Circuits, Systems, and Signal Processing (2020) 39:611–630
https://doi.org/10.1007/s00034-019-01234-7

Multi-object Detection and Tracking (MODT) Machine Learning Model for Real-Time Video Surveillance Systems

M. Elhoseny[1]

Received: 6 February 2019 / Revised: 26 July 2019 / Accepted: 3 August 2019 / Published online: 20 August 2019

Abstract
Recently, video surveillance has garnered considerable attention in various real-time applications. Due to advances in the field of machine learning, numerous techniques have been developed for multi-object detection and tracking (MODT). This paper introduces a new MODT methodology. The proposed method uses an optimal Kalman filtering technique to track the moving objects in video frames. The video clips were converted based on the number of frames into morphological operations using the region growing model. After distinguishing the objects, Kalman filtering was applied for parameter optimization using the probability-based grasshopper algorithm. Using the optimal parameters, the selected objects were tracked in each frame by a similarity measure. Finally, the proposed MODT framework was executed, and the results were assessed. The experiments showed that the MODT framework achieved maximum detection and tracking accuracies of 76.23% and 86.78%, respectively. The results achieved with Kalman filtering in the MODT process are compared with the results of previous studies.

Keywords Video surveillance · Multi objection detection and tracking (MODT) · Machine learning · Kalman filtering · Region growing

1 Introduction

In today's world, machine learning has become increasingly popular in computer vision-based applications due to the state-of-the-art results achievable in the image classification, object detection and natural language processing domains. The reasons for the popularity of machine learning are twofold: the high availability of datasets and powerful graphics processing unit (GPU) capabilities. Modern techniques satisfy the training needs of machine learning algorithms, which require both large datasets

✉ M. Elhoseny
Mohamed_elhoseny@mans.edu.eg

1 Faculty of Computers and Information, Mansoura University, Mansoura, Egypt

and extensive resources. In video observation applications, video object tracking is an essential task because it yields interrelated temporal data about the moving objects (MO) [12]. In the real world, tracking the movements of objects is conducive to a variety of applications, for example enabling better security by utilizing video feeds [6]. The most difficult tasks involved in identifying moving objects are sudden changes in illumination [28], lengthy occlusions among moving objects, shadows and non-stationary background objects [11]. Object theories act on the second layer by clustering moving features as well as performing appearance-based object segmentation [24]. To accomplish tasks such as background removal (or background subtraction in some cases), the utilized approach compares each frame with a reference or background model [4, 17]. To perform efficient tracking in video segments, one segmentation technique proposed in the literature uses the least squares tracking scheme [1]. The general prerequisites for a background removal algorithm are accurate object contour detection (spatial accuracy) and the temporal stability of the detection (temporal coherency) [6]. During video tracking, the background scene of each video clip is modelled by the minimum and maximum intensity values. Furthermore, the maximal temporal derivative [25] for each pixel is recorded and updated instantly. Each frontal object area is matched to the current object set using a combination of shape analysis and tracking [2]. 'Object discovery' is a procedure that distinguishes objects moving within a region. It is fundamental for a tracking strategy to know about the objects in the region [19]. Segmentation plays a vital role in assembling pixels into three regions, i.e. image portions related to individual surfaces, products or acknowledged parts of the objects [5]. The segmentation result is nothing but an arrangement of segments that cover an ideal image or an arrangement of contours extracted from the image [20]. Furthermore, morphological segmentation should be utilized when the intensity values of objects are similar to those of the background image [10]. In contrast, probabilistic tracking strategies such as particle channels and Kalman channels can also model the necessary data [14]. For example, the position as well as the velocity of the probability distribution functions of tracked objects can be recursively refreshed based on a current estimation [22]. The Kalman channel controls the minimum variance; it ensures that the minimum upper limit of the evaluated error variance is less than the minimum of the filter. Consequently, it can overcome parameter perturbations [13]. In the current study, the probability-based grasshopper algorithm (PGA) was chosen for detection to enhance the detection rate with the help of the dilation and erosion processes.

In the proposed method, an optimal Kalman filtering technique is used to track the MO in video frames. The reason for selecting PGA to perform the filtering optimization is to enhance the fundamental propositions of parameters in filtering used to assess the movement values. In the MO detection process, both morphological tasks and the RG procedure were utilized. Finally, the execution results of this MODT procedure were assessed and compared with the current approaches. The rest of this paper is organized as follows: Sect. 2 examines the recent literature concerning MODT approaches, and Sect. 3 discusses the proposed methodology. The experimental results are presented in Sect. 4, after which the proposed MODT process is concluded and suggestions are made for future work.

2 Literature Survey and Current Issues in MODT

Mahalingam et al. [15] proposed a video tracking strategy that is divided into three stages: detection, tracking and evaluation. A Mixture of Adaptive Gaussian (MoAG) approach was proposed to effectively accomplish segmentation. In addition, a fuzzy morphological filter model was applied to remove noise present in the foreground area-sectioned edges. Finally, the evaluation stage included feature extraction and classification. The performance of the proposed method was compared with existing strategies such as k-NN and MLP in terms of precision, recall and f-measure along with the receiver operating characteristic (ROC) curve.

A multi-type multi-object tracking algorithm was proposed by Tian et al. [21], into which online input data were fed between identification and tracking forms operations that followed the tracking-by-recognition strategy. During the recognition step, the objects were identified by locators that were balanced by tracking data. In addition, to address tracking situations with different many-sided qualities, the objects were grouped into two classes, i.e. single objects and multiple objects, and were managed with multiple techniques.

A tracking algorithm comprises two stages, as discussed by Hamuda et al. [8]. Initially, Kalman filtering is applied to anticipate the new position of an object in video groupings. In addition, a data affiliation technique was utilized to coerce each and every recognized product that shows up in each image to an appropriate crop operation. A review network was used to assess the discovery and tracking results. With the assistance of the tracking scheme, location failures were reduced, particularly in bright conditions, to such an extent that the overall discovery execution increased from 97.28 to 99.3404%.

In 2018, a tracking algorithm proposed by Anindaputri et al. [9] that consolidated both the mean shift and the particle Kalman filter to address the tracking problem. In this strategy, mean shift was utilized as the master tracker when the target object was not occluded. When an occlusion was present or mean shift tracking provided inconclusive results, the particle Kalman filter was adopted as the master tracker to enhance the tracking results.

In Chen et al. [3], the essential aspect was to perform lower-grade detection but spread the outcomes across two scales, which involved considerably less time and allowed using less expensive systems by capitalizing on the stable relationships. Using this system, it is possible to investigate different techniques to adjust execution speed and cost. This adaptability was exploited: additionally, the researcher built an ingenious plan with the detector summoned on demand; an enhanced trade-off was acquired using this approach.

According to Sahoo et al. [21], it is incorrect to correlate the movement data of a gradually moving object with those of a fast moving object. A quick and proficient segmentation algorithm was proposed to discover gradually moving objects in a video grouping. Using the proposed technique to remove a progressively moving object in a video involved three stages. In the initial step, a frame averaging strategy was proposed to identify and extricate the movement data. In the second step, valley-based thresholding was performed to segment every case in a video.

Ewees et al. [23] opined that opposition-based learning is connected to just 50% of the answers when the multifaceted nature of time is reduced. To analyse the execution of the proposed OBLGOA, six arrangements of a series of experiments were performed that incorporated 23 benchmark capacities with four design issues. The trials revealed that the results of the proposed calculation were better than those of existing, well-known algorithms in this area. Overall, the acquired outcomes demonstrated that the OBLGOA algorithm could yield focused results for optimizing engineering issues compared with existing algorithms.

2.1 Problems in MODT

The literature shows that the MODT process has a few drawbacks that are examined in this section. Object tracking can be a tedious procedure because of the measurement of the data contained in the video. Each tracking algorithm requires an ‘object location system’—either in each frame or whenever a new object appears in a frame. It is challenging to track moving targets due to occlusions and interference in the object’s vicinity such as image backgrounds and surrounding noise. These factors reduce the efficiency of movement vector fields that utilize fuzzy sets to achieve better tracking productivity [9]. Most detection techniques use information from a single frame to detect a moving object. The location-based object models work by tracking objects using their colour distributions, and they represent objects based on their colours [1]. In contour-based tracking algorithms [3, 21], objects are tracked by considering their outlines as boundary contours. Although numerous detection methods have been proposed, only a few have addressed the multiple object tracking problem. To overcome the limitations of the existing versions of detection techniques, this study attempted to devise a new MODT framework to improve the detection rate.

3 Multi-object Detection and Tracking (MODT) Methodology

This work develops a methodology for multi-object (MO) detection and tracking using an innovative technique, as depicted in Fig. 1. One complexity in MO detection is that it is difficult to identify a particular class of object among different static images. The captured input video clips were converted into several distinct frames for MODT analysis. A Gaussian filter was used to remove noise. In addition, background subtraction was performed on the newly formed frames because this process enables quick detection of objects in frames. Then, a morphological operation (e.g. MO detection algorithm) and a region growing (RG) operation were proposed to segment the objects with the help of some additional features from the video frames with noise removed. After a MO was detected, particle-based Kalman filtering with PGA techniques was applied to track the MO of the video frame. Moreover, during the testing process, the tracked objects were matched with other noiseless frames in order to follow the MO based on similarity measures. The entire model was implemented and analysed in the MATLAB platform; a detailed explanation of the proposed work is described in the following subsection.

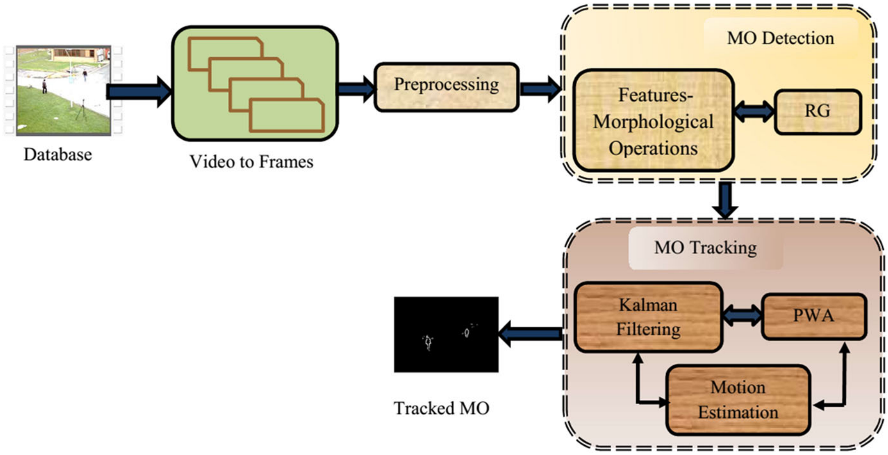


Fig. 1 Graphical representation of the proposed methodology

3.1 Video-to-Frame Conversion

Based on frame count, the input video clips were converted into multiple numbers of frames for MO detection and tracking process. Usually, a single frame is insufficient for recognizing moving objects; however, it can be utilized as a background frame to identify the MO. These frames were then considered for further pre-processing and the ODT process.

3.2 Pre-Processing: Noise Removal

During pre-processing, noise was removed from the video frames, and the nature of the images was enhanced. If the object in a video sequence scales from one frame to another frame, they are based on a relationship that fails to yield satisfactory results [18]. In the proposed model, a Gaussian filter was used to remove noise and other undesirable features during the extraction process.

Gaussian Model For each frame, the distribution of the pixel strength was calculated by representing a combination of the Gaussian function and the probability of intensity in the frame at time 't'. This process can be modelled as follows:

$$N(f) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-f^2/2\sigma^2} \tag{1}$$

where f is the interframe distance and σ is the standard deviation. This operator is a 2D convolution operator that can be utilized to 'blur' the images; then, noise and unwanted details can be eliminated. This filter alone is capable of blurring the edges and reducing the complexity and noise in the frames.

3.3 Background Subtraction

The purpose of background subtraction process is to identify the frontal areas of objects in noiseless video frames. In many videos, the movements of foreground objects occur in the area of interest [16]. Therefore, to track the motions of these objects, it is essential to extract them from the static background and recognize them before undertaking any further processing. This process is performed as follows:

$$\text{Sub Frame} = \{\text{Actual frame} - \text{Background image}\} \tag{2}$$

$$\text{Background subtraction} = \begin{cases} \text{Sub frame if} & \text{sub frame} \geq T \\ 0 \qquad\qquad \text{if} & \text{sub frame} < T_v \end{cases} \tag{3}$$

where T_v denotes the threshold value for the subtraction process, and the pixel intensity between two subsequent frames remains unchanged when it corresponds to static background. Moving objects are recognized by analysing the image pixel variations between the current video frame and the reference frame. By computing these variations, static pixels are rejected, and only those pixels that represent moving foreground objects are retained.

3.4 Feature-Based Multi-object Detection

In the morphological operation and RG segmentation model, the objects in the background-subtracted frames are identified on the basis of target objects associated with next target multiple object tracking. These are represented in Fig. 2. In this segmentation process, some additional features were utilized to recognize multiple objects in video frames. To achieve the desired MO in noiseless and background-subtracted video frames, the images were split into a small number of blocks; then, the random pixels distinguished as foreground pixels were used to create a foreground mask for the video frames.

3.4.1 Morphological Operations

Using morphological operations, the moving MO in frames was detected or selected based on features such as the size, colour, texture, shape, intensity and contrast in the video frames. In general, MO relies primarily on the associated sequences of pixel values rather than on their actual numerical values; hence, the process focuses on pairs of video frames [24]. Using this technique, the output image pixel values are based on pixel values that are similar in the input image and its neighbours.

Colour This feature is represented by a histogram and used to compute a distance measure based on the colour similarity in each frame.

Texture This feature provides information about the spatial arrangement of colours or a preferred image region.

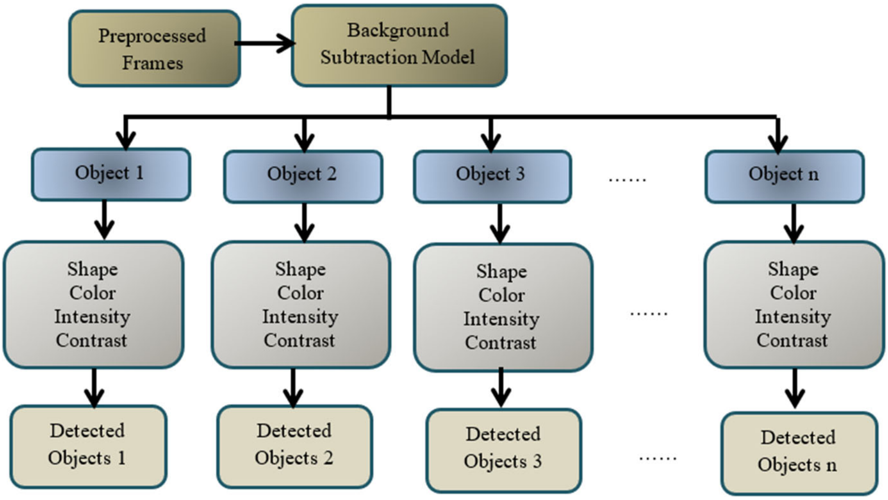


Fig. 2 Block diagram for the MO detection process

Shape This feature refers to the shape of a particular region of the image, not to the entire image.

Intensity This feature is strongly dependent on the characteristics of the images captured by the camera.

Contrast This feature is defined as the value of the contrast intensity between a pixel and its neighbours in the entire frame.

Morphological Process Procedure This operation is represented as a combination of erosion, dilation and some other processes that are explained in terms of the complement of the binary video frames.

(a) Dilation and Erosion of the Extracted Frames

Dilation and erosion are fundamental principals in mathematical morphology. In these procedures, the foreground is compressed, and the frontal area of the frames is expanded. The result increases the boundaries of the foreground pixels; hence, the regions increase in size and holes in the region decrease. The calculation is shown in Eq. (4). Erosion is the process by which objects shrink or become thinner [26]. After dilation and filling, any holes in the object (within limits) shrink, which reasonably limits object disintegration. Then, they are connected to make the boundaries of the objects more coherent for better output. This process is calculated using Eq. (5).

Here, we denote X as the input frame coordinates and Y as the set of structuring element coordinates whose origin is f. Thus, the dilation of X by Y is calculated as follows:

For dilation,

$$X \oplus Y = \left\{ f | (\widehat{Y})_f \cap X \neq \varphi \right\} \tag{4}$$

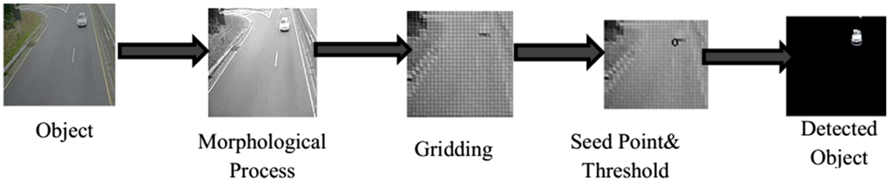


Fig. 3 Block diagram for RG

$$X \pm Y = \left\{ f|(Y)_f \cap X \neq \varphi \right\} \tag{5}$$

If the structure of an object or its parts is minimized to a width equivalent to one, it vanishes. This structuring operation assists in determining how much the image needs to be disintegrated.

(b) Opening and Closing

The organizing components are not utilized for opening and closing areas, which vary according to common morphological channels. Consequently, these two image tasks do not change the objects. Smaller or lighter objects than the threshold are expelled through zone opening. The opening of an image X by the organizing component Y is denoted by $(X \circ Y)$ and calculated as follows:

$$X \circ Y = (X(-)Y)(+)Y \tag{6}$$

Closing involves dilation followed by erosion. The closing of image X by the structuring element Y is indicated by $(X \cdot Y)$ and is described as

$$X \cdot Y = (X(+)Y)(-)Y \tag{7}$$

The primary purpose of this operation is to smooth contours while maintaining the shapes and sizes of objects. By combining these operations, this study was able to achieve better results when performing in-depth edge detection in a frame.

3.4.2 Region Growing (RG)

In MO detection, the most significant technique regarding frame segmentation is RG (see Fig. 3). The considered hypothesis relates to the neighbouring pixels within a region with analogue values. In this technique, when the intensity constraint is scrutinized using the values of neighbouring pixels, some drawbacks arise in the normal region growing method because shading occurs in real frames. Morphological operations yield the best edges for MO, which, in turn, then yields better performances. Thus, segmentation depends on a correct choice of seed point as well as on the threshold values of a particular frame region, and it includes important steps such as grinding, selecting a seed point, fixing thresholds and application to specific areas. The steps are listed and discussed as follows:

(i) **Gridding** Initially, the image is split into several blocks after applying a grid to the image.
(ii) **Seed point selection** In this step, to select a seed point for each noiseless video frame, the histogram technique is used for every pixel in the block. Pixel values lie between 0 and 255; the most frequently occurring pixel value is assigned to be the seed point. This operation is expressed as follows:

$$\text{Frame\_block}(x, y) = f_i(a : (a + s_b) - 1, b : (b + s_b) - 1) \quad \begin{matrix} \text{for}(i = 1 \ldots\ldots 10) : a \\ (j = 1 \ldots\ldots 10) : b \end{matrix} \tag{8}$$

where (x, y) represent the pixel values present in the frame.
(iii) **Identify the region threshold** Here, we set an exact threshold value for the frame block on a pixel basis to segment the MOs in videos.
(iv) **Apply RG to a particular point** After finding the seed point and threshold for the segmentation process, a specific MO is detected and used in the tracking process.

From this analysis (morphological operation and RG methods), the MO can be distinguished; subsequently, the detected objects can more easily be employed in the tracking model.

3.5 Multi-object Tracking by Optimal Filtering

After identifying the objects in frames, a tracking approach was used to locate similar MOs in each frame of single-video segments. The explanation behind the proposed PGA with the Kalman filtering technique is based on the direct application of general object and camera movements. The occurrences of non-inflexible objects, silhouettes or contour portrayals represent excellent paths for both parametric and nonparametric models that can be utilized to indicate movements.

3.5.1 Kalman Filtering

The Kalman filter is a recursive estimator; if the evaluated state from the previous timestep is available, then the current estimation can be expected to provide an estimate of the current state. In general, filtering is a production technique used to recognize the MO among the detected objects. After the outcome of the next measurement is observed, the estimates are updated using a weighted average, with more weight being allotted to estimates that have higher certainty [27]. In the tracking model, two essential parameters in the video frame, i.e. centre (c_1, c_2) and size (i, j), are presented. This concept is clarified by the evaluator that predicts and updates the states of an extensive range of linear processes. Consider a tracking system m_i, the state vector of which indicates the position, velocity and dynamic behaviour of the object and the subscript i indicates the discrete time. The task is to estimate the value of m_i from the measurement u_i. This process has four essential steps, as discussed in the following sections.

(i) Evaluation of the Centres of Objects
The object centres were measured on the basis of the transition matrix and the state vector with Gaussian noise using the probability values of the object state estimation process.

$$m_i = Wm_{i-1} + u_{i-1} \tag{9}$$

$$Q_i = Wq_{i-1}W^T + S \tag{10}$$

where W denotes the transition matrix, m_i indicates the state at time $i-1$, u_{i-1} denotes the Gaussian noise and S denotes the updated vector position.

(ii) Measurement and Time Model
This measurement calculation and updating process were based on the Gaussian noise measurement of each tracked object. The objective is to measure the posterior estimation, which is a linear combination of the prior estimate and the new measurement, u_i:

$$u_i = P * m_i + g_i \tag{11}$$

Updated measurement

$$\mathrm{KF}_i = Q_{\vec{i}} W^T \left(W Q_{\vec{i}} W^T + Z\right)^{-1} \tag{12}$$

$$Q_{\vec{i}} = (1 - \mathrm{KF} * W_i) Q_{\vec{i}} \tag{13}$$

where KF is the Kalman gain computed over the measurement update equations. The time and measurement equations are calculated recursively with the previous posterior estimate to predict a new prior estimate. The recursive behaviour when evaluating the states is one of the features of the Kalman filter [12]. To reduce the $Q_{\vec{i}}$ error rate, an optimization technique was used; the traditional parameter calculation-based optimization methods were derived under the assumption that the function to be resolved is continuous and differentiable.

3.5.2 Probability-Based Grasshopper Algorithm (PGA)

Optimization refers to achieving the best solution (maximized or minimized) in a solution space with regard to some predefined criteria. The modelling of GOA imitates the behaviour of grasshopper swarms in nature to solve optimization problems [12]. A grasshopper position is updated using the probability function, and the solution is based on three essential processes in the Kalman filtering model.

Objective Function To achieve the best filtering performance, three parameters, including the system model parameter, error variance matrix $Q_{\vec{i}}$ and measurement noise u_i, were used. The fitness function is as follows:

$$\text{Fitness} = (|Q_i| + |u_i| + |P_0|) \leq \alpha, \tag{14}$$

where P_0 is the error covariance matrix. To optimize the parameters in the above equation, the GA procedure was utilized. Then, using this procedure, the optimal settings were selected to track the MO in frames.

Updating the Position for PGA Three fundamental components to be used in the simulation are social interaction, the effects of gravitational force, and weather condition wind shifts or wind advection [12]. These components serve to simulate the development of the grasshopper swarm; however, the primary components are the grasshoppers themselves, which are denoted by the following equation:

$$N_i = I_{si} + F_{gi} + A_{wi} \tag{15}$$

Each component in the above equation is discussed in the following section along with the proposed probability function.

Social Interaction (I_{si}) This component starts from the grasshoppers themselves and is represented in Eq. (16).

Gravity Force (F_{gi}) Despite the merits of the function, it cannot be applied between grasshoppers; the gravitational force calculated for each neighbour grasshopper is expressed in Eq. (17).

Wind Advection (A_{wi}) Nymph grasshoppers have no wings; therefore, their movements are highly related to the wind direction, as expressed in Eq. (18).

Probability The mathematical procedure of the interval-based probability function is very similar to that of the classical or traditional probability function. This computation is easily simplified and compared with other range-based probability functions. The calculation is performed based on Eq. (21).

Mathematical Expression

$$I_{si} = \sum_{\substack{j=1 \\ j \neq i}}^{N} l_1\left(e_{ij}\right)\widehat{e}_{ij} \tag{16}$$

$$F_{gi} = -l_2 \hat{e}_g \tag{17}$$

$$A_{wi} = l_3 \hat{e}_w \tag{18}$$

$$N_i =_1 \sum \left(\left(|d_j - d_i|\right)d_j - d_i/e_{ij}\right) - l_2\hat{e}_g + l_3\hat{e} \tag{19}$$

In this case, this mathematical model cannot be utilized straightforwardly to solve the optimization problems because the grasshoppers rapidly find comfortable places

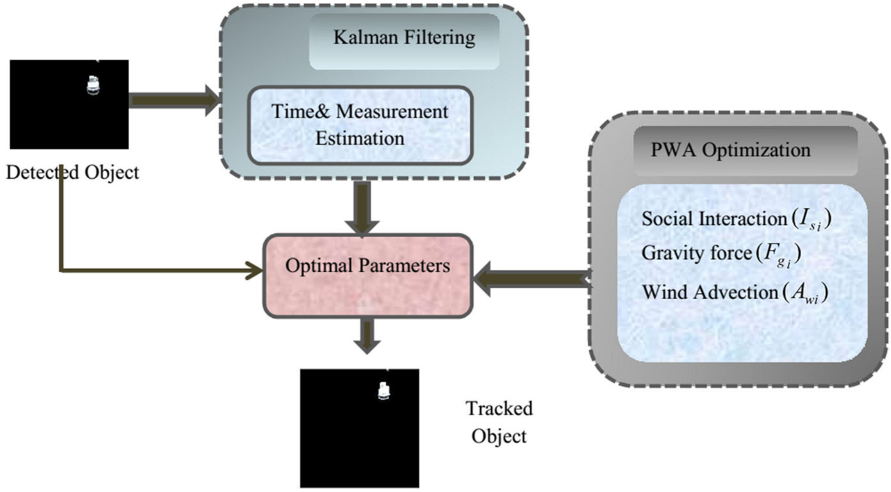


Fig. 4 Optimal filtering for MO tracking

and the swarm will not converge to a predetermined point. Thus, the updated solution is

$$N_i = pr\left(\sum_{\substack{j=1\\ j\neq i}} pr(H_b - L_b/2)\big((|d_j - d_i|)d_j - d_i/e_{ij}\big) - l_2\hat{e}_g + l_3\hat{e}\right) \tag{20}$$

and the probability equation is

$$\Pr_{ij} = \{P(N_i(Q,m) \ \leq\ 0 = \int_{N_i(Q,m)} f(m)\mathrm{d}m \ = (0,q) \tag{21}$$

In the preceding equations, H_b and L_b indicate the upper and lower bound values, respectively, pr is a decreasing coefficient to shrink the comfort, repulsion and attraction areas, and $f(m)$ is a joint probability density function. From this analysis, the optimal parameters were found in the Kalman filter with the minimum error rate for tracking the objects as shown in Fig. 4.

3.5.3 Optimal Parameter-Based Motion Estimation Process

Based on the optimal parameters of the motion model and measurement equation matrix, m_i is an eight-dimensional system state vector expressed as follows:

$$m_i = \big((m * n_{o,i})(l * h_i)(\upsilon_{(m,i),(n,i),(l.i),(h,i)})\big) \tag{22}$$

where $m{*}n_{o,i}$ denote the horizontal and vertical coordinates of the frame, respectively, whereas $l * h_i$ denote the length and width of the frame, respectively. Finally, the term

v indicates the speed. Then, the state and measurement equations of the motion model are defined in the next frame, and the Kalman filter can be used to assess the object location.

3.6 Matching Score for Tracked MO

Each moving object is portrayed by its centroid and tracking window: the horizontal and vertical centroid coordinates and the area of the object in the frame are separate. The current work considers the centroid distance as the Minkowski distance, which is described as follows:

$$\text{Similarity}_{(i,j)} = \left(\frac{\sum_{k=1}^{d} |m_{ik} - n_{ik}|^{1/d}}{\max|m_{ik} - n_{ik}|} \right)^{d} \tag{23}$$

This similarity measure for the MO tracking and detection process explains the variation in the Minkowski distance metric, where 'd' is taken as the point of confinement. This distance can be utilized for factors that are both ordinal and quantitative.

Overall steps for MO detection and tracking system

Step 1: *Video clips are converted into several frames*

Step 2: *The Gaussian filter is applied to remove noise in the generated frames.*

Step 3: *Object detection process. The background is subtracted from the noiseless frames.*

Step 4: *MO detection*

Morphological behaviour is initially evaluated for each object in the frames.

Based on the morphological operation, the RG segmentation approach is used to segment the MO in each frame.

Step 5: *Track the detected MO using optimal Kalman filtering*

Evaluate the time and measurements of the MO

Optimize the three parameters for filtering using PGA

Update the new solution to minimize the error value

Predict social interaction, gravity force and wind advection using a probability function

Based on the optimal parameters, estimate the motions of the tracked objects

To make predictions, the Minkowski distance is used to find similar objects in each frame in the videos

End

(a) PETS09-S2L1 **(b)** CAR

Fig. 5 Database

4 Results and Analysis

The proposed MODT model was implemented in MATLAB 2016a with a system configuration that included an Intel i7 processor and 4 GB of RAM. This analysis was carried out on various databases, the results of the proposed work were examined first, and then, its performance was compared with existing works. For the experiments, the parameters considered were as follows: batch size: 8, learning rate: 0.02, epoch or step size: 10,000, score threshold: 0.7, minimum dimension: 600 and maximum dimension: 1024.

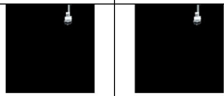
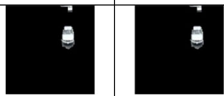
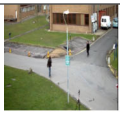
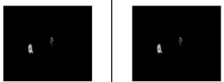
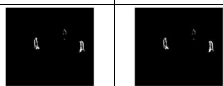
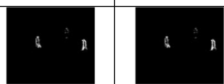
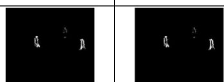
4.1 Database Description

Two videos were considered using the proposed model. These are described in the following section [7]. The initial video is PETS09-S2L1, in which the task is to estimate the crowd density in Regions R1 and R2 in each frame of the sequence. The coordinates of the entry and exit lines are given below, for reference. The second is the CAR video, which has various visible interactions. Several agents are engaged with different simultaneous grouping, grouped and splitting events while partially or wholly occluded (Fig. 5).

Performance metrics

Precision	$\text{Pre} = \frac{\text{TP}}{\text{FP+TP}}$
Recall	$\text{Recall} = \frac{\text{TP}}{\text{FN+TP}}$
F-measure	$F\_M = \frac{2(\text{Pre}\times\text{Recall})}{\text{Pre+Recall}}$
MO detection accuracy	$\text{Detection Accuracy} = \frac{\text{Correctly detected object}}{\text{No. of frame}}$
MO tracking accuracy	$\text{Tracking Accuracy} = \frac{\text{Correctly traacked object}}{\text{No. of frame}}$
Similarity	$\text{Simi} = \frac{\text{TP}}{\text{TP+FP+FN}}$
Error	$\text{Error} = \sum\left(A_i - A_i'\right)$

Table 1 Results for the proposed MODT

Database	Input Frame	Frame Number	Pre-processing	Detected MO	Tracked MO
		12			
		14			
CAR		16			
		18			
		20			
		22			
PETS09-S2L1		5			
		10			
		15			
		20			
		25			
		30			

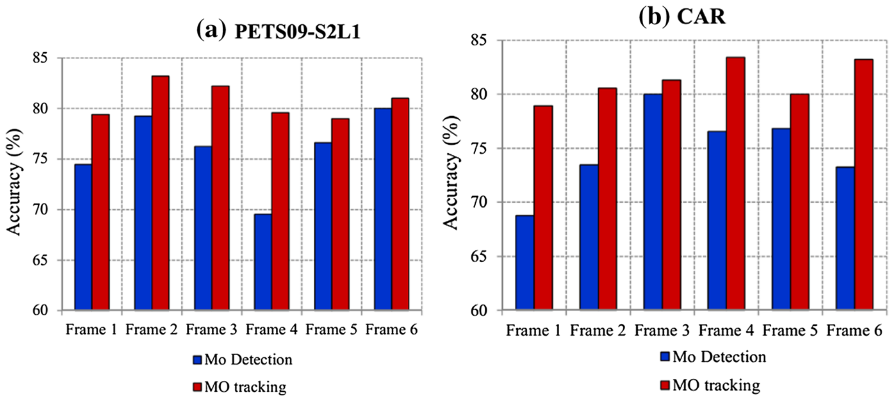


Fig. 6 Frames versus object detection and tracking accuracy

Table 2 Performance results for the proposed MODT

Measures/frames	Precision	Recall	F-measure	Error analysis	Similarity
(i) PETS09-S2L1					
Frame 12	67.67	64.45	92.45	9.45	14.3
Frame 14	64.67	65.9	90	10.56	13.4
Frame 16	67.77	62	95	11.56	15.6
Frame 18	60.88	73.3	93.34	9.75	10.5
Frame 22	62.22	69.4	95.22	12.22	13.52
Frame 24	57.78	59.7	93.6	11.5	12.06
(ii) CAR					
Frame 5	72.4	58.7	93.5	9.44	12.55
Frame 10	52.2	72.4	92.3	11.4	13.45
Frame 15	69.4	67.56	90	9.45	14.53
Frame 20	56.5	62.44	89.67	12.44	17.5
Frame 25	58.99	58.5	88.5	8.56	14.5
Frame 30	65.6	57.5	95.55	7.67	13

Table 1 demonstrates the MO detection, tracking and pre-processed results of the proposed technique, i.e. a morphological operation with RG and optimal Kalman filtering. This table shows selected frames from the analysed results on two databases, PETS09-S2L1 and CAR. For PETS09-S2L1, the considered frames were 12, 14, 16, 18, 20, 22 and 24, and the detected and tracked objects were evaluated; a similar procedure was followed for the CAR videos. The current system counts the number of objects that enter the scene. It additionally recognizes objects that are entering or leaving the scene. Based on the optimal Kalman filter, the detected objects were tracked using linear motion. Although severe occlusion occurred, the objects were not tracked separately.

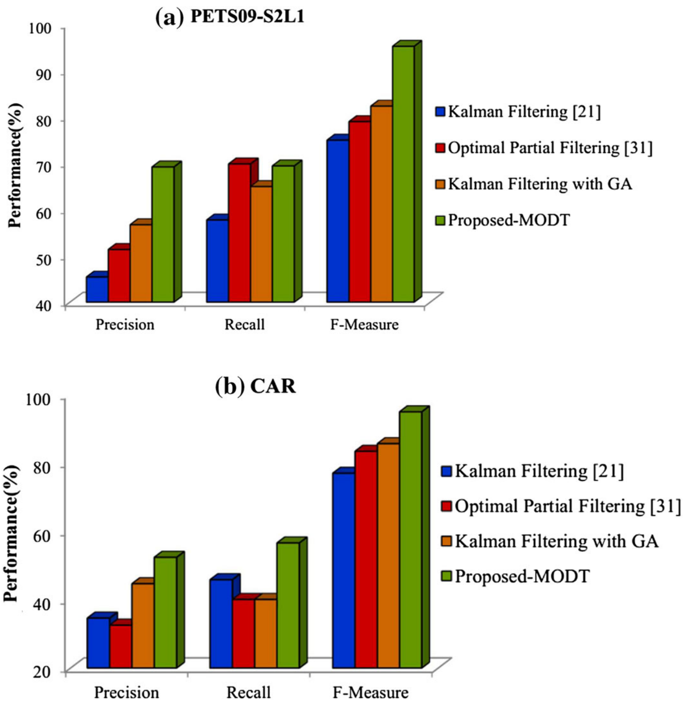


Fig. 7 Performance analysis

Figure 6 demonstrates the MO detection and tracking accuracy of the proposed framework in light of the selected frames in the database, Fig. 6a for PETS09-S2L1 and Fig. 6b for the CAR video. The filtering method productively segmented the video objects with better accuracy considering their multiple highlights. For instance, on the CAR video, frame 3, MODT achieved it's a detection accuracy of 73.232% and a tracking accuracy of 83.22%. Furthermore, on PETS09-S2L1 data frame 6, the detection and tracking accuracies were 76.23 and 82.22%, respectively. As Fig. 6 shows, the object tracking accuracy is improved by the proposed method.

Table 2 lists all the performance results of the proposed MODT strategy for the two databases. In database 1 (PETS09-S2L1), the frame size was 22, and the precision, recall and F-Measure were 62.22%, 69.4% and 95.22%, respectively. Furthermore, the minimum error achieved was 9.75 in frame 18 with the minimum similarity value. Table 2ii lists the CAR video results, i.e. the performance results of the proposed MODT system on database 2. Here, the optimal results for frame 15 are shown, in which the performance rates for precision, recall and F-Measure were 52.22%, 56.47% and 95.22%, respectively, the minimum error in the frame was 9.46, and the minimum similarity was 14.52. The object detection failure might be due to a failure of the tracking part.

Comparative analyses of the proposed MODT with existing systems are shown in Figs. 7 and 8. The comparison techniques considered were Kalman filtering [8], optimal partial filtering with a morphological operation [22] and Kalman filtering

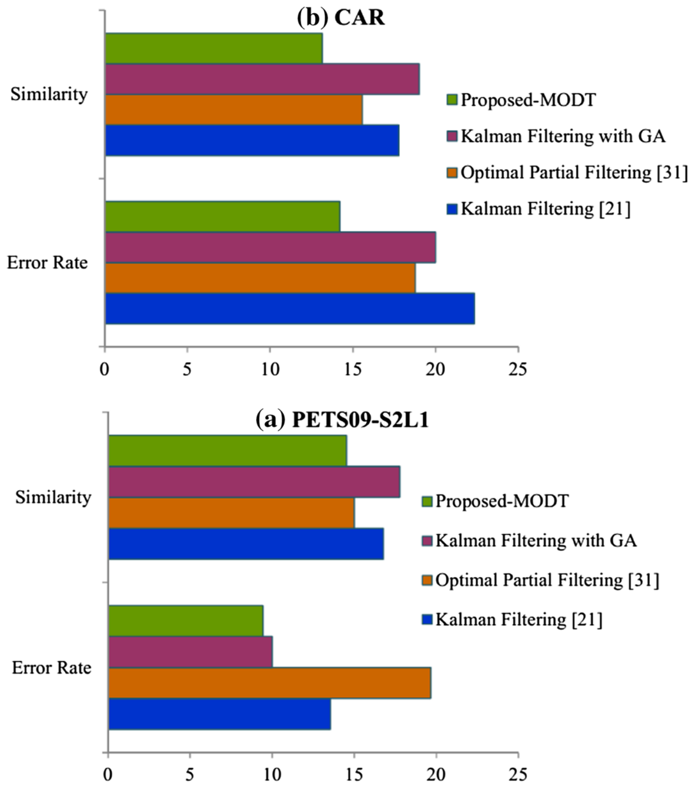


Fig. 8 Error rate and similarity measure

with GA techniques. The precision, recall and F-measure scores are shown in Fig. 7, where the precision of the proposed method is 69.22%. When compared to [8, 22], the minimum differences were 4.22 and 11.2%, respectively. Moreover, measures such as recall and F-measure produced similar results favouring the MODT model. The proposed algorithm was executed on three datasets involving randomly selected or acquired video clips. Figure 8 shows the CAR database with a minimum error rate of 14.22 and a similarity of 13.14 on the CAR video. The proposed MODT approach outperformed all the existing works by providing better person detection and tracking results.

5 Conclusions

In this paper, a new MODT framework was presented to track objects in videos effectively. In the proposed methodology, the optimal Kalman filtering technique was used to track the MO in video frames. The proposed algorithm detected the moving objects in images without noise even under low-level illumination. There may be some concern that the linear motion estimation might be insufficient for objects that move in complicated nonlinear ways. When object movement is moderate, the estimated movements based on the assessed position between successive frames at that point achieve accurate tracking. The experimental results demonstrated that the morphological operation with the RG model is robust and improves the object detection rate (to 76.23%). Correspondingly, the Kalman filtering calculation with the PGA framework achieved the highest tracking rate (86.78%) as well as the smallest error rate and similarity. The main limitation of the proposed work is that the detection rate is quite low and should be further increased. In future work, motion estimation techniques could be applied to the MODT analysis to achieve a better detection rate.

References

1. H.A. Abdelali, F. Essannouni, L. Essannouni, D. Aboutajdine, Algorithm for moving object detection and tracking in video sequence using color feature. in *2014 Second World Conference on Complex Systems (WCCS)* (2014), pp. 690–693
2. R. Assaf, A. Goupil, V. Vrabie, T. Boudier, M. Kacim, Persistent homology for object segmentation in multidimensional grayscale images. Pattern Recogn. Lett. **112**, 277–284 (2018)
3. K. Chen, J. Wang, S. Yang, X. Zhang, Y. Xiong, C.C. Loy, D. Lin, Optimizing video object detection via a scale-time lattice. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 7814–7823
4. A.A. Ewees, M.A. Elaziz, E.H. Houssein, Improved grasshopper optimization algorithm using opposition-based learning. Expert Syst. Appl. **112**, 1–31 (2018)
5. Q. Feng, B. Gao, P. Lu, W.L. Woo, Y. Yang, Y. Fan, X. Qiu, L. Gu, Automatic seeded region growing for thermography debonding detection of CFRP. NDT E Int. **99**, 36–49 (2018)
6. S.M.A. Hasan, K. Ko, Depth edge detection by image-based smoothing and morphological operations. J. Comput. Des. Eng. **3**(3), 191–197 (2016)
7. https://motchallenge.net/vis/PETS09-S2L1
8. E. Hamuda, B. Mc Ginley, M. Glavin, E. Jones, Improved image processing-based crop detection using Kalman filtering and the Hungarian algorithm. Comput. Electron. Agric. **148**, 37–44 (2018)
9. I.A. Iswanto, B. Li, Visual object tracking based on mean-shift and particle-Kalman filter. Proc. Comput. Sci. **116**, 587–595 (2017)
10. S. Kuppuswamy, B. Panchanathan, Similar object detection and tracking in H. 264 compressed video using modified local self similarity descriptor and particle filtering. Int. J. Intell. Eng. Syst. **10**(5), 95–104 (2017)
11. S. Li, C. Zeng, Y. Fu, S. Liu, Optimizing multi-graph learning based salient object detection. Signal Process. Image Commun. **55**, 93–105 (2017)
12. J. Luo, H. Chen, Y. Xu, H. Huang, X. Zhao, An improved grasshopper optimization algorithm with application to financial stress prediction. Appl. Math. Model. **64**, 654–668 (2018)
13. L. Liang-qun, Z. Xi-yang, L. Zong-xiang, X. Wei-xin, Fuzzy logic approach to visual multi-object tracking. Neurocomputing **281**, 139–151 (2018)
14. E. Maekawa, S. Goto, Examination of a tracking and detection method using compressed domain information, in *2013 IEEE Picture Coding Symposium (PCS)* (2013), pp. 141–144
15. T. Mahalingam, M. Subramoniam, A robust single and multiple moving object detection, tracking and classification. Applied Comput. Inf. (2018). https://doi.org/10.1016/j.aci.2018.01.001
16. S.S. Mohamed, N.M. Tahir, R. Adnan, Mohamed, background modelling and background subtraction performance for object detection, in *2010 6th IEEE International Colloquium on Signal Processing & its Applications* (2010), pp. 1–6
17. S.P. Patil, Techniques and methods for detection and tracking of moving object in a video. J. Innov. Res. Comput. Commun. Eng. **4**(5), 8116–8120 (2015)
18. S. Pollnow, N. Pilia, G. Schwaderlapp, A. Loewe, O. Dössel, G. Lenis, An adaptive spatio-temporal Gaussian filter for processing cardiac optical mapping data. Comput. Biol. Med. **102**, 1–39 (2018)

19. K.R. Reddy, K.H. Priya, N. Neelima, Object detection and tracking: a survey, in *2015 IEEE International Conference on Computational Intelligence and Communication Networks (CICN)* (2015), , pp. 418–421
20. K.S. Ray, S. Chakraborty, An Efficient Approach for Object Detection and Tracking of Objects in a Video with Variable Background (2017). arXiv preprint arXiv:1706.02672
21. P.K. Sahoo, P. Kanungo, S. Mishra, A fast valley-based segmentation for detection of slowly moving objects. Signal Image Video Process. **7**, 1–8 (2018)
22. K. Srinivasan, P. Balamurugan, Object detection and movement tracking in H. 264 compressed video based on optimal particle filtering technique. J. Comput. Theor. Nanosci. **13**(11), 8171–8181 (2016)
23. M. Teutsch, W. Krüger, Detection, segmentation, and tracking of moving objects in UAV videos. in *2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance (AVSS)* (2012), pp. 313–318
24. Y. Tan, L. Liu, Q. Liu, J. Wang, X. Ma, H. Ni, Automatic breast DCE-MRI segmentation using compound morphological operations. Fourth IEEE Int. Conf. BMEI **1**, 147–150 (2011)
25. K.K. Verma, P. Kumar, A. Tomar, Analysis of moving object detection and tracking in video surveillance system, in *2015 2nd IEEE International Conference on Computing for Sustainable Global Development (INDIACom)* (2015), pp. 1758–1762
26. J. Xu, L. Zhao, S. Zhang, C. Gong, J. Yang, Multi-task learning for object keypoints detection and classification. Pattern Recognit. Lett. **13**, 1–10 (2017)
27. Z. Zhou, D. Wu, Z. Zhu, Object tracking based on Kalman particle filter with LSSVR. Optik Int. J. Light Electron Opt. **127**(2), 613–619 (2016)
28. D. Zhou, V. Frémont, B. Quost, Y. Dai, H. Li, Moving object detection and segmentation in urban environments from a moving platform. Image Vis. Comput. **68**, 76–87 (2017)