compared with the existing baseline approaches.

This new concept of dynamic tracking of numerous mobile objects takes advantage of GMM in the segmentation of objects. It also handles the constraints of traditional background subtraction methods towards the appropriate detection of moving objects. The study also further improvises the tracking model considering the potential features of the Kalman filter towards predicting the centroid of each track for motion-based tracking, through which it has also handled the track assignment problem. The experimental outcome further justifies how the formulated concept of LoM considers directionality movement that cost-effectively performs association among identified moving objects and performs tracking considering trajectory formulation. It also shows better identification performance by the tracking module with cost-effectiveness when compared with the baseline approaches. Unlike baseline studies, the proposed strategy offers a much lower response time with considerable processing execution and iterations.

2. METHOD

This part of the study formulates the analytical design modeling of the proposed cost-efficient dynamic tracking model which is capable of tracking multiple video objects with higher accuracy and computational efficiency. The study formulates the flow of the design with analytical research modeling to realise the working scenario of the proposed approach. It also involves a set of functional modules which operates on fulfilling the design requirements of the proposed system.

The block-based architecture of the proposed system in Figure 1 exhibits that it considers of a set of operational modules where the first module is associated with video I/O initialization where it constructs a video reader object and read the video file. Here the functionality constructs a reference object (Ov) which basically computes different attributes which is further discussed in the consecutive sections. Further it also initializes two players which are P1 and P2 respectively to visualize the computation of foreground mas and the video file sequence of (Vf). Further the system also constructs explicit functionalities to initialize the operations corresponding to Gaussian based detector for foreground and binary large objects (BloB) analyzer which also considers the reference object from the video sequence. Further the study also employs a dynamic mobile object detection module which basically constructs system objects to read the video file input sequence and also detect the foreground object. Here the study also enhances the operations of precise object detection by incorporating morphological operations which performs pre-processing over the data and make it suitable for video analysis for Blob Analyzer. The proposed strategy further applies GMM to perform precise object segmentation from the complex video scenes. The approach also considers initialization of tracking module where it constructs structure array fields. Finally, the study applies a Kalman filter to enhance the prediction of new location of track where the computation of centroid calculation and updating bounding box also evolves. Finally, the proposed system strategy also handles the track assignment problem for detected mobile object and here also use Kalman filter approach to perform detections to track assignment. It has to be noted that the entire process also minimizes the cost of track allocation where the track depicts the contextual LoM aspect for the mobile object. Further the proposed strategy performs the updating operations with respect to updating attributes and exhibits the final tracked mobile objects from the complex video scenes. It has to be noted that the core strategy of the proposed tracking module is to effectively locate the moving object or multiple objects over progressive time for a given V_f. Here the in the core strategy of the proposed system identifies the association problem and detects an object across multiple frames of a video stream. The core strategy of the proposed system also considers the fundamental principle of baseline models of tracking where the core philosophy is to initially detect the objects of interest in the video frame and further performing prediction to construct the LoM of object trajectories over the next consecutive frames of a video sequence. The proposed study handles the problem of data association by estimating the predicted locations and further associate the detections across the frames to formulate the trajectories for the LoM for respective objects.

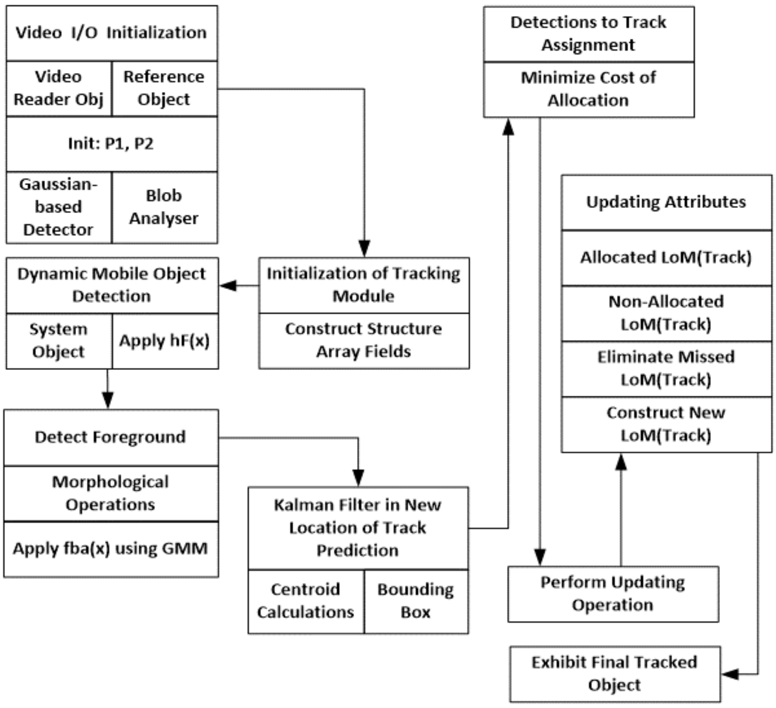


Figure 1. Architecture of the proposed MOT framework

2.1. Video input-output initialization

The computing process involved in the proposed in the proposed cost-effective dynamic tracking model initially employs a functionality for video input-output initialization. Here the system initially considers the input video (V_f) from the surveillance system. The information related to O_v is handled while constructing a reference object (O_v). Here the system employs a functionality of $f_{VR}(V_f) \rightarrow O_v$ which helps constructing this object. This phase of computation also comes under data exploration corresponding to the input V_f. The computation of the reference information corresponds to V_f. The exploration of the reference constructed object of Ov shows that the current time (C_t) refers to time stamp required to read the frame correspond to V_f. Here the tag attribute basically refers to as a reference to identify the Ov such as [tag ➔ O_v]. This is an optional name-value pair argument for the computation of the reference object from the video file. Here the user data (UD) is also constructed as an optional name-value pair attribute where it refers to a generic field to hold any new information which can be added to the reference object Ov. The processing and the computation of the V_f with the functionality of f VR(x) constructs a reference object Ov which holds the following properties as shown in the following Figure 2. The location attribute of path (P) contains the reference path to locate the video file. The general property of the reference object also includes the name of the video file nV_f which is associated with the object O_v. Here the duration (t) considers the total length of the V_f. The computed reference object of the V_f also consists of other important information related to video properties. Here in the Table 1 the attribute of b_p refers to the bits amount correspond to unit of pixel in the respective V_f. The attribute (*Fr*) also refers to the frame rate of the V_f computed in frame/s. It also computes the height (*h*) of the i^{th} frame ($frame_i$) of V_f in pixels along with width (*w*) of the ($frame_i$) in pixels. It also computes the number of frames ($frame_n$) along with the video format type.

The structure of Ov is finally constructed considering its essential properties to understand the input video data. The challenges arise in the conventional systems in detection of moving objects from the dynamic video scenes. In the problem of tracking the moving objects from the video sequences, segmentation of the dynamic region in the real-time synchronization is a quite challenging task because of various reasons which include complex and moving background, occlusion, motion blur, illumination variations and many more other factors. Therefore, to handle individual challenges many custom background subtraction methods is being evolved. The Table 1 further provides some of the important information about the properties of the V_f through Ov. The inference of Table 1 shows the important properties of V_f explored through the object and its associated methods of Ov.

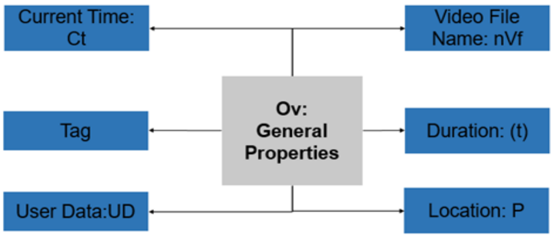


Figure 2. General properties: Ov

Table 1. Important properties of Vf

Sl. No	Property Name
1	Bits / Pixel (bp)
2	Frame Rate (Fr)
3	Height (h)
4	Width (w)
5	Number of Frames (n)
6	Video Format

In these methods the fast learning in the dense environment is the main focus of research. The explicit algorithm for the video input-output initialization as in Algorithm 1. The numerical algorithm modeling initially considers the video sequences through the video file (V_f) and initially creates two player objects as P1 and P2 for foreground mask and original video sequences respectively. The study further employs initialization and creation of an explicit function: function for the foreground detector(*ffd)* takes input parameter set as {Number of Gaussians (Ng), number of frames for the training (NTf), percentage of the minimum background ratio (MBr)} to construct the detector (D) to get advantages of the GMM [32], [33].

Algorithm 1: For video input-output initialization

```
1.  Input: Vf
2.  Output: D,B
3.  Begin
4.  Initialization of players
        a.  P1← foreground Mask
        b.  P2← Vf
5.  D←ffd(Ng,NTf,MBr)
6.  B←fba(BOp, AOp, COp, MBa)
7.  End
```

2.2. Computation measures of binary large object

The idea of GMM plays a crucial role to influence the outcome of background subtraction for the detection of moving objects. The idea of background subtraction allows in detecting the moving objects from dynamic video scenes. Which is applied in this proposed study considering GMM.

Idea of GMM: It has been observed that different background objects could more likely appear at the same pixel location of over a specific period of time. This arises a challenge of single-valued background model. Several researchers talks about the design and modeling of multi-valued background model which can easily cope with the multiple background objects appearing in video scenes [34], [35]. The model provides better description of both foreground and background values by describing the probability of observing a certain pixel value (x_t) at a specific time of (t). The method GMM computes each pixel within a temporal window (w) considering k number of mixtures of either single or multi-dimensional Gaussian distribution. Here if the value of k is larger that tends to stronger ability to deal with the disturbance background. If the sequence is observed with $x = \{x_1, x_2 \dots . x_t\}$ for a given pixel. Then the probability computation for observing a current pixel value at time t can be represented with the following mathematical (1).

$$P(x_t) = \sum_{i=1}^{k} \omega_{i,t} \; \eta(x_t \,, \mu_{i,t}, \Sigma_{i,t}) \qquad (2)$$

Here k represents the number of gaussian distributions which represents description for one of the observable foreground or background objects. In practical instances k value is likely to be reside within the range of $3 \leq k \leq 5$. The computation of Gaussians remains multi-variate for the purpose of describing the red, green, and blue values. Here $\mu_{i,t}$refers to the computation correspond to the mean value of ith gaussian in the mixture of models at the instance of t. Also $\Sigma_{i,t}$ computation denotes the covariance matrix of the ith gaussian at the time t. It has to be noted that here k is determined considering the computing aspects of both memory and computational power. Here the estimation of $\omega_{i,t}$ also denotes the factor of weight associated with ith Gaussian in the time instance of t. The principle here follows that the factor $\sum_{i=1}^{k} \omega_{i,t} = 1$ and $\eta(x_t \,, \mu_{i,t}, \Sigma_{i,t})$ considered to be Gaussian probability density function.

$$\eta\left(x_t \,, \mu_{i,t}, \Sigma_{i,t}\right) = \frac{1}{2\pi^{n/2}|\Sigma|^{1/2}} e^{-1/2 (x_t-\mu_t)^T \Sigma^{-1} (x_t-\mu_t)} \qquad (3)$$

The system modeling also considers the beneficial features associated with GMM. The background modeling of a grayscale image considers the value of n=1 and $\Sigma_{i,t} = \sigma^2{}_{i,t}$. However also when the modeling is applied on an RGB components then, it updates the values of n =3 and $\Sigma_{i,t} = \sigma^2{}_{i,t} I$. This computation of $\Sigma_{i,t} = \sigma^2{}_{i,t} I$ basically assumes the form of covariance matrix. Additionally, the system evaluates the incoming frames in real time, and GMM modifies its parameters in step-by-step response to the changing pixel value. Additionally, the pixels are mapped using a thresholding approach and the Gaussian model. The system further modifies the weights of the Gaussian components if a match is identified. This is how the background model estimation according to the distributions is carried out, and background pixel categorization is possible. The functionalities defined in the modeling of *ffd* (Ng, NTf, MBr) basically aims to form the foreground detector considering effective segmentation of background subtraction. The formation of the foreground detection object basically enables the potential features of GMM in which it compares the color or grayscale video frame with a background model as discussed in the (2) and (3).

This computational process enables a classification criterion to understand whether a certain pixel belongs to a part of background or foreground. This computational process is essential for background subtraction algorithms as this data exploration and pre-processing stage also helps eliminating the redundant attribute from the data and make it suitable for further computational analysis with truthful, accurate and complete information about the foreground object. Here the foreground mask (M_f) is computed which is associated with the D. And the algorithm correspond to background subtraction here efficiently computes the foreground objects (O_f) from the frame sequence of the $V_{f.}$ another explicit function for the purpose of analyzing the properties of connected regions is being used as function for BlobAnalyser (fba) that takes parameters as in set {Port for the bounding box (Bop), port for output area (AOp), Port for output centroid (COp, Minimum blob area (MBa)} that yield the blob (B). The underlying idea behind Blob analysis is to explore the statistics for labelled region in the binary frame of the video sequence. It basically helps segmenting the objects from the video sequence. The description of the Blob analysis can be seen in Figure 3.

The method of Bob analysis basically refers to analyzing the shape features associated with objects. Here the implications of the method Bob analysis basically identify the group of connected pixels which are more likely related with the moving object. The idea of Bob analysis is to explores the pixels connectivity and construct the Blob through the function fba(x). The connectivity among the pixels is represented with Blob. Firstly, the process computes the statistics associated with blob and further analyse the information of Blob which correspond to geometric characteristics which include points of borderline, and perimeter. These ideas and the standard methods are further incorporated in designing the object detection and tracking methodologies in the proposed system's context.

Figure 3. Blob analysis description

In the computation of statistics blob, the system analyses the output of AOp which represents a vector of pixels in the labeled regions. Here COp refers to an *N*-by-2 matrix of centroid coordinates c(x,y) which could be represented with the following matrix (3). Here N represents the number of Blobs. Here [x,y] represents the centroid coordinates. Here $[x_1,y_1]$ ➔ $[x_N\ y_N]$ implies that there are two blobs then the row and column coordinates of their centroids are $[x_1,y_1]$ and $[x_N\ y_N]$ respectively.

$$\mathrm{COp} = \begin{bmatrix} x_1 & y_1 \\ x_N & y_N \end{bmatrix} \tag{4}$$

The process of computation for the measure of Blob (B) also analyse the parameter MBa which refers to another *N*-by-4 matrix which is of [x,y] dimension. Here also N represents the number of blobs whereas [x,y] denotes the upper left corner of the bounding box. The analysis of the blob considering statistics returns a blob analysis system object (B). The analysis of B also constructs the significant properties of centroid, bounding box, label matrix and blob count in the output which are referenced with B. Finally, this computation process extracts the shape features of the objects of interest from the video sequence.

2.3. Initialization of the tracking module

The formulated design of the dynamic tracking model further constructs an empty structure array of tracking module T_m with six different fields. Which could be shown with the Figure 4. The structure array basically initializes six different fields such as (*ID*), Kalmar filter (*KF*), Age (a), bounding box (*Bx*), total visible count measure (*tVC*), and consecutive invisible count measure (*cIC*).

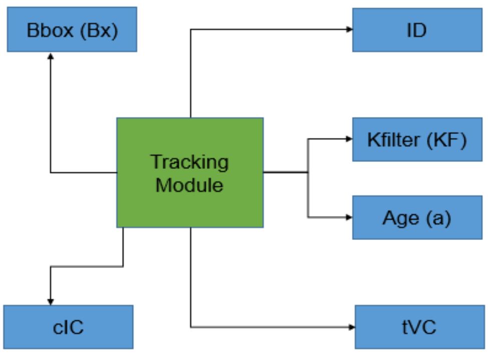


Figure 4. Structure array fields of T_m

The system also formulates a functionality to initiate the structure for initialization of array of tracks. Here each individual track $T_i \in T_m$. Here each track $T_i$represents the structure corresponding to the moving object appearing in the V_f. The design requirement for the tracking module in the proposed moving object detection and tracking strategy is to formulate the structure fields in such a way so that the state of the tracked object (T_O) can be maintained appropriately. Here ID refers to the integer ID of the track, Bx represents the current bounding box associated with the object. KF represents a Kalman filter object which is used for motion-based tracking. a refers to the frame count since the first detection of T. The consecutive visible count measure refers to the number of frames in which the track was detected. cIC represents the number of counts of consecutive frames for which the track was not detected. The process of computation of state correspond to the information utilized for detection of track allocation, track expiry and display.

2.4. Object detection module

The computing process further considers identification number of the next track (T_{ID}) and initiates the process of detecting moving objects considering a logical function $\mathrm{hF(x)}: \forall \mathrm{x} \in \mathrm{V_f}$. Here the function $hF(x)$ is a logical function which considers a set of objects associated with the video file (V_f) to read. The function basically returns a logical value from the set of $l \rightarrow \{1,0\}$. If the function hF(x) returns the value 1 that implies that there is a video frame F_i available to read. The process further also applies another function of rF(x) which reads the video frame from the file then the process further detects the binary mask (Bm) from the F_i. The binary mask is of same size of the input F_i. Here the reading of the frame considers constructing of system object (obj). The process of detecting objects from the F_i enables another explicit function of $dO(x)$, here the function considers the input of F_i and process it to generate three distinct attributes which are $\{c, Bx, m\}$. Here c refers to the centroid calculation considering the detected objects, Bx is bounding box of computation of the detected object followed by the measure of mask (m). The initial computation of the function $dO(x)$ considers the video frame sequence of F_i and identify the mask Bm and computes a logical matrix $Lm(r, c)$. Here the computing function of binary mask computation basically performs motion segmentation considering an explicit method of $\mathrm{ffd}(x)$ [32]. The following analytical algorithm, Algorithm 2, basically modeled to present the proposed work-flow associated with object detection from video where the advantageous factors of the method GMM us utilized to perform blob analysis.

The computed mask further undergoes through pre-processing operations as defined by morphological operations. The morphological operation here subjected to eliminate redundant attributes of pixels and also fill the missing gaps in the blobs for the resulting mask Bm. The process further performs morphological operation (MO) over $Lm(r, c)$. It applies two functions such as I_1 and I_2 to perform the morphological operations where I_1 opens the $Bm[Lm]$ and performs morphological operation over it with respect to structuring element of size $[s \times s]$ and update the values of Bm. The process also further applies another function of I_2 for morphological close operation over Bm considering dilation followed by erosion [33]. Finally, another function of I_3 helps filling the image regions and gaps and make the updated Bm suitable for effective blob analysis. The customized function of $\mathrm{dO(x)}$ finally returns three attributes of $\{\mathrm{c, Bx, m}\}$ and terminates the process of execution.

Algorithm 2: For object detection from video

```
Input:V_f
Output:{c, Bx, m}
Begin
1. Define:dO(x), construct system object (obj)
2. Define: hF(x): ∀x ∈ V_f
3. While (F_i = 1)
4.      rF(x) → F_i
5. End
6. Return: l → {1,0}
7. Bm[Lm] ← ffd(x): ∀F_i , Lm(r, c)
7. MO→ Bm[Lm(r, c)], for {I_1, I_2, I_3}
8. Apply fba(x): ∀x ∈ Bm (1), (2) for GMM
9. Return {c, Bx, m}
End
```

2.5. Prediction module for new position of line of movement

The core strategy developed in the proposed system targets appropriate identification and tracking of mobile objects from a complex set of scenes. Here the scenes are captured from a camera which is mounted in static position. The formulated tracking module further considers $T_i \in T_m$ and apply a function of $P_{NT}(x)$ over the tracks with the inclusion of Kalman filter approach to predict the new location of the LoM. Here the system considers the computation of Bx considering the updates on T_i for LoM and further initially predict and estimate the current location of the track of LoM considering the function of $P_{NT}(x)$ it optimizes the process of prediction of centroid P_{ci} considering the approach Kalman Filter (KF). The computation process can be represented in (5).

$$P_{ci} \leftarrow P_{NT}(x): \forall x \in T_i, KF \tag{5}$$

Here the computation of prediction of centroid basically determines the current location attributes of the T_i considering Kalman filter object. The further computation considers shifting of the Bx in such a so that its center lies in the P_{ci}. It is achieved with the (6).

$$P_{ci} = P_{ci} - {}^{Bx(k)}/_{2} \quad (6)$$

The function further updates the new location of the T_i with respect to the LoM for P_{ci}. The proposed system also explores the shape-based features of the target object which further assist in optimal estimation of motion associated with the identified object on its LoM. The next computational process performs LoM allocation to the identified objects of interest.

2.6. Line of movement allocation to the identified objects

In the functional module of the proposed system the estimation of the new position of track (LoM) is predicted considering the approach of Kalman filter over the progressive $F_i \in V_i$. In this stage of computation, the proposed model the appropriate allocation of LoM to the identified moving objects take place along with the cost evaluation. The system here employs another function of $A_{LoM}(x)$ which computes the number of identified objects nIO from the c_i and compute the cost of assignment $Cost_{alloc}$ considering the (7).

$$Cost_{alloc} = A_{LoM}(x): \forall x_1 \rightarrow T, x_2 \rightarrow KF, x_3 \rightarrow c \quad (7)$$

Finally, the optimized estimator of this function solves the allocation problem of identified objects to the track or LoM for multiobject tracking. Also compute four different attributes such as allocated LoM, non-allocated LoM and non-allocated identifed objects. The Algorithm 3 shows the design strategy of the tracking module which has got influenced from the [36], [37] for solving the problem of allocation of detections to tracks during multiobject tracking.

Algorithm 3: For multi-object tracking
Input: $T_i \in T_m$
Output: FO_T
Begin
1. Init T_i. Bx
2. Update $Bx \leftarrow T_i(Bx)$
3. Compute current location of LoM

$$P_{ci} \leftarrow P_{NT}(x): \forall x \in T_i, KF \quad (5)$$

4. Predict the new position of LoM

$$P_{ci} = P_{ci} - {}^{Bx(k)}/_{2} \quad (6)$$

5. Update T_i with respect to the LoM for P_{ci}
6. LoM Allocation to identified objects
7. Evalutate Cost

$$Cost_{alloc} = A_{LoM}(x): \forall x_1 \rightarrow T, x_2 \rightarrow KF, x_3 \rightarrow c \quad (7)$$

8. Update allocated LoM, Non-Allocated LoM
9. Eliminate Missed LoM, Construct New LoM
10. Exibit Final Tracked Objects (FO_T)
End

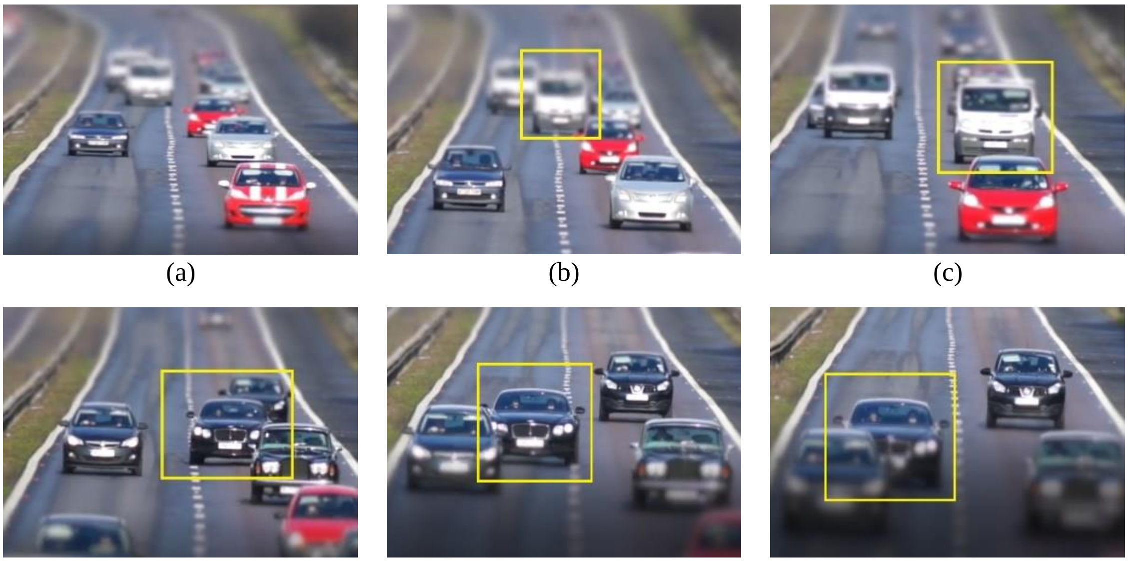
Once the cost evaluation metric is computed for solving the assignment problem, further the process executes updating of allocation of LoM. Here the algorithm strategy estimates the location of the detected objects considering another approach based KF. Here the KF based method basically performs correction of the moving object's location considering LoM. Here the finetuning of LoM for a detected object also takes place where predicted Bx is replaced with the detected Bx. Finally, the age corresponds to T_i is updated with visibility. Finally, the proposed algorithm strategy computes the updated allocated LoM, non-allocated LoM, eliminate the missed LoM and construct new LoM prior exhibiting the FO_T attribute. It can be seen that the design strategy of the proposed MOT module is quite simplistic and less-iterative which has also enhanced the computing speed of analytical operation of the algorithm. The methods are computationally lesser complex which perform the tracking operations for the implemented idea and also offers cost effective MOT. The next section further discusses experimental outcome obtained from the simulation of the proposed strategy for multi-object tracking over complex video sequence.

3. RESULTS AND DISCUSSION

This section discusses about the simulation study outcome obtained from implementing the proposed multiple objects tracking framework for dynamic video scenes. The study implementation of the analytical algorithms is scripted over MATLAB numerical computing environment supported by 64-bit conventional windows system. The study also considers different set of multiple mobile object-oriented datasets as referred from [38]. It has to be noted that this proposed study is the continuation of our previous research works [39], [40].

This phase of the study basically judges the outcome of the proposed system and exhibits its effectiveness in terms of visual and comparative performance analysis from both accuracy of tracking and cost point of view. The initial experimental analysis considers moving object detection and tracking for a single test object. In this regard the system considers the case of two-lane system of roadway where the idea is to track a single moving vehicle attempting to change the lane. The study considers tracking of a white and a black vehicle which are moving and attempted to change the lane which is further shown in the Figure 5.

The analysis and interpretation of the visual outcome of Figure 5 highlights that the white vehicle was initially moving over its assigned left lane where it has been detected considering the proposed tracking module Figure 5(a). However, it has suddenly shifted to the right lane and continued its journey over the right lane as tracked by the proposed tracking module Figures 5(b)-5(c). A similar tracking outcome is also found in the case of black vehicle which has changed its lane from right to left and continued its journey on the left lane of the roadway Figures 5(d) to 5(f). It has to be noted that the tracking of the target mobile object from a very complex dynamic scene is achieved effectively by the proposed tracking module even in the presence of partial occlusion between the target vehicle and other similar vehicles over the frame sequence. The outcome clearly shows that for a single mobile test object the proposed tracking module has achieved higher accuracy in tracking the fast-moving object. However, the performance assessment is further extended for multiple moving objects as well which is further shown in the Figure 6.

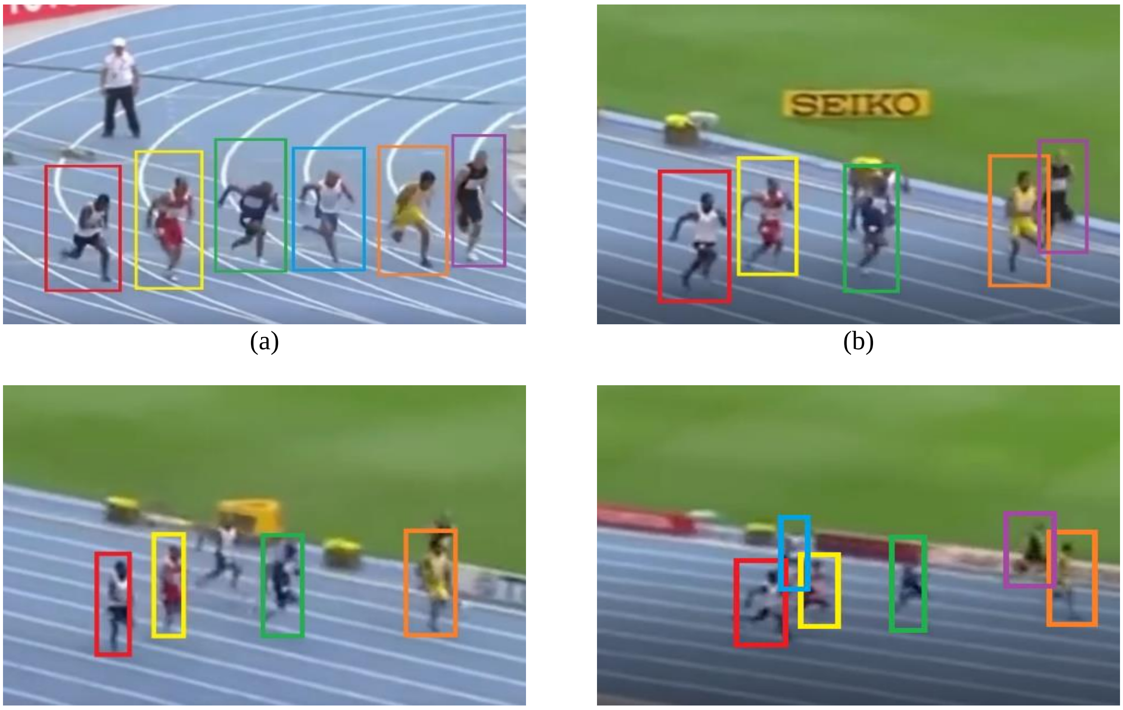


(a) (b) (c)

(d) (e) (f)

Figure 5. Tracking of a single test object: (a) no tracking of white vehicle, (b) tracking of white vehicle in the middle of roadway, (c) tracking of white vehicle in the right lane, (d) tracking of black vehicle in the right lane, (e) in the left lane, and (f) continued its journey on the left lane

Another test instance in the proposed study model is considered where identification and tracking of multiple mobile objects are performed considering the proposed MOT framework. The Figure 6 clearly shows that the multiple mobile objects are distinctly indexed initially in Figure 6(a) whereas in the sequence of other frames the detection and tracking is slightly affected due to occlusion. However, in Figures 6(b)-6(d) majorly features are positively determined and in the end the accuracy of tracking also improved irrespective of the presence of partial occlusison. It can also be seen that the proposed study model retains a proper balance between the performance accuracy of tracking and computational complexity which is further illustrated in the following comparative Table 2.



(a) (b) (c) (d)

Figure 6. Tracking of multiple test objects in the presence of occlusions: (a) tracking of multiple objects distinctly indexed, (b) occlusion between two running objects, (c) major occlusion between two running objects, and (d) occlusion between the three running objects

Table 2. Comparative analysis based on observations

Approaches	Accuracy (%)	Response time	Number of processing steps	Iterativeness	Cost evaluation
Cheng *et al*. [15]	96.00	Slow	Higher	Higher	No
Abdelali *et al*. [27]	92.50	Faster	Higher	Very Higher	No
Chen *et al*. [30]	93.3	Medium	High	Medium	No
Aslam and Sharma [32]	95.1	High	Higher	Higher	No
Proposed tracking	96.22	Very less	Very less	No	Yes

The interpretation of the observational outcome from the Table 2 shows that the proposed system offers comparatively better performance of tracking along with balancing the cost factors where it also obtained considerable response time along with executional steps which doesn’t involve much complex procedure. The cost evaluation also shows how the proposed tracking model has addressed the assignment of detections to track problem effectively while minimizing the cost factors. The insights from the comparative study outcome shows that when compared with the approaches in [15], [27], [30], [32] the proposed tracking model attains considerably better tracking accuracy which is approximately 96.22% and comparable with the exsiting baseline models. Also, the critical findings of the study shows that the proposed model is found to be better in terms of response time, interativeness, complexity and cost of compuatation factors. Another strength factor of the study model is that it is capable of providing better accuracy even in the presence of low ir medium size of video data.

4. CONCLUSION

The study introduces an effective computational framework for multi-object tracking where it considers tracking a set of mobile objects from a given dynamic video scene. The study attempts to provide a simplistic design schema for the proposed system. It aims to detect moving objects in each frame precisely and precisely track the identified objects' movement over successive frames, even in partial occlusion. The study also handles the problem of assigning the detection to each track, considering an efficient distance computation using the Kalman filter. The strategic modelling performs the detection of moving objects considering the background subtraction method, which is based on GMM, and the Blob analysis further generates the group of connected pixels for the moving object, which is further considered to determine the association of detections of the moving objects for its LoM. The contribution of the proposed model is as follows: i) unlike the existing system, it offers a simplistic design modelling of tracking model, which attains better accuracy of LoM for moving objects without compromising the computational performance; ii) it basically enhances the computation operation with object-oriented design modelling of system objects and also performs better foreground detection and lump analysis, iii) the proposed system also performs contextual attribute based LoM analysis for the directionality of movement of an object that assists in effective tracking of multiple objects over successive frame sequence, and iv) the inclusion of optimal estimator in the proposed system not only reduces the noise but also offers effective management of allocated and non-allocated LoM to balance the cost factors which also addresses the assignment problem in dynamic tracking. Overall, it is pretty clear that the simplistic study model of the proposed system retains a better balance between accuracy and computation cost while performing detection and tracking of a mobile object over dynamic video scenes. It has to be noted that the study considered specific form of dataset for the evaluation of the proposed tracking model and also considered specific volume of dataset to study the effectiveness of the system. The model has not been evalauated under increasing number of samples. The future scope of the research aims to implicate the study model towards accomplishing better public safety and security by considering faster, more reliable and accurate object tracking among the interconnected smart cities.

REFERENCES

[1] M. H. Sedky, M. Moniri, and C. C. Chibelushi, "Classification of smart video surveillance systems for commercial applications," *IEEE International Conference on Advanced Video and Signal Based Surveillance*, vol. 2005, pp. 638–643, 2005, doi: 10.1109/AVSS.2005.1577343.

[2] Y. Wang, "Development of AtoN real-time video surveillance system based on the AIS collision warning," *ICTIS 2019 - 5th International Conference on Transportation Information and Safety*, pp. 393–398, 2019, doi: 10.1109/ICTIS.2019.8883727.

[3] T. Zhang, B. Ghanem, and N. Ahuja, "Robust multi-object tracking via cross-domain contextual information for sports video analysis," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 985–988, 2012, doi: 10.1109/ICASSP.2012.6288050.

[4] F. Wu, S. Peng, J. Zhou, Q. Liu, and X. Xie, "Object tracking via online multiple instance learning with reliable components," *Computer Vision and Image Understanding*, vol. 172, pp. 25–36, 2018, doi: 10.1016/j.cviu.2018.03.008.

[5] J. Gwak, "Multi-object tracking through learning relational appearance features and motion patterns," *Computer Vision and Image Understanding*, vol. 162, pp. 103–115, 2017, doi: 10.1016/j.cviu.2017.05.010.

[6] M. Weber, M. Welling, and P. Perona, "Unsupervised learning of models for recognition," *Computer Vision - ECCV 2000*, vol. 1842, pp. 18–32, 2000, doi: 10.1007/3-540-45054-8_2.

[7] M. A. Naiel, M. O. Ahmad, M. N. S. Swamy, J. Lim, and M. H. Yang, "Online multi-object tracking via robust collaborative model and sample selection," *Computer Vision and Image Understanding*, vol. 154, pp. 94–107, 2017, doi: 10.1016/j.cviu.2016.07.003.

[8] M. Han, W. Xu, H. Tao, and Y. Gong, "An algorithm for multiple object trajectory tracking," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2004, doi: 10.1109/CVPR.2004.1315122.

[9] 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, doi: 10.1016/j.cviu.2016.07.012.

[10] S. Huang, S. Jiang, and X. Zhu, "Multi-object tracking via discriminative appearance modeling," *Computer Vision and Image Understanding*, vol. 153, pp. 77–87, 2016, doi: 10.1016/j.cviu.2016.06.003.

[11] D. B. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, 1979, doi: 10.1109/TAC.1979.1102177.

[12] J. Prokaj, M. Duchaineau, and G. Medioni, "Inferring tracklets for multi-object tracking," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 37–44, 2011, doi: 10.1109/CVPRW.2011.5981753.

[13] J. D. H. Resendiz, H. M. M. Castro, and E. T. Leal, "A comparative study of clustering validation indices and maximum entropy for sintonization of automatic segmentation techniques," *IEEE Latin America Transactions*, vol. 17, no. 8, pp. 1229–1236, 2019, doi: 10.1109/TLA.2019.8932330.

[14] K. Minhas *et al.*, "Accurate pixel-wise skin segmentation using shallow fully convolutional neural network," *IEEE Access*, vol. 8, pp. 156314–156327, 2020, doi: 10.1109/ACCESS.2020.3019183.

[15] L. Cheng, J. Wang, and Y. Li, "ViTrack: efficient tracking on the edge for commodity video surveillance systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 33, no. 3, pp. 723–735, 2022, doi: 10.1109/TPDS.2021.3081254.

[16] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006, doi: 10.1109/TIT.2005.862083.

[17] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006, doi: 10.1109/TIT.2006.871582.

[18] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006, doi: 10.1109/TIT.2006.885507.

[19] W. Xing, Y. Yang, S. Zhang, Q. Yu, and L. Wang, "NoisyOTNet: a robust real-time vehicle tracking model for traffic surveillance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 2107–2119, 2022, doi: 10.1109/TCSVT.2021.3086104.

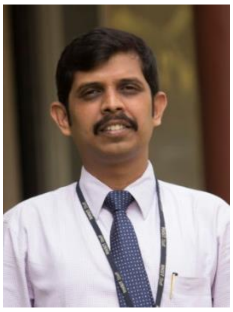
[20] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, 2015, doi: 10.1109/TPAMI.2014.2345390.

[21] M. Danelljan, G. Bhat, F. Shahbaz Khan, and M. Felsberg, "ECO: Efficient convolution operators for tracking," *30th IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2017, pp. 6931–6939, 2017, doi: 10.1109/CVPR.2017.733.

[22] J. Valmadre, L. Bertinetto, J. Henriques, A. Vedaldi, and P. H. S. Torr, "End-to-end representation learning for correlation filter based tracking," *30th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5000–5008, 2017, doi: 10.1109/CVPR.2017.531.

[23] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "SIAMRPN++: Evolution of siamese visual tracking with very deep networks," *The IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 4277–4286, 2019, doi: 10.1109/CVPR.2019.00441.
[24] H. Fan and H. Ling, "Siamese cascaded region proposal networks for real-time visual tracking," *The IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 7944–7953, 2019, doi: 10.1109/CVPR.2019.00814.
[25] S. Yun, J. Choi, Y. Yoo, K. Yun, and J. Y. Choi, "Action-decision networks for visual tracking with deep reinforcement learning," *30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pp. 1349–1358, 2017, doi: 10.1109/CVPR.2017.148.
[26] D. Zhang and Z. Zheng, "High performance visual tracking with siamese actor-critic network," *Proceedings - International Conference on Image Processing, ICIP*, vol. 2020, pp. 2116–2120, 2020, doi: 10.1109/ICIP40778.2020.9191326.
[27] H. A. I. T. Abdelali, H. Derrouz, Y. Zennayi, R. O. H. Thami, and F. Bourzeix, "Multiple hypothesis detection and tracking using deep learning for video traffic surveillance," *IEEE Access*, vol. 9, pp. 164282–164291, 2021, doi: 10.1109/ACCESS.2021.3133529.
[28] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Fluids Engineering, Transactions of the ASME*, vol. 82, no. 1, pp. 35–45, 1960, doi: 10.1115/1.3662552.
[29] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *The IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.
[30] J. Chen, Z. Xi, C. Wei, J. Lu, Y. Niu, and Z. Li, "Multiple object tracking using edge multi-channel gradient model with ORB feature," *IEEE Access*, vol. 9, pp. 2294–2309, 2021, doi: 10.1109/ACCESS.2020.3046763.
[31] L. Chen, H. Zheng, Z. Yan, and Y. Li, "Discriminative region mining for object detection," *IEEE Transactions on Multimedia*, vol. 23, pp. 4297–4310, 2021, doi: 10.1109/TMM.2020.3040539.
[32] N. Aslam and V. Sharma, "Foreground detection of moving object using Gaussian mixture model," *2017 IEEE International Conference on Communication and Signal Processing, ICCSP 2017*, pp. 1071–1074, 2017, doi: 10.1109/ICCSP.2017.8286540.
[33] R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 532–550, 1987, doi: 10.1109/TPAMI.1987.4767941.
[34] F. Wang, F. Liao, Y. Li, and H. Wang, "A new prediction strategy for dynamic multi-objective optimization using Gaussian mixture model," *Information Sciences*, vol. 580, pp. 331–351, 2021, doi: 10.1016/j.ins.2021.08.065.
[35] X. Lin, C. T. Li, V. Sanchez, and C. Maple, "On the detection-to-track association for online multi-object tracking," *Pattern Recognition Letters*, vol. 146, pp. 200–207, 2021, doi: 10.1016/j.patrec.2021.03.022.
[36] M. L. Miller, H. S. Stone, and I. J. Cox, "Optimizing murty's ranked assignment method," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 3, pp. 851–862, 1997, doi: 10.1109/7.599256.
[37] J. Munkres, "Algorithms for the assignment and transportation problems," *Journal of the Society for Industrial and Applied Mathematics*, vol. 5, no. 1, pp. 32–38, 1957, doi: 10.1137/0105003.
[38] L. Wen *et al.*, "UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking," *Computer Vision and Image Understanding*, vol. 193, 2020, doi: 10.1016/j.cviu.2020.102907.
[39] K. S. Kumar and N. P. Kavya, "An efficient unusual event tracking in video sequence using block shift feature algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 7, pp. 98–107, 2022, doi: 10.14569/IJACSA.2022.0130714.
[40] K. S. Kumar and N. P. Kavya, "Compact scrutiny of current video tracking system and its associated standard approaches," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 12, pp. 398–408, 2020, doi: 10.14569/IJACSA.2020.0111249.

BIOGRAPHIES OF AUTHORS



Sunil Kumar Karanam holds the Bachelor of Engineering in Computer science and Engineering. Along with a M.Tech. degree from VTU Belagavi. He is currently an assistant professor at Department of Computer Science and Engineering, BMS College of Engineering, Bull Temple Rd, Basavanagudi, Bengaluru, Karnataka, India. His research includes meta-heuristics, network security, object tracking and surveillance, machine learning, data mining, deep learning, and computer vision. He can be contacted at email: sunilkaranamresearch2020@gmail.com.

Narasimha Murthy Pokale Kavya holds Bachelor of Engineering in Computer Science and Engg. along with M.S. in software systems and Ph.D. in computer science from VTU Belagavi. She has vast experience of 26 years in the field education and research. She is currently a Professor in Department of Computer Science and Engineering, RNSIT, Bengaluru. She has published around 90 research papers in reputed international journals including IEEE, Elsevier, Springer (SCI and Web of Science). Has 94+ citations in Google scholar as of Jan 2024. Her main areas of expertise are machine learning, artificial intelligence, and big data analytics. She can be contacted at email: kavya.np@rnsit.ac.in.