International Journal of Computational Intelligence Systems, Vol. 9, No. 5 (2016) 888-899

Hybrid tracking model for multiple object videos using second derivative based visibility model and tangential weighted spatial tracking model

Felix M. Philip[1*], Rajeswari Mukesh[2]

[1]Research Scholar,
Department of Electronics and Communication Engineering,
Hindustan University, Chennai.
mfelixphilip@gmail.com

[2]Professor and HOD,
Department of Computer Science and Engineering,
Hindustan University, Chennai.

Received 30 January 2016

Accepted 16 May 2016

Abstract

In the area of video surveillance, tracking model for multiple object video is still a challenging task since the objects are usually affected with inter-object occlusion, object confusion, different posing, environment with heavy clutter, small size of objects, similar appearance among objects, and interaction among the multiple objects In order to alleviate these challenges, literature presents different tracking models using spatial and visual information. Accordingly, in this paper, we have developed a hybrid tracking model for tracking the multiple objects from the videos using twofold architecture. At first, visibility model for tracking is proposed based on the second derivative model, which considers the second derivative function to predict the objects. Secondly, a spatial tracking model is proposed using tangential weighted function. Finally, these two contributions are effectively included in the hybrid tracking model for multiple object tracking and the performance analysis is carried out using two videos from UCSD dataset. From the results, we proved that the proposed hybrid tracking model achieves the Multiple Object Tracking Precision (MOTP) of 99% than the other exiting tracking models.

Keywords: hybrid tracking, visibility model, spatial tracking, tangential weighed, second derivative.

1. Introduction

In the research area of video analysis [1,2], automatic detection and tracking of objects in video data is one of the challenging task in video surveillance, behaviour modelling, security applications and traffic control. In essence, many tracking algorithms are used to estimate the individual objects derived from the previously computed direction of objects. Although, multi-object tracking is the most important research area in computer vision, in which many algorithms given in Ref.3-8 are proposed to solve the general tracking problem without considering the challenges of multi-object tracking. Most of the algorithms given in Ref.3-6 have focused on the common crisis of tracking, without particularly addressing the challenges of a multiple objects. Traditional tracking methods classically assume a static background or easily discernible moving objects, and, as a result, are limited to scenes with relatively few constituents [9,10].

In the conventional methods, the process of multi-object tracking can be carried out in two ways, such as data association and state estimation. The task of multi-object tracking is mainly partitioned to locate multiple objects, yielding their individual trajectories and maintaining their identities given an input video [11].

However, multi-object visual tracking is a challenging task for maintaining several problems such as, inter-object occlusion, object confusion, different posing, environment with heavy clutter, small size of objects, similar appearance among objects, and interaction among the multiple objects. To address the aforementioned problems, several filtering methods are used in multi-object tracking. For instance, kalman filtering is one of the efficient methods to address the multi-object tracking problem[12] when the number of objects remains small. It is a linear quadratic estimation algorithm that performs the series of measurements observed over time.

Even though the kalman filtering has some major advantages in spatial tracking, it has some serious drawbacks like, i) it can be used only for linear state transitions, ii) It is applicable, when the number of objects remains small. Moreover, the shift algorithm is another one filtering method to overcome the multi-object tracking problem. Then, particle filtering can address the limitations of aforementioned kalman Filter [13]. Particle filter techniques provide a well-established method for generating samples from the required distribution without considering the state-space model. However, the aforementioned particle filter does not perform well, when applied to very high-dimensional systems. This paper proposes a new visual tracking approach and spatial tracking by reflecting some aspects of spatial selective attention.

This paper aims to design and develop a system for multi object tracking using hybrid tracking model to handle the above discussed challenges. Accordingly, a hybrid tracking model is proposed for multiple object videos using second derivative based visibility model and tangential weighted spatial tracking model. The first step is to detect the objects presented in the first frame through object segmentation algorithm, where, K-means clustering algorithm is utilized. Once the objects are clearly segmented, the tracking is performed using the hybrid tracking algorithm. In the hybrid tracking, the spatial tracking and visual tracking methods are effectively combined using weighted average formulae. In the visual tracking mechanism, the similarity of the objects is compared among the frames and the objects are tracked using the second derivative based visibility model. In the spatial tracking method, the location of spatial points is predicted using the proposed tangential weighted spatial tracking model. Then, these two tracked outputs are combined to obtain the final tracked results. The rest of this paper is organised as follows: Section 2 reviews several existing approaches for tracking multi-objects video. Section 3 presents the motivation behind the proposed approach. Then, the proposed methodology of hybrid tracking model is described in section 4. Extensive experimental results on publicly available UCSD datasets are given in Section 5. Finally, section 6 concludes this paper.

2. Literature review

In this section, we review the summary of the several multi-object tracking methods with their importance. Many researchers deal with the multi-object tracking in video using different methods. Early work by J. Berclaz *et al.*[14] have developed a method for multi-object tracking with K-shortest path optimization to track the object, which is difficult to handle intersecting trajectories. Additional work by X.Zhou *et al.*[15] dealt with the improved spatial colour appearance with interferences using gaussian mixture probability hypothesis density (GM-PHD) filter, in which computational cost is high. Then, to improve the confidence of sampling and perform the iteration effectively, X. H.Xia *et al.*[10] have implemented the Markov chain Monte Carlo-based multi-object visual tracking method. W. Hu *et al.*[16] addressed the multi-feature joint sparse representation to automatically focus on the visible parts of an occluded object, in which many parameters need to set and the semantic corrections between the different features are not modelled.

I. Ali and M. N. Dailey[17] have developed the confirmation-by-classification method to detect and track multiple humans in high-density crowds in the presence of extreme occlusion. To capture the Interdependence of multiple influence factors, X. Liu *et al.*[18] have implemented the discriminative structure prediction model. To capture both the global and local spatial layout information, Log-Euclidean Riemannian Subspace and Block-Division approach was developed by W. Hu *et al.*[19], in which detections was not flexible for high level environment change. In this paper, we present an approach for hybrid tracking of multi-object video using second derivative based visibility model and tangential weighted spatial tracking model. Table 1 shows the summary of the related works.

Table 1. Literature review

Author	Approach	Advantages	Problem identified
J.Berclaz *et al.*[14]	K-Shortest Path Optimization	far simpler formally and algorithmically	Difficult to handle intersecting trajectories
X.Zhou *et al.*[15]	Gaussian mixture probability hypothesis density (GM-PHD) filter	improved spatial colour appearance with interferences	large computational cost once a large number of noises are tracked
X. H.Xia *et al.*[10]	Markov chain Monte Carlo-based multi-object visual tracking	improve the confidence of sampling and perform the iteration effectively	weaknesses in the colour model and very heavy occlusion
W.Hu *et al.* [16]	multi-feature joint sparse representation	Automatically focuses on the visible parts of an occluded object	many parameters need to be set and the semantic corrections between the different features are not modelled
W. Hu *et al.*[19]	Log-Euclidean Riemannian Subspace and Block-Division Appearance Model	captures both the global and local spatial layout information	detection is not flexible for high level environment change
I.Ali, M.N. Dailey[17]	confirmation-by-classification method	It detects and tracks multiple humans in high-density crowds in the presence of extreme occlusion	detector's output is unreliable
X. Liu *et al.*[18]	discriminative structure prediction model	It can capture the interdependence of multiple influence factors	Risk of labelling error

3. Motivation behind the approach

3.1 Problem definition

When public anxiety about crime and terrorist activity increases, the importance of security is on the rise and video surveillance systems are increasingly well-known tools for monitoring, management, and law enforcement in public areas. Since it is complicated for human operators to monitor surveillance cameras continuously, there is strong attention in automated analysis of video surveillance data. Some of the important problems include pedestrian tracking, behaviour understanding, irregularity detection, and unattended personal belongings detection.

- The task of multi-object tracking is to estimate the state and the number of objects or group of objects from sequence of noisy observations.
- Tracking of multiple objects are extremely challenging for the researchers because overlapping detections and dynamic occlusions cause complicated issues when the tracking to be done through visual information.
- The real time videos containing the walkways were variable, ranging from sparse to very crowd. This scenario is difficult to track down. Also, the movement of every object is variable which is difficult to predict even if the system utilizes the spatial prediction methods.
- In multi object tracking, existing works fail to do tracking due to variable object size, colour and deformation of objects. Also, it finds difficulty in detecting the overlapping objects through arbitrary movement.
- In existing works given in Ref.3-6, 14, 15, 17, 18, multi object tracking is done either through visual tracking method or spatial tracking methods but the uncertainty of object movement and occlusion caused severe performance degradation in both the methods.

4. Proposed Methodology: Hybrid tracking model for multiple object videos

The block diagram of the proposed method is shown in Fig. 1.The steps involved in our proposed hybrid tracking model as follows:

i) Read the input video[20]
ii) Extract the object area from two dimensional matrix using K-means clustering algorithm with respect of texture and neighbour pixels.
iii) Apply the proposed second derivative based visible tracking model in the extracted frame to track the object.
iv) Again, apply the proposed tangential weighted tracking model for same extracted frame.
v) Combine the both visibility model and spatial tracking model for input frame
vi) Finally, we get the tracked output.

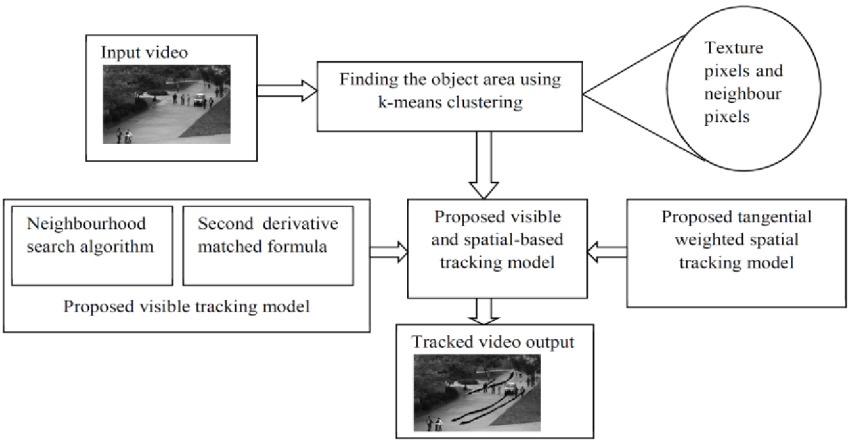


Fig. 1. Block diagram of the proposed method

4.1 Finding the objects from key frame using k means clustering

In our work, the objective is to track the multiple objects presented in the video. So, we should identify the objects' presence in the video initially. Suppose, if we want to track four objects from the input video, we should identify four objects. To identify four objects from the video, we have to apply K-means clustering by giving k value as four. Here, K-means clustering identify the location of four objects from the video. The value of k should be given by the user based on the number of objects to be tracked. Here, K value does not affect the tracking accuracy because K value here is about to detect the objects. At first, the input video is read and the key frame is extracted. From the key frame, the objects presented within the frame should be detected. The detection of area of the objects is performed using K-means clustering which segment the object area from the key frame. K-means clustering algorithm[21-23] is an important algorithm to solve the clustering problem. While tracking the multimodal object in video by reducing the occlusions as well as interference, K-means clustering algorithm is favoured in this paper. Before applying this K-means algorithm, at first, we should extract the features from the frame using Local Binary Pattern (LBP) modal[24].

Feature construction: For constructing the frame feature, the input video V is read and extract the frame V_i, which is defined as the i -th frame of the video shown as follows.

$$V = \{V_i; \quad 1 \le i \le n\} \tag{1}$$

Each and every frame, the pixel vales can be represented as shown below.

$$V = \{v_i^{\,j,k}; \quad 1 \le j \le N; 1 \le k \le M\} \tag{2}$$

Where, j is a column of pixels that can be varied from 1 to N, and k is a row of pixels varying from 1 to M. The performance of object detection using only the pixels of input video image is not accurately predictable. In order to improve the detection performance, LBP[25] feature is used. The aforementioned LBP method is an efficient texture operator which describes the small scale appearance of input image, in which image pixels are characterized by thresholding [26] of each pixel against the centre pixel that shows the result in the form of binary values. Here, i_c is a centre pixel value of the eight surrounding pixels. While representing the binary number, if the neighbourhood pixel is greater than the centre pixel, we represent that binary number as one. If the neighbourhood pixel is less than the centre pixel, representation of binary number is zero.

$$S(x) = \begin{cases} 1 & if\ x > 0 \\ 0 & if\ x < 0 \end{cases} \tag{3}$$

Then, the corresponding pixel value is estimated by converting the aforementioned binary values into decimal number.

$$v_i^{\,L}(j,k) = \sum_{a=0}^{7} S(i_a - i_c)2^a \tag{4}$$

Where, i_c is a centre pixel (j,k) value of the eight surrounding pixels. Assume that the location of i_c is (x,y). Then, the location of i_0 to i_7 are (x+1, y), (x-1,

y+1), (x-1, y), (x-1, y-1), (x, y-1), (x+1, y-1), (x, y+1), (x+1, y+1). Then, the new frame is generated using LBP feature, in which the range of pixel values is same as that of the input frame, which is represented as follows:

$$V = \left\{ v_i^{\ L}(j,k); \quad 1 \le j \le N;\ 1 \le k \le M \right\} \tag{5}$$

Finally, the two dimensional matrix is formed by combining both the original frame and LBP generated frame is shown below:

$$f_{k,l} \in V_i, V_i^L \tag{6}$$

$$F = \left\{ f_{k,l} \quad 1 \le k \le M \times N;\ 1 \le l \le 2 \right\} \tag{7}$$

Where, V_i is an original frame, V_i^L is a LBP generated frame and $f_{k,l}$ is represented as the combination of both original frame and generated LBP frame.

K-Means clustering: After obtaining the two dimensional matrix, the objects can be extracted using K-means clustering which considers texture and colour feature to segment the objects. Then, we can group the data set the aforementioned two dimensional matrix using K-means clustering algorithm. Based on the inherent distant from each other, K-means clustering algorithm[22] groups the input objects into multiple groups based on the features into K number of groups. Where, K is represented as any positive integer. The grouping done by K-means clustering algorithm minimizes the sum of squares of distance between object and the corresponding cluster centre. This is the type of partitioned clustering algorithm which determines all clusters at once. The steps involved in k means clustering algorithm as follows:

i) Initialise the cluster centres in two dimensional matrix based on the range of needed clusters C which can be represented as follows:

$$R = \left\{ R_{i,j};\ \quad 1 \le i \le C; 1 \le j \le 2 \right\} \tag{8}$$

ii) Randomly select the cluster centre $R_{i,j}$ in the two dimensional vector.

iii) We should do the matching between every $f_{k,l}$ with R by euclidean distance (ED), which shows the similarity of calculated two elements and influence the shape of the clusters.

$$ED(f_{k,l}, R) = \sqrt{\sum_{i=1}^{n} \left(f_{k,l} - R \right)^2}; 0 \le K \le m \tag{9}$$

Where, $f_{k,l}$ is a two dimensional matrix with both original frame and LBP generated frame $ED(f_{k,l}, R)$ is euclidean distance between $f_{k,l}$ and R.

iv) Find the minimum distance of $f_{k,l}$ with R, represented as $R_{new}(t)$

$$R_{new}(t) = \frac{1}{|C_1|} \sum_{f \in C_i} f_i^{k,l} \qquad \forall i \tag{10}$$

v) Go to step (ii), until $R_{new}(t+1)$ is equal to $R_{new}(t)$

vi) Then, the number of objects selected based on the cluster value C, shown as:

$$R_1, R_2, ...R_C$$

4.2 Visual tracking using the proposed second derivative model and neighbourhood search

Visual tracking[27] is a fundamental method for analysing video motion processing, data mining, visual surveillance, vision based control, human computer interactions. Object appearance model of visual tracking model is based on region colour histograms, kernel density estimates, gaussian mixture model etc. However, multi-object visual tracking has been in still research because of couple of factors, such as inter-acting each other posing inter-object occlusion, object confusion and low probability of detection. In order to overcome these aforementioned drawbacks in visual tracking, in this project we propose visual tracking algorithm based second derivative model and neighbourhood search, namely SDVM (Second derivative visual model).

Neighbourhood search algorithm: Neighbourhood search algorithm[28] is one, in which the object can be tracked as follows:

i. Fixing the fixed reference point R_i^t in key frame using template function.

$$R_i^t \in f(x, y) \tag{11}$$

Where, R_i^t is represented as the reference point of the key frame.

ii. Extraction of the reference point in new frame based on the key frame reference point.

$$R_i^{t+1}(x, y) = f\left(r^k\right) \tag{12}$$

Where, R_i^{t+1} is a extracted reference point, and $f\left(r^k\right)$is a function to generate sub image.

iii. Find the euclidean distance between the key frame and extracted new frame pixels.

$$ED(R_i^t, R_i^{t+1}) = \sqrt{\sum_{i=1}^{n}\left(R_i^t - R_i^{t+1}(x,y)\right)^2} \; ; 0 \le K \le m \tag{13}$$

iv. Increase or decrease the reference point by unity to find the new location.

$$r_i^{t+1}(x,y) = \{x \pm i, y \pm j\} \tag{14}$$

Where, (i, j) denotes the increasing or decreasing reference point

v. Extract the new region from the input image and find the ED between the input image and the extracted region.

vi. Perform the step 4 and 5 until it reaches the predefined search threshold.

vii. After completing step 5, the reference point which has minimum ED is taken as the final region of detected image.

$$R_i^{t+1}(x,y) = \arg\left(\underset{K=1}{Min}\, P_K\right) \tag{15}$$

Proposed matching formulae: The aforementioned neighbourhood search algorithm is unsuccessful to track the image if there is an interaction between multiple objects and frequent occlusions. In order to overcome these drawbacks, the second derivative model with neighbourhood search algorithm is proposed in this work. The next spatial location of the object using second derivative-based visible pixels as follows:

$$P_K = \alpha \times ED\left(R_i^t, R_i^{t+1}(x,y)\right) + (1-\alpha)\times ED\left(\frac{\partial^2 R_i^t}{\partial x^2}, \frac{\partial^2 R_i^{t+1}(x,y)}{\partial x^2}\right) + (2-\alpha) ED\left(\frac{\partial^2 R_i^t}{\partial y^2}, \frac{\partial^2 R_i^{t+1}(x,y)}{\partial y^2}\right) \tag{16}$$

$$\frac{\partial^2 R_i^t}{\partial x^2} = \frac{\partial^2 f(x,y)}{\partial x^2} = W_{xx}(t) \; ; \; \frac{\partial^2 R_i^t}{\partial y^2} = \frac{\partial^2 f(x,y)}{\partial y^2} = W_{yy}(t) \tag{17}$$

$$\frac{\partial^2 R_i^{t+1}(x,y)}{\partial x^2} = \frac{\partial^2 f(x,y)}{\partial x^2} = W_{xx}(t+1) \; ; \quad \frac{\partial^2 R_i^{t+1}(x,y)}{\partial y^2} = \frac{\partial^2 f(x,y)}{\partial y^2} = W_{yy}(t+1) \tag{18}$$

Where, R_i^t is a reference point in key frame, $R_i^{t+1}(x,y)$ is a reference point of the next frame, $W_{xx}(t)$ and $W_{xx}(t+1)$ indicates the second derivative of key frame and next frame based on x direction, $W_{yy}(t)$ and $W_{yy}(t+1)$ indicates the second derivative of key frame and next frame based on y direction. ED is an euclidean distance between key frame and next frame of the reference point. This proposed matching formula utilizes second order derivative like, $W_{xx}(t)$ and $W_{yy}(t)$. The formulae does not include $W_{xy}(t)$ and $W_{yx}(t)$ for similarity computation. The important reason is that the second derivatives of $W_{xx}(t)$ and $W_{yy}(t)$ can demolish the noisy information as much when compared with $W_{xy}(t)$ and $W_{yx}(t)$.

4.3 Spatial tracking using proposed tangent weighted spatial tracking model

In this section, we describe the spatial tracking method [9], which predicts the next spatial location of the object using the tangential weighted spatial model (TWSM). Here, the object is represented in the form of x,y coordinate in key frame. According to the direction of changing x,y coordinates, the corresponding object of the next frame can be predicted. Then, the location of the same object in next frame is predicted according to the difference between the coordinates of the objects in key frames with next frame. By doing this tracking process using spatial method, the object in the frame can be accurately tracked.

In this model, the centre point value of the objects is selected for further calculation. The selected centre point of the cluster can be represented as: $r_i^t .....r_C^t$

Where, r_i^t is a i-th object in the t-th frame. Using the aforementioned centre point of the object r_i^t, we can do the spatial tracking method using the following formulae which is the Exponential Weighted Moving Average (EWMA) model[29].

$$tr_i^{t+1} = \alpha r_i^t + (1-\alpha) tr_i^t \tag{19}$$

Where, tr_i^{t+1} is an i-th object in the $t+1$-th frame. Then, the object r_i^t is belongs to (i, j) Where, α is the smoothing factor which is computed based on equation 22. It is a function of exponential, in which each pixel intensity value of the output image is equal to the basis value raised to the value of the corresponding pixel values in the input image is represented as follows:

$$\alpha = \left(1 - \exp^{\Delta T / \tau}\right) \tag{20}$$

The smoothing factor contains two parameters called, time constant and sampling time interval. Here, we have given the time constant as two and sampling time interval as 1. These two values are fixed based on trial and error computation. Here, τ is a Time constant, α is a smoothing factor and ΔT is sampling time interval. While doing this spatial tracking using exponential function, several problems may occur during tracking process. More importantly, the major problems that can be addressed by the hyperbolic tangential function are, i) it can easily represent the transition from one state to another , ii) when estimating the time series data, it can susceptible to rare events, iii) it can handle fluctuations optimally. Therefore, to the best of our knowledge, in order to improve the tracking performance, this is the first ever research work where, tangent weighted method is preferred instead of exponential function.

Tangential weighted function: Commonly, hyperbolic functions have many useful applications in engineering, such as electrical transportation, superstructure, and aerospace. In this section, to improve the performance of spatial filtering, tangential weighed function is preferred instead of exponential function. The hyperbolic tangent is a function that can easily represent the transition from one state to another. The aforementioned tangential weighted function operates element wise on arrays, which has the advantage of non-linear and linear function with high training speed. The prediction of next location using the proposed tangential function is as shown as:

$$tr_i^{t+1} = \alpha r_i^t + (1-\alpha) tr_i^t \tag{21}$$

$$\alpha = 2(1 - \tanh + C) \tag{22}$$

Where, r_i^{t+1} is a i-th object of $t+1$–th frame and α is a smoothing factor, $\tanh$ is the hyperbolic tangential function and C is the weighted constant. It varies in the range of 0 to 1. The weighted constant is important to avoid the fluctuation in tracking. The value fixed here is 0.75 which is found out using trial and error method.

4.4 Integration of visual and spatial tracking model

In this section, we integrate both visual and spatial tracking model, which demonstrates the better tracking performance. Moreover, second derivative model used in visual tracking and tangential weighed method using in spatial tracking used to overcome the multi-object tracking limitations such as, frequent occlusions, similar appearance of objects, and interaction between multiple objects. The results of the visibility model R_i^{t+1} and spatial tracking model tr_i^{t+1} are integrated to obtain the final tracking T_i^{t+1} results shown as:

$$T_i^{t+1} = \left\lfloor C\left\{R_i^{t+1}\right\} + \left\{tr_i^{t+1}\right\} \right\rfloor \tag{23}$$

Where, R_i^{t+1} is a reference point selected in the visual representation, $C\left\{R_i^{t+1}\right\}$ is a centre point of the object which we have selected from the reference area, and $\left\{tr_i^{t+1}\right\}$ is the centre point of the object in spatial tracking. Fig. 2 shows the algorithmic representation of multi-object tracking videos.

	Input: Multi-object video, V
	Output: Tracked object, T
	Parameters: R_i^{t+1} →selected reference point in the visual representation
	tr_i^{t+1} → centre point of the object in spatial tracking
	T_i^{t+1} →Final tracking results
	Procedure
1	**Begin**
2	Read the key frame from the video, V_i
3	Extract the LBP from key frame
4	Generate two dimensional feature matrixes $f_{k,l}$
5	Detect the objects from the video using k-means clustering algorithm
6	**For** all the objects
7	Track the object using the formula of P_K in visual representation.
8	Find the cluster centre of tracked object $C\left\{R_i^{t+1}\right\}$
9	Track the object using spatial tracking method based on the formulae of tr_i^{t+1}
10	Integrate $T_i^{t+1} = \left\lfloor C\left\{R_i^{t+1}\right\} + \left\{tr_i^{t+1}\right\} \right\rfloor$
11	**End** for
12	**End**

Fig. 2. Algorithmic description

5. Results and discussion

This section presents the experimentation of the proposed hybrid model of tracking with two different videos and a detailed comparative analysis with three different metrics.

5.1 Experimental set up

The proposed hybrid tracking model is implemented using Matlab 8.3 (R2014a) with a system configuration of 4GB RAM Intel processor and 64 bit OS.

Dataset description: We have used the two different videos from UCSD datasets[20]. The first video contains movement of multiple persons with bicycle and the second video contains the movement of multiple persons with truck. The dataset is utilized from Ref.20 which have the ground truth information so it would be easy to evaluate the proposed hybrid model.

Methods employed for experimentation: Here, we consider four works such as, second derivative visual method (SDVM), tangential weighed spatial model (TWSM), GM-PHD filtering, proposed hybrid model for the experimentation. The proposed hybrid model is newly proposed in this paper. TWSM is also newly proposed here for spatial tracking. SDVM is also newly proposed for visual tracking. GM-PHD filter[15] is the existing works considered here for experimentation.

Evaluation metrics: Three different metrics are utilized here to evaluate the system, 1) tracking number, 2) tracking distance, iii) MOTP. The first metrics is used to ensure that the tracking algorithm is more accurate with starting and ending frame of the objects and its presence over with the true number of frames. The second and third parameters are utilized to find the deviation of 2D location from the ground truth versus the tracked results. Tracking distance is the euclidean distance measurement between the ground truth results with the tracked output. The euclidean distance is computed based on the location of pixels in the ground truth and tracked results. MOTP (Multiple Object Tracking Precision)[30] can be spontaneously expressing the tracking precision, which denotes exact positions of estimated persons, which is represented as follows:

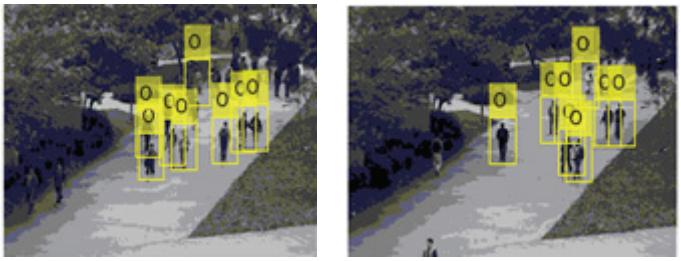
$$MOTP = \frac{\sum_{i,t} D_t^i}{\sum_t m_t} \tag{24}$$

Where, m_t is represented as the number of matches found at time t and D_t^i is the distance between the objects and matching hypothesis. It is the total error in estimated position for matched object-hypothesis pairs over all frames, averaged by the total number of matches made.

5.2 Experimental results

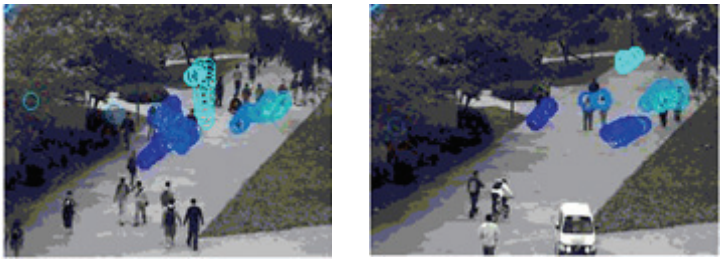
This section shows the experimental results of detected objects. Fig. 3(a) shows the sample video frame, in which several pedestrians are walking. Walking pedestrians are detected using k-means clustering algorithm. Then, the detected pedestrians are represented in yellow mark. After applying the clustering algorithm, fig. 3(a) represents the eight detected objects from the first frame of the video. The similar type of behaviour can be seen in fig. 3(b).

Fig. 4(a) and fig. 4(b) shows the tracked objects in video 1 and video 2. After detecting the object, several steps are used to track the objects. Nearest neighbourhood algorithm is one of the method to track the objects, using the aforementioned detected points. Fig. 4(a) and fig. 4(b) shows the object tracking performance of the last frame of the video.



Video 1 detected objects Video 2 detected objects

Fig. 3. a) Detected objects in video 1, b) detected objects in video 2

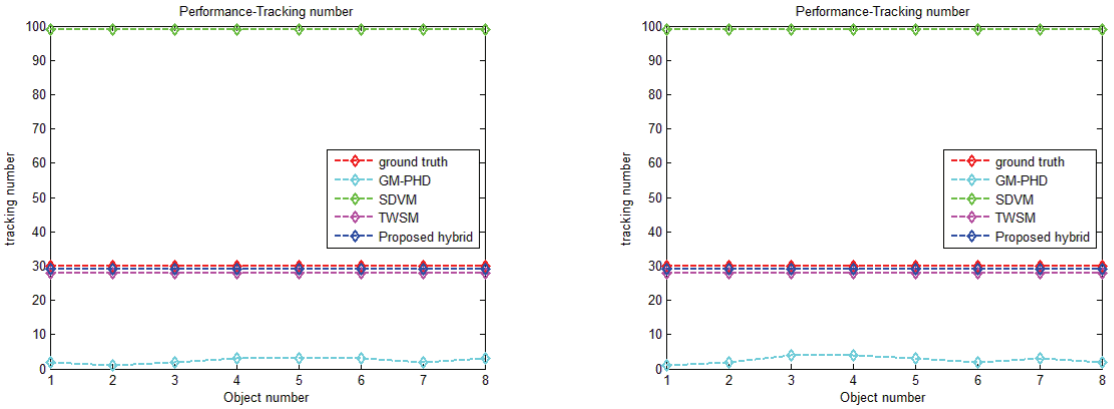


Video 1 tracked object Video 2 tracked object

Fig. 4. a) Tracked results of video 1, b) tracked results of video 2

5.3 Comparative analysis using tracking number

This section presents the analysis of four different methods such as, second derivative visual method (SDVM), tangential weighed spatial model (TWSM), GM-PHD filtering and proposed method using tracking number. Fig. 5(a) represents the analysis of the tracking number in the first dataset. Here, eight objects are considered for the analysis of tracking number, in which ground truth (red) denotes the accurate presence of objects in total frames of the input videos. Based on the ground truth information, every object should present only in 30 frames. From the fig. 5(a), the SDVM model shows tracking number of 99 which means that the every objects are presented in 99 frames instead of 30 frames. GM-PHD filtering shows the tracking number of objects is one to three, which is not matched with ground truth. Consequently, tangential tracking model shows the tracking number of 28 instead of 30 but the proposed hybrid tracking model shows the tracking number as 29 which is better than any other methods taken for comparison. Fig. 5b shows the analysis of the tracking number in the second dataset. While analysing the dataset 2, every objects should present in 30 frames founded on the ground truth representation. However, fig. 5b shows the tracking number for SDVM, TWSM, GM-PHD filtering methods are 99 frames, 28 frames and six frames respectively. Accordingly, our proposed method shows the tracking number as 29 which is better than the other methods of SDVM, TWSM, GM-PHD filtering.



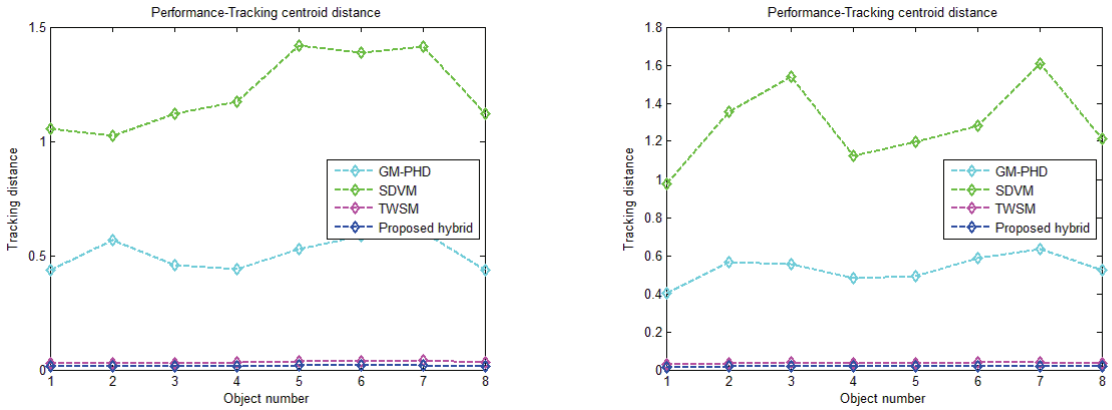
Dataset 1 tracking number analysis Dataset 2 tracking number analysis

Fig. 5. Analysis using tracking number, a) dataset 1, b) dataset 2

5.4 Comparative analysis using tracking distance

In this section, fig. 6 shows the comparative analysis of four algorithms using tracking distance. Tracking distance is computed by finding the distance between the original paths of the object with tracked paths by the different algorithms. If the tracking distance shows minimum value, then the corresponding method is chosen as the better algorithm. While analysing the fig. 6a, we can understand that the minimum tracking

distance is achieved when the proposed method is used for every objects. For every eight objects, the tracking distance performance of the proposed method is in the order of 0.0145, 0.0132, 0.0151, 0.0160, 0.0187, 0.0196, 0.0203, and 0.0155 respectively. Further, by analysing this figure, SDVM and GM-PHD filtering methods shows the maximum tracking distance in the order of 1.0557 and 0.4282 for first object. TWSM method shows the tracking distance as 0.0308, which is better than both SDVM and GM-PHD filtering performance. Fig. 6b shows the performance of second dataset using tracking distance, which indicates that the SDVM method generates maximum tracking distance as 1.2136. Then, the tracking distance of particle filtering lies in between the second derivative-based method and the proposed method that is shown as 0.5315. From the fig. 6, we can conclude that, the proposed method reaches the minimum tracking distance between the original objects and tracked objects for achieving better tracking performance.



Dataset 1 tracking distance analysis

Dataset 2 tracking distance analysis

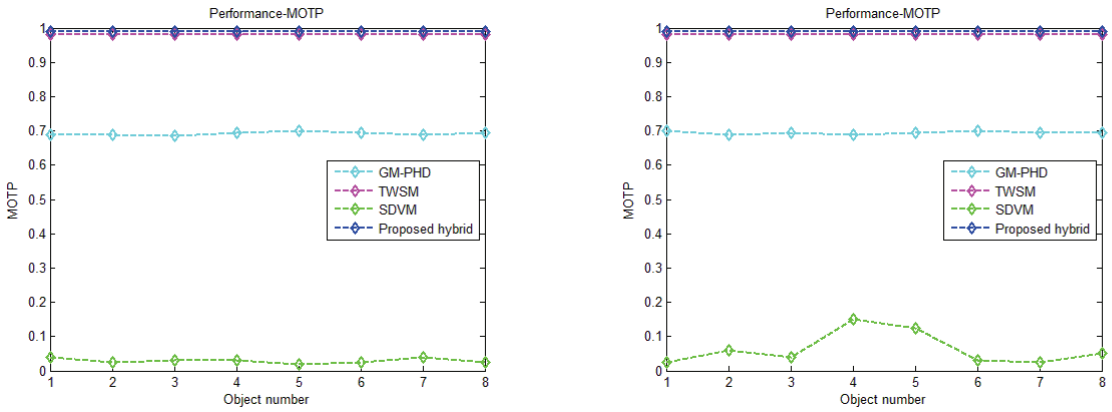
Fig. 6. Analysis using tracking distance, a) dataset 1, b) dataset 2

5.5 Comparative analysis using MOTP

Here, the performance is compared using MOTP (multi object tracking precision) for various filtering models such as, SDVM, TWSM, GM-PHD filtering and our proposed method. Fig. 7(a) shows the MOTP analysis of different filters, in which our proposed method generates better tracking performance of multiple objects. Moreover, comparing with TWSM and GM-PHD filtering, the proposed method predicts the multi-object tracking performance effectively as 99%. When analysing the fig. 7a, the SDVM filtering is not suitable for multi-object tracking, which shows the least tracking performance as 40%.Then the MOTP performance of TWSM and GM-PHD filtering is shown in fig. 7a as 98% and 69%. Fig. 7b shows the MOTP analysis of second dataset using various filters, which indicates the SDVM filtering produces the least multi-object tracking performance as 25% compared to other filtering methods of TWSM, GM-PHD filtering. Then, the MOTP performance of TWSM and GM-PHD filtering is shown as 98% and 68% for second object in fig. 7b. From these results, we can see that our proposed method outperforms the other methods in tracking.

6. Conclusion

In this paper, we presented a hybrid tracking model to track the multiple objects using spatial and visual information. Here, spatial tracking model is developed by incorporating the tangential function within weighted average model. In the visual tracking, neighbourhood search algorithm and second derivate model are included to predict the next location of the objects using intensity. Finally, these two models are effectively integrated to find the final decision on the predicted location. The experimentation was performed with two videos from UCSD datasets and the results are compared with the existing methods using tracking number, distance and MOTP. The results clearly proved that, the proposed method reaches the maximum MOTP of 99% as compared with other existing methods. In future, the proposed model can be strengthened to track multiple objects in low resolution videos.



Dataset 1 MOTP analysis

Dataset 2 MOTP analysis

Fig. 7. Analysis using MOTP, a) dataset 1, b) dataset 2

References

1. M. K. Geetha and S. Palanivel, Video Classification and Shot Detection for Video Retrieval Applications, *International Journal of Computational Intelligence Systems*. **2**(1) (2009) 39-50.
2. C.M. Pun, H.M. Zhu and W.Feng, Real-Time Hand Gesture Recognition using Motion Tracking, *International Journal of Computational Intelligence Systems*. **4**(2) (2011) 277-286.
3. M. Hofmann, M. Haag and G. Rigoll, Unified hierarchical multi-object tracking using global data association, *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*. (2013) 22-28.
4. M. Pätzold and T. Sikora, Real-time person counting by propagating networks flows, *in: AVSS*, (2011) 66–70.
5. M.D. Breitenstein, F. Reichlin, B. Leibe, E. K. Meier and L.V. Gool, Robust tracking-by-detection using a detector confidence particle filter, *in:ICCV*. (2009).
6. V. Eiselein, D. Arp, M. Pätzold and T. Sikora, Real-time multi-human tracking using a probability hypothesis density filter and multiple detectors, *in: AVSS*. (2012).
7. B. Leibe, K.Schindler, N.Cornelis and L.V.Gool, Coupled Object Detection and Tracking from Static Cameras and Moving Vehicles, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **30**(10) (2008) 1683-1698.
8. Z.Jin and B.Bhanu, Single camera multi-person tracking based on crowd simulation, *International Conference on Pattern Recognition (ICPR)*, (2012) 3660 – 3663.
9. L. Kratz and K.Nishino, Tracking pedestrians using local spatio-temporal motion patterns in extremely crowded scenes. *IEEE Trans Pattern Anal Mach Intell*. **34**(5) (2012) 987-1002.
10. X. H.Xia, W.Y. Nan, Z. Wei, Z. Jiang, Y. X.Fang, Multi-object visual tracking based on reversible jump Markov chain Monte Carlo, *IET Computer Vision*. **5**(5) (2011) 282–290.
11. W.Luo, J.Xing, X.Zhang, X.Zhao and T.K. Kim, Multiple Object Tracking: A Literature Review, *Computer Vision and Pattern Recognition*, 2015.
12. J. Black, T. Ellis, and P. Rosin, Multi-View Image Surveillance and Tracking, *IEEE Workshop on Motion and Video Computing*, (2002).
13. J. Vermaak, A. Doucet, and P. Perez, Maintaining Multimodality through Mixture Tracking, *Proc. IEEE Int'l Conf. Computer Vision*, (2003) 1110-1116.
14. Jerome Berclaz, Francois Fleuret, Engin Turetken, Multiple Object Tracking Using K-Shortest Paths Optimization, *IEEE transactions on pattern analysis and machine intelligence*. **33**(9) (2011) 1806-1819.
15. X.Zhou, Y.Li, B.He and T.Bai, GM-PHD-Based Multi-Target Visual Tracking Using Entropy Distribution and Game Theory, *IEEE transactions on industrial informatics*. **10**(2) (2014) 1064-1076.
16. W. Hu, W. Li, X.Zhang and S.Maybank, Single and Multiple Object Tracking Using a Multi-Feature Joint Sparse Representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **37**(4) (2014) 816 – 833.
17. I.Ali and M.N. Dailey, Multiple human tracking in high-density crowds, *Image and Vision Computing*. **30** (2012) 966–977.
18. X.Liu, D. Tao, M.Song, L.Zhang, J.Bu, and C.Chen, Learning to Track Multiple Targets, *IEEE transactions on neural networks and learning systems*. **26**(5) (2014) 1060 – 1073.
19. W.Hu, X.Li, W.Luo, X.Zhang, S.Maybank, and Z.Zhang, Single and Multiple Object Tracking Using Log-Euclidean Riemannian Subspace and Block-Division Appearance Model, *IEEE transactions on pattern analysis and machine intelligence*. **34**(12) (2012) 2420-2440.

20. UCSD datasets http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm.
21. T.Kanungo, D.M .Mount and N.S.Netanyahu, An efficient k-means clustering algorithm: analysis and implementation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **24**(7) (2002) 881-892.
22. W.Zhong, G.Altun, R.Harrison, P.C. Tai and Y.Pan, Improved K-means clustering algorithm for exploring local protein sequence motifs representing common structural property, *IEEE Transactions on NanoBioscience*. **4**(3) (2005) 255-265.
23. S.Ilhan, N.Duru and E.Adali , Improved Fuzzy Art Method for Initializing K-means, *International Journal of Computational Intelligence Systems*. **3**(3) (2010) 274-279
24. Z.Guo, D.Zhang and D.Zhang, A Completed Modeling of Local Binary Pattern Operator for Texture Classification, *IEEE Transactions on image processing*. **19**(6) (2010) 1657-1663.
25. D.H. Kim, H.K. Kim, S.J. Lee, W.J Park, S.J.Ko, Kernel-Based Structural Binary Pattern Tracking, *IEEE Transactions on Circuits and Systems for Video Technology*. **24**(8) (2014) 1288-1300.
26. S.C. Pei and C.M.Cheng, Color image processing by using binary quaternion-moment-preserving thresholding technique", *IEEE Transactions on image Processing*. **8**(5) (1999) 614-628.
27. W.J.Heng and K.N. Ngan, "The implementation of object-based shot boundary detection using edge tracing and tracking", *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems*. **4** (1999) 439-442.
28. Kanungo, Tapas, Q.Zheng, "Estimating degradation model parameters using neighborhood pattern distributions: an optimization approach", *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **26**(4) (2004) 520-524.
29. Y.Zheng , Wuhan, China, D.S.Wong, Y.W.Wang, H. Fang, Takagi–Sugeno Model Based Analysis of EWMA RtR Control of Batch Processes With Stochastic Metrology Delay and Mixed Products, *IEEE Transactions on Cybernetics*. **44**(7) (2014) 1155-1168.
30. Multiple objects tracking performance from http://www3.ntu.edu.sg/home/JSYUAN/index_files/papers/04270203_Yang_Yuan_Wu_CVPR07.pdf