HYBRID TRACKING MODEL BASED ON OPTIMIZED DEEP LEARNING FOR OBJECT DETECTION AND TRACKING USING THE MULTI-VIEW CAMERAS

Mr. Digambar Hanmant Dewarde

Research Scholar,
Department of Computer Science & Engineering,
KLE Dr. MSSCET Belgaum, Karnataka –India.
dewarde.digambar73@gmail.com

Dr. S. Kotrappa

Professor, Department of Computer Science & Engineering,
KLE Dr. MSSCET Belgaum, Karnataka – India.
kotrappa06@gmail.com

Abstract

Video surveillance renders great services to humans formonitoring the children, crime prevention and monitoring the patients. Yet, detecting the presence of human objects in the surveillance video is found to be hectic challenge due to various constraints, such as the view point of the camera, illumination and occlusion. In this research, a hybrid model is proposed to restrain the issues, like illumination and occlusions for effective recognition of human objects in the surveillance video. In this proposed hybrid model, the object in the frame are tracked with the aid of the rectangular bounding box, proposed search source-based Deep LSTM, and modified deep sort algorithm, where the effect of vanishing gradient is minimized and performance is boosted. The search space algorithm trains the Deep LSTM and establishes an effective trade-off between the exploration and exploitation phases with a better hyperparameters that impacts directly on the tracking performance. At the same time, the modified deep sort algorithm solves the assignment problem through incorporating the motion information, which in turn increases the tracking efficiency. The performance evaluation and comparative analysis are executed to demonstrate the effectiveness of the proposed Hybrid model. The proposed Hybrid model shows impressive results in terms of MOTA, MOTP and the tracking error of98.5%, 98.1% and 0.0135 respectively.

Keywords: Video surveillance, human tracking, Bounding box, Deep sort, Deep LSTM

1. Introduction

The most significant task assigned in the domain of computer vision is the human tracking, which aim to recognize the location of the human at any edge of the world (Damotharasamy, 2020). Home land security (Chowdhry, et al., 2015) , gait characterization (Ran, et al., 2010), gender classification (Chen & Hsieh, 2012) , crowd flow analysis (Eshel & Moses, 2008),accident prediction, crime prevention, close-packed crowd (Lai, et al., 2013), fall detection of elderly(Thome, et al., 2008),driver assistance (Akhlaq, et al., 2012), smart video surveillance, monitoring the children at home(Liu, et al., 2017), monitoring the patients and to monitor the human activities (Reddy & Shah, 2013) are some of the application of the human tracking (Damotharasamy, 2020) .Yet detecting the presence of humans in the surveillance video is the hectic challenge in computer vision due to the various constraints, such as the variation in the human poses,the view point of the camera, illumination and occlusion. The deviation in exterior appearance of human is experienced with respect to their poses along with the variation in stand point as the human anatomy is comprised of various joints (Dilawari, et al., 2020).The camera should be fixed such that the view point not exceeds the angle of 45◦ to find out the human beings at any orientation yet it can recognize the human in a moderately upstanding posture.Furthermore, the human shape evaluation demands sufficient resolution and the width of the human image areneeded to be more than 24px. The humans that are detected in the surveillance are needed to be enclosed in the rectangular box in the video or the image in order to represent the presence of the humans. The dynamic background, severe occlusion and different illumination

increase the perplexity of the multiple human tracking. The objects in the video or image are represented through the points, which relay upon the past position of the each blob that comprise the position of the data and the motion (Karpagavalli & Ramprasad, 2020). The target tracking is one of the significant and the challenging premise of the image processing in the video sequence. Hence, many researchers made their effort to explore the target tracking algorithm for video classification (Li, et al., 2017; Zhang, 2019).

The human tracking algorithm is characterized into two significant methodologies, specifically known as discriminative and the generative methodologies. The most analogues regions in the image are scrutinized in order to recognize the target model in the Generative methodology (Damotharasamy, 2020). The fundamental idea is to accomplish a representation that can precisely portray the framework, and then to make a deduction between the current edge picture and the foundation model, and afterward select an edge(Zhang, 2019). The appearance of the humans varies with respect to the view point along with clothing and the visible part of the human (Wu & Nevatia, 2007). Hence, some additional complexities were experienced in tracking the human after the prior determination. The precise multiple human recognition outcomes are employed in the tracking algorithm to track the human. Toobtain the efficient tracking result in the crowd, the outcomes from the multiple human recognition system are integrated with the motion data(Karpagavalli & Ramprasad, 2020). The holistic feature is extracted in the whole –based tracking strategy and these extracted features are utilized to track the object. The whole-based tracking strategy shows prevalent execution when the article is generally furnished in the scene. However, it experiences some impudent in the presence of the occlusion (Damotharasamy, 2020).Hence, the recognition execution of the subsequent models isn't acceptable in complex scenes, while gathering adequate and exacting comments through bouncing boxes in huge scope and different situations is complex and requires high cost.

The generative strategy utilizes the classifiers to differentiate the closer view objects from the background sectors which are developed to suppress the drawbacks of the generative methods. Now-a-days various classification algorithms are developed to track the human in the videos or images (Damotharasamy, 2020). The researchers utilized the background subtraction method(Zhang, et al., 2020) in order to update the quantity of the models, learning rates and the weights. The background subtraction method is dependent on mixture Gauss strategy to segregate the targeted human from the background. Further the gradient direction histogram of combined edge bearing aggregation and feature is utilized in the method to segment the region of interest through the feature description.The classification and the recognition of the humans in the particular video or image are executed through the Support vector machine (SVM) (Zhang, et al., 2020). The edge variation between the two adjacent planes of RGB3 planes are estimated in the detection method based on the deep learning technique. Then, the estimated edge variation in each pixel is compared with the superimposed calculation(Zhang, et al., 2020).The additional features in the aforementioned strategy enhances the recognition performance in the two well accepted datasets namely, Crowd Human Dataset(Shao, et al., 2018) and the COCO persons dataset (Lin, et al., 2014).Furthermore, the method provides the perception about the enhancement of the recognition performance through the external features of the object. In order to obtain the insight about performance enhancement the learned discriminative features are scrutinized and compared with the genuine features(Wang, et al., 2019). The Cam shift algorithm method is another efficient method employed to attain the target tracking. Yet the computational delay, low recognition and the tracking efficiency due to the multiple algorithms are the main drawback experienced in the method (Damotharasamy, 2020) .Hence, the exploration in the tracking, object recognition and the human tracking domain based on deep learning is required so as to restrain the issues experienced in the above said methodologies.

This research aims to develop advanced and well-organized tracking models, which handle the above-said issues, like occlusion and illumination. In the proposed Hybrid model, a set of standardized cameras are utilized to track the multi-view image. The captured images are then subjected to pre-processing, where the keyframe extraction and contrast enhancement is carried out. The pre-processed keyframes are then subjected to the object tracking using the proposed search source-based Deep LSTM, modified Deep sort algorithm, and bounding box model.The major contribution of the system is enlisted below.

- ✓ *Visual tracking using bounding box model:*Generally, a rectangular bounding box is utilized for tracking the objects in the key-frame in order to represent the direction of the model. The pixel co-ordination and the ground truth are the main parameters considered in the determining the target of the bounding box.
- ✓ *Object tracking through Search space-based Deep LSTM:* The Deep LSTM is generally employed in the object tracking system to minimize the vanishing gradient and to avoid the performance degradation. In this research, the performance is further boosted through training the deep LSTM with the proposed search source algorithm, which enabled the global optimal convergence.
- ✓ *Proposed modified Deep sort algorithm:* The modified deep sort algorithm is developed with the integration of the entropy factor along with the standard deep SORT algorithm, which solves the assignment problem to incorporate the motion information and enables effective object tracking.

The organization of this article is as follows: section 2 comprises of the review of the literature, section 3 elucidates the major challenges experienced in the research. Section 4 elucidates the framework of the proposed technology, section 5 elucidates the Results and discussion of the research and section 6 concludes the research.

2. Review of the Literature

This section elucidates the review of some conventional methods that are adapted for tracking the humans in the multi-view camera. Zhang, et al., (2020) developed a human tracking and recognizing system based on deep learning technique. The target is detected through the background subtraction method which is dependent on the well organized hybrid Gaussian background model. The holes that are encompassed in the foreground domain are eliminated through the morphological filtering. The Advantages of the deep learning based tracking systems are the enhanced position accuracy and the highest recognition and the tracking efficiency. The issues related to the training of deep learning is the main drawback experienced in the human tracking system.Karpagavalli & Ramprasad, (2020)designed a scheme for the automatic recognition and the tracing of the human in the multiple cameras, which is based on Adaptive Hybrid Gaussian Mixture Model. The supremacy of the system is that it effectively manipulates the issues related to the background variation and gradual illumination. Furthermore the system is automatically revamped in regular interval to cope up with biased occlusion. Though the system is efficient in tracking the human it is not suitable for tracking in the rain and the night vision. Wu & Nevatia, (2007) developed a system that can recognize and track the partially occluded human beings utilizing the Bayesian Combination of Edgelet dependent Part Detectors. The edge let features are employed to upgrade the debilitated classifiers. The advantages considered in the methods are that it can recognize and track the human both in moving and standing position. Furthermore the method obtain reduced false alarm rate. Some of the biological structure of the human is not distinguished in this system. The other drawback of the system is that it does not possess any prompt for motion segmentation. Moreover it requires the well adapted classifiers in order to achieve enhanced performance and accuracy. Damotharasamy, (2020) developed an effective algorithm to track the human beings in the video surveillance. The method utilizes the orientation of the gradient and the sub space learning to track down the human. The reduced reconstruction error and ease of segregating the human from the background domains are considered as the advantages. Yet the system fails when the humans is occluded and appears in various illumination environment. Chen, et al., (2020) developed a semi-administrated human recognition system. The system explores both the labeled and unlabeled data to restrain the optimization issues so as to enhance the object recognition capacity of the system. The system requires an appropriate and well organized training to improve the performance of the classifiers. The system enables the feature extraction only from the last layer of the convolutional layer for the estimation of IOU without providing any data with regards to the for-ground domain. Wang, et al., (2019)devised the human tracking system with the aid of deep learning techniques. The system can effectively detect the humans in the highly occluded domains. The drawback experienced in the system is that the detection module and the segmentation module remain solitary with each other.Zhang, (2019) explored the cam shift algorithm for effective tracking of the humans in the multi-view surveillance camera. The reduced computational area and the enhanced robustness are listed as the advantages of the cam shift algorithm method. The complexity of the algorithm is the prime disadvantages of the system. Dilawari, et al., (2020) devised three visual perception algorithms, based on Counter, histogram of oriented gradients (HOG) and speeded up robust features (SURF).The improvement of annotation accuracy and the reduction of the human efforts are the prime advantage of the system. The latency occurred in the Counter, HOG, SURF and Deep learning in mining database and feature extraction is the main drawback of the system.Liu, et al., (2020) devised an recognition algorithm to determine the human motion in the video sequence. The issues such as shape variation and local occlusion are restrained through the hidden Markov model algorithm process.The ViBe algorithm is utilized to perform the background subtraction. Improved recognition rate and good anti-interference performance are the prime advantage of the method. The variation of color and hog performance with respect to the target appearance is one of the drawbacks of the system.

2.1 Challenges

The major challenges of the research include:
This section elucidates the challenges that experienced in the recognition and tracking the human in the surveillance video.

- The small scale detectors are utilized in the Edge let dependent part detectors to observe the human postulations. However, the small scale part detector fails to distinguish some of the human parts such as shoulder, head. Hence, it is preferable to use large scale part detectors (Wu & Nevatia, 2007) .
- The degradation is observed in the recognition accuracy of the global-based sparse classification due to the Gaussian assumptions that are consigned in the noise distribution (Liu, et al., 2020) . This is the main drawback that is experienced in the shape variation and local occlusion in the human recognition.
- The reorganization of the conventional methods is required to restrain the issues that are occurred through the climatic conditions such as rain and night vision. Furthermore the computational complexity is also need to be reduced so as to implement the human recognizing and tracking method in real –time application(Karpagavalli & Ramprasad, 2020).
- The background frame estimations were negatively influenced through the fractional or full impediments. The impediments are occurred when the moving target is obscured by the tree or moving through the foot

path. Hence, the background dynamics that are responsible for unpredictable and intermittent variations are required to be restrained (Dilawari, et al., 2020).

- Even though the human tracking system based on tracker system effectively track down the human in the complicated attributes fails in motion blur. Therefore, the exploration of human tracking within the motion blur is required (Damotharasamy, 2020).
- The irrelevant data enhance the computational time and the recollection rate of the model, which in turn makes the system ill-adapted for real-time application.Thus the irrelevant data enhance the perplexity of tracking the human in the surveillance. Deep multi-layer network is utilized for saliency prediction, which makes the strategy more feasible in tracking the objects (Gajjar, et al., 2017).
- Track down the multiple objects from the video segment is found to be more complex in the online demonstration for managing the particles, which is under undetermined condition.
- The strategy that utilized in (Gamage, et al., 2018) fails to track the object without any interruptions for the persons who posses indiscriminate walking patterns. Hence, to track the persons without any interruptions the comparison is accomplished between the angular distance and the walking direction.

3. Proposed Hybrid tracking model to track the video objectsusing the Multi-viewsto tackle occlusion

This section elucidates the proposed framework to recognize and track the objects in the multi-view videos.The issues related to the single-view cameras, such as occlusion and illumination are greatly reduced through the multi-view cameras. Thus, the proposed tracking model is developed, which deals with the occlusions, minimizes the false alarms, and is tolerant for the pose variations. In particular, the multi-view cameras handle the occlusion, which is the common problem associated with the object tracking. In other words, an object in a video frame, disappear in the next frame, which badly affects the tracking performance as the same object may appear in the latter frames, leaving the problem of associating the older track of the object along with the features of the trajectory.Figure 1 shows the proposed model for object detection and tracking using the multi-view cameras based on the hybrid tracking model. The multi-view videos on the same scene are interpreted for performing the object tracking. Initially, videos from the individual cameras are subjected to the contrast enhancement and key-frame extraction, which assures smooth processing of the further steps. Let us suppose, there is n number of videos for which the keyframe extraction is done, which facilitates the selection of keyframes for object detection and tracking such that the computational complexity is very much reduced. From the keyframes, the objects are located and tracked for which the proposed hybrid tracking model that is developed using the visual tracking model, proposed search source-Deep LSTM-based tracking, and modified deep sort algorithm-based tracking is used. The tracked location of the objects in the video keyframes is fused using the weighted average fusion model, which yields the final tracked location of the objects in the video.

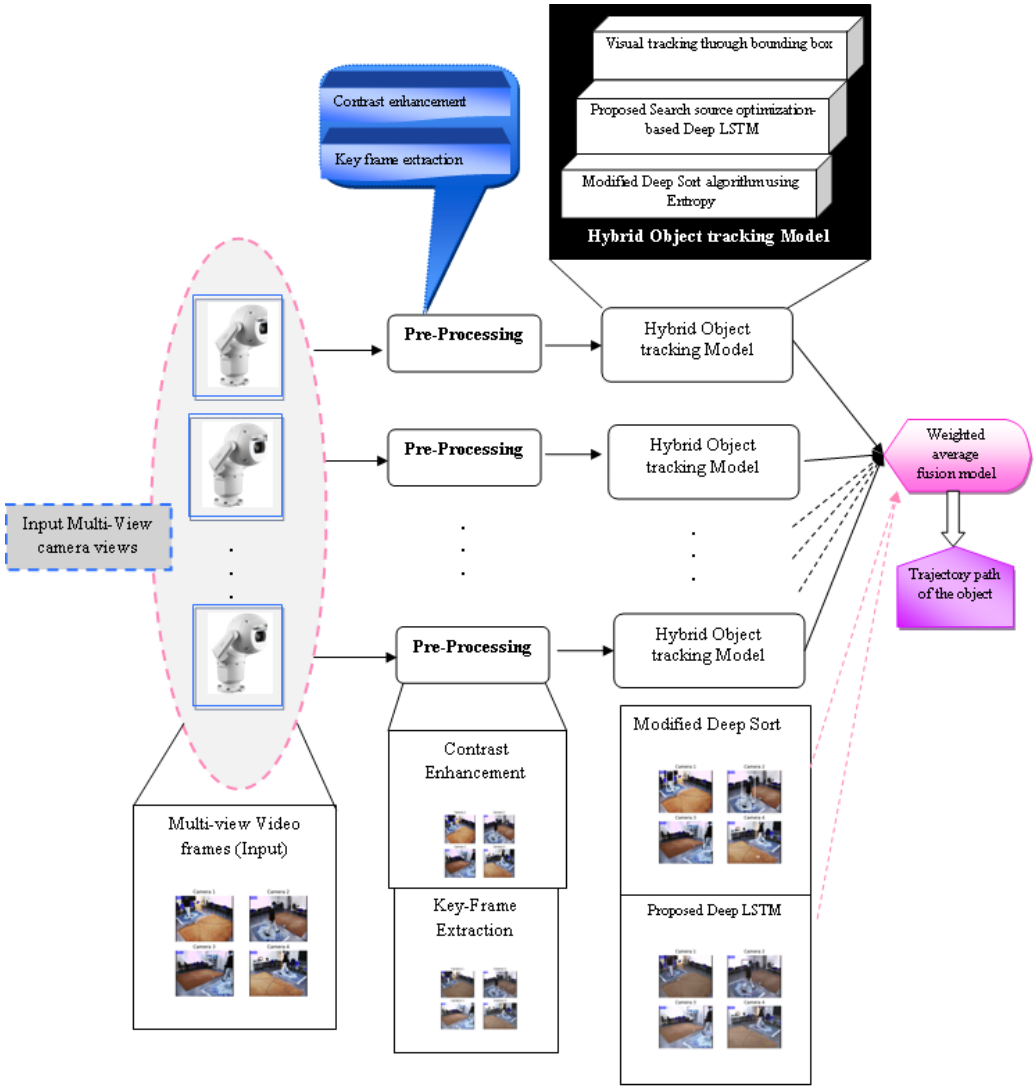


Figure 1. Proposed hybrid tracking model for object tracking using multi-view cameras

Let D be the database consisting of n number of videos required for tracking the object. The input video consist of n videos and it is given by,

$$V = \{V_1, V_2, ..., V_i, ..., V_n\} \tag{1}$$

where, i represents the i_{th} video and V corresponds to the individual video in the database.

3.1 Pre-processing

The pre-processing step enables the effective processing of the video frames in the successive steps of tracking and the considered pre-processing steps include the contrast enhancement and key-frame extraction. The significance of pre-processing lies in the intention to reduce the computational complexity of the tracking process through reducing the irrelevant data. The in-depth explanation of the steps follows.

Contrast enhancement: One of the important steps utilized for the subjective analysis of the image quality is Contrast and differentiation is modeled by the distinction in the reflected luminance from two adjoining surfaces. As such, contrast is the distinction in visual properties, which makes an object recognizable from the background and other objects. Contrast, in terms of visual perception is regulated as variations in the brightness and color of the object with other objects. Visual framework is more susceptible to the contrast while comparing with absolute luminance. Hence, little consideration is paid to impressive changes in the illumination conditions. In this research, histogram equalization is used for the contrast enhancement of the video frames recorded by the cameras.

*Key frame extraction:*Thekey-frame extraction is employed in the proposed human tracking method to extract the group of frames that posses the better representations of images. The key frame extraction removes the irrelevant

and repeated frames without affecting the salient features of the shot. Each video V_i consist of a set of frames that is represented as f_j .

$$V_i = \{f_1, f_2, ..., f_j, ..., f_m\} \tag{2}$$

where, j represents the j_{th} frame. The key-frames are selected based on the Euclidean distance measure, which computes the distance between the neighboring frame f_j and the reference frame f_i such that the neighboring frame j is selected if-and-only-if the Euclidean distance between the j^{th} frame and i^{th} frame is minimal. The key-frame selection is formulated as,

$$ED = \begin{cases} 1 & ; \quad \partial[f_i, f_j] < \partial_{Thres} \\ 0 & ; \quad Otherwise \end{cases} \tag{3}$$

where, ED is referred as the Euclidean distance and ∂_{Thres} belongs to the threshold value of distance employed for selecting the key-frames. The selected key-frames are given by,

$$V_i = \{f_1, f_2, ..., f_j, ..., f_n\} \text{such that } (n < m) \tag{4}$$

The key frames are employed for the object tracking, which minimizes the computation complexity associated with the processing of the individual video frames.

3.2 Proposed hybrid tracking model using visual tracking and deep tracking models from the multi-view videos

This section elucidates the proposed model for tracking the occluded objectsof the same scene in the surveillance videos. The proposed tracking model comprised of three phases: visual tracking using the bounding box model, proposed search source-based Deep LSTM classifier and the modified deep sort algorithm, which are elaborated in the following subsections.

*3.2.1 Visual tracking using bounding box model:*Generally, a rectangular bounding box is utilized for tracking the objects in the key-frame in order to represent the direction of the model(Yuan, et al., 2020) . This sub-section elucidates the object tracking in the video frames using the rectangular box. The pixel co-ordination and the ground truth are the main parameters considered in the determining the target of the bounding box.The center location of the h_{th} object in the j_{th} frame is denoted as, $\left(A_{[j]}{}^{j}, B_{[j]}{}^{h}\right)$.The position of h_{th} object in the next frame changes as the object moves within a certain limit 'L'. Therefore, the centre position of the h_{th} object in the $(j+1)_{th}$ frame is denoted as, $\left(A_{[j+1]}{}^{h} \pm L, Q_{[j+1]}{}^{h} \pm L\right)$.The features are extracted by the bounding box method, where parallel processing is enabled to classify the object and recognize the bounding box.The input is denoted as 'S' training sets of $\left\{\left(P_{coor}{}^{j}, G^{j}{}_{truth}\right)\right\}$where, the training sets vary in the range of $\{0, 1, 2 \ldots \text{S}\}$.The pixel coordinates P in the bounding box is given by,

$$P_{coor}{}^{j} = \left(P_{coor(a)}{}^{j}, P_{coor(b)}{}^{j}, P_{coor(c)}{}^{j}, P_{coor(d)}{}^{j}\right) \tag{5}$$

The ground truth of the bounding box is given by,

$$G_{truth}{}^{j} = \left(G^{j}{}_{truth(a)}, G_{truth(a)}{}^{j}, G^{j}{}_{truth(a)}, G_{truth(a)}{}^{j}\right) \tag{6}$$

The target of the regression is given by t ,

$$t_a{}^{j} = \frac{\left(G_{truth(a)}{}^{j} - P^{j}{}_{coor(a)}\right)}{P_{coor(c)}{}^{j}} \tag{7}$$

$$t_b{}^{j} = \frac{\left(G_{truth(b)}{}^{j} - P_{coor(b)}{}^{j}\right)}{P_{coor(d)}{}^{j}} \tag{8}$$

$$t_c{}^{j} = \log\left(\frac{G^{j}{}_{truth(c)}}{P_{coor(c)}{}^{j}}\right) \tag{9}$$

$$t_d^{\,i} = \left(\frac{G_{truth(d)}^{\,j}}{P_{coor(d)}^{\,j}} \right) \tag{10}$$

Using the bounding box model, the location of the video objects in the successive keyframes is tracked and the tracked location using the visual tracking model is denoted as, Tt_1.

*3.2.2 Object tracking through Search space-based Deep LSTM:*This section enumerates the object tracking from the multi-views based on the proposed search space-based Deep LSTM. One of the prime issues that degrade the performance of the deep networks during object tracking is the vanishing gradient, which is defined as the state in which the input of the different hidden layers reduced exponentially with respect to the variations in the time steps. The LSTM is generally employed in the object tracking system to minimize the vanishing gradient and to avoid the performance degradation. Adeep neural network, which comprises the layers of LSTM known as Deep-LSTM is accomplished in this research to obtain a better tracking performance.Deep–LSTM utilizes the advantage of both the LSTM recurrent network and Deep networks. The diagrammatic representation of the LSTM is illustrated in the Figure 2.

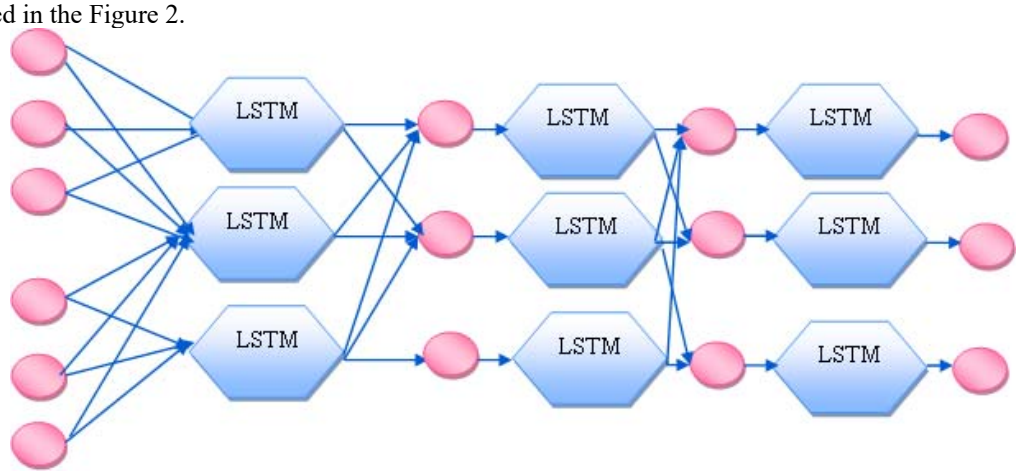


Figure2. Diagrammatic representation of LSTM

The Deep-LSTM framework comprises of some interior state cells, which remains as the short term and long term memory cells. These memory cell influences the outcomes of the Deep-LSTM network. The memory cell of the Deep-LSTM consists of various fractional units with unique objectives. The block diagram representation of the LSTM is illustrated in the Figure 3.

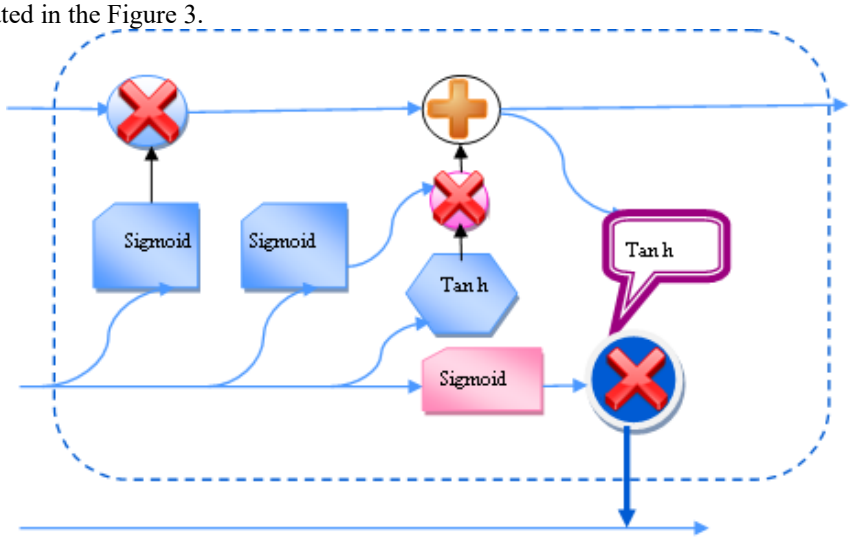


Figure 3. Block representation of LSTM

By incorporating the idea of memory cell, Deep LSTM algorithm vanish the gradient problem and the output depends on the states of these cells.The data predicted is non-linear and it is carried out by the sigmoid function, which improves the prediction accuracy. The weighted sum of k_m and r_{m-1} is passed through tanh function.The input node sustains the input of the deep networks and the input obtained from the previous hidden layers. The input node l_m receives input k_m from the input layer of the deep network and from the previous hidden states

r_{m-1} of the node in time steps.The sigmoid is one of the non-linear functions and it is mainly employed to enhance the prediction accuracy. Hence, a weighted sum of both the output is fed to the $\tanh$ function. The outcome of $\tanh$ function is represented in the following expression as,

$$l_m = \tanh\left(k_m \bullet W_{lk} + r_{m-1} + biasinputnode\right) \tag{11}$$

The input gate sigmoid function restricts the flow of input when if the net value is found to be zero. The input gate sigmoid function enables the flow of output if the net value is found to be 1.The sigmoid activation function is termed as the input function and it is given by,

$$u_m = \sigma\left(k_{m\bullet} W_{lk} + r_{m-1} W_{lr} + biasinputnode\right) \tag{12}$$

The third state, which is known as interior state is a node that comprises of a self-loop recurrent edge of both the linear activation function and unit weight. The updated function of internal state v_m with a self-loop recurrent edge is given by,

$$v_m = u_m \Theta l_m + r_{m-1} \tag{13}$$

The forget state $\left(x_m\right)$ is a subunit to reinitiate the internal state of memory, which is given by,

$$x_m = \sigma\left(k_m \bullet W_{rk} + r_{m-1} W_{lr} + biasforget\right) \tag{14}$$

Finally, the output is given by the state 'O_m' as,

$$O_m = \sigma\left(k_m W_{Ok} + r_{m-1} W_{Or} + biasoutputgate\right) \tag{15}$$

Now, the final output of the memory cell,

$$r_m = \tanh\left(V_m\right) \Theta O_m \tag{16}$$

$$v_m = l_m \Theta u_m + v_{m-1} \Theta x_m \tag{17}$$

In the deep LSTM, training the internal model parameters is a hectic challenge for which proposed search source algorithm is utilized. The internal model parameters to be trained include the weights and the biases, where the weights are given by, $W = \left\{W_{lk}, W_{lr}, W_{rk}, W_{Ok}, W_{Or}\right\}$ and the biases correspond to the input gate, forget gate, and output gate. The deep insight into the proposed search source algorithm is enumerated below.

Search Source Algorithm for training the Deep LSTM: The search source optimization is based on the searching mechanisms of Hymenopterans, which is developed with the hybridization of the optimal search mechanism for food and source secretion characteristics of the honey bees (Dorigo, et al., 2006) and ants (Karaboga & Ozturk, 2011) . The search source algorithm exhibits stronger exploration ability than the standard optimizations operating under the nature-inspired and artificial computing platforms. The search source optimization is based on the foraging mechanisms of the employer and onlooker hornets, which provides a chance to learn from individuals with better performance. The control operators used include the global and local search and the search mechanism employed in search space overcome the oscillation phenomenon in employed hornets. They adapted intelligent learning mechanism to accelerate the convergence rate of the worst employee hornets. The intelligent learning schemes accelerate the convergence of the employer hymenopterans, spectatorhymenopterans and outriderhymenopterans. The usage of special turbulent operator helps to balance global and local searches, which further establishes an effective trade-off between the exploration and exploitation phases.

A) *InitializationPhase:* The initialization of the control parameters and the population of search source optimization are done in the first step. The control parameters for the optimization, like colony size, $Q = a_1$, limit for scout, $Y = \dfrac{Q * I}{2}$, dimension of the problem $I = 2$, and position be denoted as, $Xx_{i_1 j_1}$.

*Initialization of the Food source:*The total feasible solutions are randomly initialized and the search process is initiated with the randomly set search space. The position for the j_1^{th} variable for the i_{th} food source with upper and lower bound values for the search space is given by,

$$Xx_{i_1 j_1} = Lq_{i_1 j_1} + \left(uq_{i_1} - Lq_{i_1 j_1}\right) * Rrand\left(0,1\right) \tag{18}$$

where, $i_1 = 1,2,.....N_q$ correspond to the food sources and $j_1 = 1,2,.....I$ denotes the search dimension involved in optimization process. Let $Rrand\left(0,1\right)$ is the uniformly generated random number in the interval $\left[0,1\right]$ and the upper and lower bounds of the search process is denoted as, uq_{i_1} and $Lq_{i_1 j_1}$.

*B) Division of employed hymenopterans:*In the proposed optimization, the optimal search space is decided based on the three groups of searchers namely, employer, spectator and outrider hymenopterans. It is evident that in the employer phase, each employer is associated with a food source. However, the employer hymenopterans search for the new food sources and the updates the location of the new sources with the spectator hymenopteransand the search is based on their memory of the neighboring searches. The employer hymenopterans find the new source of food using the equation,

$$v^{t+1}{}_{i_1 j_1} = Xx^{t}{}_{i_1 j_1} + \phi * \left(Xx^{t}{}_{i_1 j_1} - Xx^{t}{}_{k_a j_1} \right) \tag{19}$$

where, $Xx^{t}_{i1\,j1}$ in the above equation represents the neighbors selected randomly within the range $[1, N]$, ϕ represents the randomly generated value $(-1,1)$, and t represents the current iterations, $v^{t+1}{}_{i_1 j_1}$ signifies the new location of the food, and $Xx^{t_a}{}_{k_a j_1}$ refers to the j_1^{th} component of another food. Greedy algorithm is used todetermine survivability of the candidate solution $v^{t+1}{}_{i_1 j_1}$ with respect to $Xx^{t}_{i_1}$. Rearranging equation (25), we get,

$$v^{t+1}{}_{i_1 j_1} = X^{t}{}_{i_1 j_1} \left(1 + \phi_{i_1 j_1}\right) - \phi_{i_1 j_1} X_{xk_a j_1}{}^{t} \tag{20}$$

The equation (33) portrays the optimal search mechanism, which exhibits the local optimal convergence. However, the local optimal convergence is a major problem, which is handled effectively through the integration of the source secretion characteristics along with the optimal search mechanism. Thus, the source secretion characteristics is highlighted as,

$$v_{x i_1 j_1}{}^{t+1} = \left(1 - \phi_{i_1 j_1}\right) \bullet X^{t}{}_{x i_1 j_1} + \phi_{i_1 j_1} \tau_0 \tag{21}$$

Rearranging the equation (34), we get,

$$\left(\frac{v_{x i_1 j_1}{}^{t+1} - \phi_{i_1 j_1} \tau_0}{1 - \phi_{i_1 j_1}} \right) = X^{t}{}_{x i_1 j_1} \tag{22}$$

On substituting (22) in (20),

$$v_{x i_1 j_1}{}^{t+1} = \left(\frac{v_{x i_1 j_1}{}^{t+1} - \phi_{i_1 j_1} \tau_0}{1 - \phi_{i_1 j_1}} \right) \left(1 + \phi_{i_1 j_1}\right) - \phi_{i_1 j_1} X_{xk_a j_1}{}^{t} \tag{23}$$

$$v_{x i_1 j_1}{}^{t+1} = \frac{1}{2\phi_{i_1 j_1}} \left\{ \left(\frac{\phi_{i_1 j_1} \tau_0}{1 - \phi_{i_1 j_1}} \right) \left(1 + \phi_{i_1 j_1}\right) + \phi_{i_1 j_1} X_{xk_a j_1}{}^{t} \right\} \tag{24}$$

Equation (34) is the final updated equation for the proposed search source algorithm, which inherits both the characteristics, like optimal search mechanism and source secretion characteristics that boosts the global optimal convergence and renders effective diversification and intensification. Thus, the employer bees return to their hives and communicate the optimally discovered location of the food sources with the other members through a waggle dance. The survivability of the new food sources is checked based on the below fitness function. If F_{i_1} is better than $Xx^{t}_{i_1 j_1}$ then, the candidate solution $v^{t+1}{}_{i_1 j_1}$ is preserved and $Xx_{i_1 j_1}$ is discarded else, $v^{t+1}_{i_1 j_1}$ is abandoned, which is modeled as $F_{i_1} < Xx^{t}_{i_1, j_1}$.

*C) Division of onlooker hymenopterans:*The onlooker hymenopterans receive the nectar information of the available food source from the employed hymenopterans and tracks the fresh availability based on their accessible location. Hence, to find the best food source, classical route wheel strategy selection is used, where the solution with the greatest probability is selected. The selection probability on each food source is given by,

$$P_{i_1}{}^{t} = \frac{F_{i_1}}{\sum_1^N F_{i_1}} \tag{25}$$

The objective function used in this research is the prediction error that replaces the general objective function shown in equation (26). The fitness function is given by,

$$f_a^{t_a} = \begin{cases} \dfrac{1}{1+f_a} & ; \quad if\ f_a \geq 0 \\ 1+A_a B_b S_s(f_a) & ; \quad if\ f_a < 0 \end{cases} \tag{26}$$

where, F_{i_1} refers to the fitness measure, and f_a symbolizes the objective measure of food source i_1. As per the objective of the proposed search space optimization, the objective is to minimize the prediction error of the deep LSTM, which is formulated as,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Z_i - Z_i^{\Lambda}\right) \tag{27}$$

where, n represents the number of the training samples, z_i represents the predicted samples and z_i^{Λ} represents the actual samples. Thus, the food source is selected using the update equation, which is same as in the employed hymenopterans and the survivability of the equations is verified using the greedy selection algorithm.

D) Division of outrider hymenopterans: The update in the solutions is done in the above two phases and when a situation arises such that there is no further update in the best solution. Under such condition, the employed hymenopterans corresponding to the best solution becomes the outrider hymenopterans, while the outrider becomes the employed hymenopterans. Here, the position $Xx_{i_1 j_1}^t$ is initialized randomly in the entire search space and the limit of the food is reinitialized to zero.

E) Termination: The steps are repeated for the maximal number of t in order to reveal the effectiveness of the proposed optimization. Algorithm1 shows the pseudo code of the proposed search source algorithm.

Algorithm 1. ***Search Source Algorithm***

1	**Input:** $Xx_{i_1 j_1}$
2	**Output:** $v^{t+1}{}_{i_1 j_1}$
3	Set the parameters with maximum number of iterations as Xx
4	Initialize the population of food source $Xx_{i1,j1}$
5	Where $i = 1,2,3.....,N_q$ and $J_1 = 1,2,.....I$
6	For each food source i, j , set $i = 0$ and $j = 0$
7	Set iteration = 1 /* iteration counts ABC iteration*/
8	Repeat
9	/*hymenopterans bee phase*/
10	For $i = 1 to N_q$
11	Generate new food source using the equation (19)
12	Apply greedy selection
13	If improvement in the food source $trail_i = 0$ Otherwise $trail_i = trail_i + 1$
14	End
15	Find the probability value of food source using equation (25)
16	/*Spectator bee phase*/
17	Initialize $t = 0$ and $i = 0$
18	Repeat
19	If random < $P_{i_1}^t$
20	Generate a new food source for onlooker using the equation (19)
21	Apply greedy selection
22	If food source $Xx_{i_1 j_1}$ did not improve then $trial_i = trial_i + 1$,otherwise $trial_i = 0$
23	t= t+ 1
24	Until (t= $Xx_{i_1 j_1}$)
25	End
26	/* outrider bee phase */
27	If max ($trial_i$)> limit then

28	Replace the food source with a new randomly generated food source using Equation(18)
29	Store the best food source
30	iteration = iteration + 1
31	Until (iteration = maximum iteration)
32	End

Thus, the proposed search source-based deep LSTM tracks the location of the objects in the video and the tracked location using the proposed search source-based deep LSTM is denoted as, Tt_2.

*3.2.3 Object tracking using the proposed modified deep sort algorithm:*In the tracking system, it is difficult to determine the position of the target as there is incorporation between the motion information. Therefore, modified Deep sort algorithm is proposed in this research to solve the assignment problems to incorporate motion information.Modified deep sort algorithm is the advancement of the Simple Real Time Tracker (SORT)algorithm, where a conventional single hypothesis tracking of algorithm is adopted with recursive kalman filtering through framing the data with its pixel value. Assume a general tracking scenario, where the camera is un-calibrated and thebenchmarks for multiple object tracking are the most common step in consideration with filtering the framework and it is explained with the help of the kalman filtering technique.The proposed modified deep sort algorithm not only tracks the distance and velocity of the objects, but also computes the deep features for the bounding box and tracks the objects in the video based on the similarities between the deep features corresponding to the bounding boxes in the video frames. The in-depth idea of proposed modified deep sort algorithm is given below.

*Kalman filtering for track handling:*The hectic challenge is regarding the object tracking, when there is a lack of ego-motion availability and the camera utilized to capture the scene is incalculable. Hence, the tracking structure is represented as the eight dimensional variables,which represent the position of the bounding box, and the other variables express the velocities of the image co-ordinates. The Kalman filter is utilized with the persistent velocity motion and the linear observation model in which the bounding coordinates represented-above are taken as the immediate perceptions of the object. The number of frames is counted for each track since it attains the ultimate associate measurement. The counter is reset to zero when the track obtains the ultimate associate measurements and the counter is incremented in the course of Kalman prediction.

Assignment issues:The problems prevailed when there is association between the Kalman prediction and the measurements are newly measured. The squared Mahalanobis distance is utilized to resolve the issues and create the corporation between the newly appeared measurements and the Kalman prediction. The modified deep sort algorithm is used to solve the assignment problem to incorporate the motion information. The newly arrived measurements is given by,

$$g^{(1)}(s,z) = (g_z - w_s)^T R_s^{-1} (g_z - w_s) \tag{27}$$

Here, the s_{th} track is distributed into measurement space by (w_s, R_s) and the z_{th} bounding box detection is done using g_z .The Mahalanobis distance is used to estimate the uncertainties through determining the number of standard deviation.The Mahalanobis distance considers the uncertainty estimation through measuring the number of the standard deviation in which the determination is far away from the mean track. Moreover, the filter eliminates the superfluous associations and the decision vector is done with the help of an indicator.

$$o^{(1)}_{s,z} = 1\left[g^{1}_{(s,z)} \le C^{(1)}\right] \tag{28}$$

where, $C^{(1)} = 9.4877 - Threshold\ value$. Furthermore, the unpredicted motion of the camera is responsible for the accelerated displacement in the image plane, which in-turn makes the Mahalanobis distance unconcerned metric to track the occlusions.The second metric measurement is the smallest cosine distance between the s_{th} track and the z_{th} detection is given by,

$$g^{(2)}(s,z) = \min\left\{1 - H_z F_{Hu}^{(i)} \Big/ H^{(i)}{}_u \in K_i\right\} \tag{29}$$

Again a binary variable is introduced to indicate the association is admissible to the metric,

$$E_{(S,Z)}^{(2)} = \left[g_{(s,z)}^{(2)} \le C^{(2)}\right] \tag{30}$$

Both the Mahalanobis distance metrics and the cosine distances are compliment to each other as the Mahalanobis distance only elucidates the information related to the location of the object, whereas the cosine distance provides the information about the appearance of the object. In combination, both the metrics complement each other by serving different assignment problems. The cosine distance considers appearance information that is particularly useful to recover identities after long term occlusions, when motion is less discriminative.Now, the consideration is carried out by combining both the metrics as weighted sum,

$$J_{S,Z} = \lambda_g^{(1)}(s,z) + (1-\lambda) g^{(2)}(s,z) \tag{31}$$

where, λ controls the combined cost association, which is modeled using the entropy function that depends on the probability function. The integration of the entropy in the modified deep sort algorithm is highlighted in equation (31), where λ is evaluated based on the entropy concept, which ensures the effective similarity matching between the objects detected in the video frames that further boosts the performance of object tracking. The entropy measure is formulated as,

$$E_A(e_i) = \sum_{i=1}^{u_1(e_i)} P_{i_1}^t \log P_{i_1}^t \tag{32}$$

where, $P_{i_1}^t$ is the individual points in the bounding box that detects the object and the deep features are extracted for the individual points in the detected objects between the frames so that the tracking of the objects is enabled.Thus, the tracked trajectory using the proposed deep sort algorithm is denoted as Tt_3 .

3.3 Weighted average fusion for object trajectory

The tracked trajectory using the three tracking models is integrated to generate the final trajectory path of object in the video. The trajectory path of the multi-view video is then considered so that the occlusion of the objects is tackled effectively. The final trajectory path of the proposed hybrid model is formulated as,

$$Tt = \frac{Tt_1 + Tt_2 + Tt_3}{3} \tag{33}$$

where, Tt represents the tracked response of an object in the video-frame.

4. Results and Discussion

The results and discussion of the proposed Hybrid tracking model is elucidated in this section. The performance evaluation and the comparative analysis are implemented in this research so as to manifest the supremacy of the proposed Hybrid tracking model.

4.1 Experimental setup

The proposed Hybrid tracking model is executed in the PYTHON, which operates in the Desktop with 4GB RAM and Windows 10 Operating system.

*Dataset description- CAVIAR4REID dataset:*The Caviar4reid dataset is one of the dataset which is utilized evaluating the humans in public places. This datasets consists of several video sequences which are recorded in the outdoors of the shopping complex in Lisbon. It consists of 26 sequences of video clippings, which was captured with the resolution of 384x288 pixels. These video clippings include people entering the shop, exiting the shops, walking alone, chatting with other persons and window shopping. Then, 72 of the pedestrian is selected from the video clippings among them 50 pedestrians were selected from camera view and 22 from the single camera view. Some constrains like occlusions, light condition and resolution change make them challenge for re-identification task.

4.2 Simulation results

This section briefly enumerates the simulation results of the proposed hybrid tracking model to recognize and track the object in the multi-view camera. For the effective tracking of the objects from the videos and to tackle the occlusion effects, the videos from the four cameras are utilized in this research. The videos obtained from the four cameras are depicted in the figure 4.These gathered video sequences were first subjected to the pre-processing, where the contrast enhancement and the key frame extraction is done. The pre-processed images are then exposed to the proposed Hybrid tracking model for effective tracking of the objects in the video sequences.

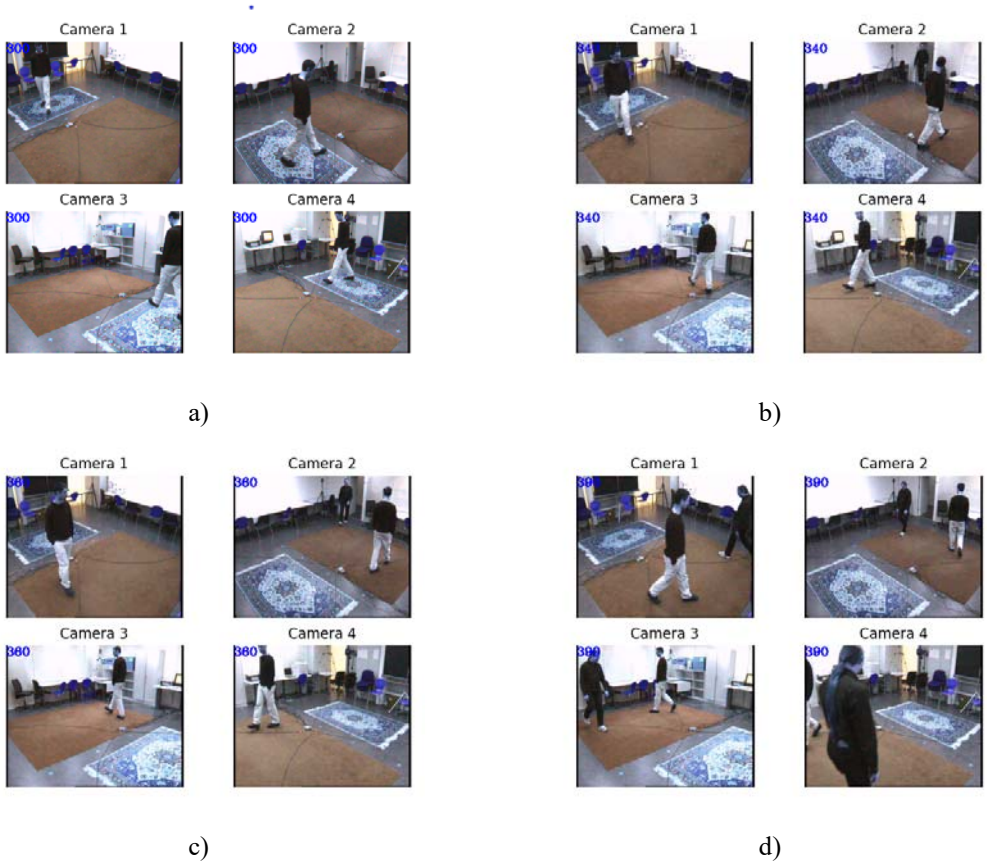


Figure 4.Multi-view cameras employed for occlusion handling during video object tracking a) Frame_300, b) Frame_340, c) Frame_360, and d) Frame_390

The results obtained from the proposed Hybrid tracking model are demonstrated in the figure 5. The rectangular bounding box is utilized in order to determine the position of the object in the multi-view camera. The vanishing gradient is greatly reduced through the proposed search space optimization based on Deep LSTM. Moreover, the modified Deep sort algorithm is employed in the proposed Hybrid tracking model for the effective tracking of the object by restraining the incorporation between the motion information.The simulation result shows that proposed Hybrid tracking model that tracked the objects in the multi-view camera. The tracked objects are susceptible to the similarity matching of the bounding boxes using the proposed deep LSTM that assures that the objects in the four cameras are same.

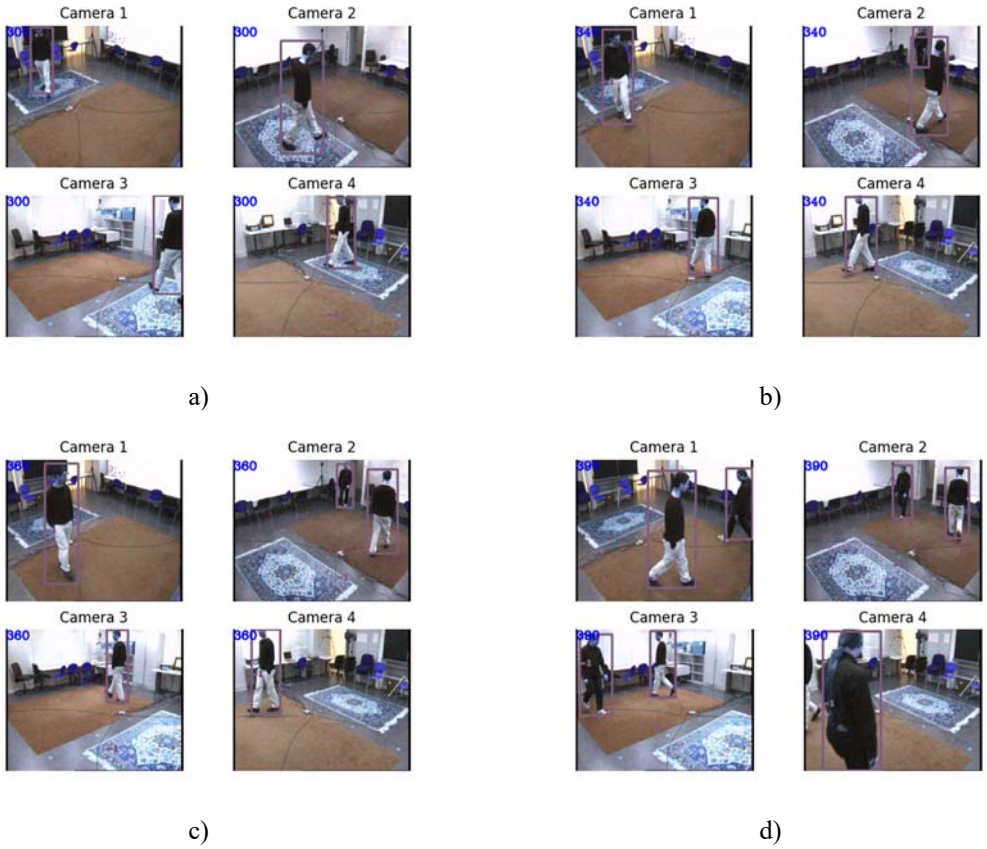


Figure 5. Results of tracking using the proposed Hybrid tracking model, a) Frame_300, b) Frame_340, c) Frame_360, and d) Frame_390

4.3Performance metrics

The metrics, such as MOTA, MOTP and Tracking error are evaluated to recognize and track the objects from the multi-view camera videos. The deep insight into the metrics is given by,

Multi-objective Transmission Precision (MOTP): MOTP is characterized as the standard distance between the anticipated object area and the ground-truth object area. The MOTP is mathematically expressed as,

$$MOTP = \frac{\sum Pos_{error}}{\sum mat_m} \tag{34}$$

where, $\sum Pos_{error}$ represents the total position error and $\sum mat_m$ represents the number of matches made.

Multi Objective Tracking Accuracy (MOTA): MOTA is characterized as the sum of ratio of misses is the sequences, ratio of false positive and the ratio of mismatches to the total number of object present in all frames. The MOTA is mathematically expressed as,

$$MOTA = 1 - \frac{\sum_t Fal_n + Fal_p + Id_s}{\sum_t Tot_o} \tag{35}$$

where, Fal_n represents the false negative, Fal_p represents false positive, Id_s represents identity switch, and Tot_o is the object present in frame.

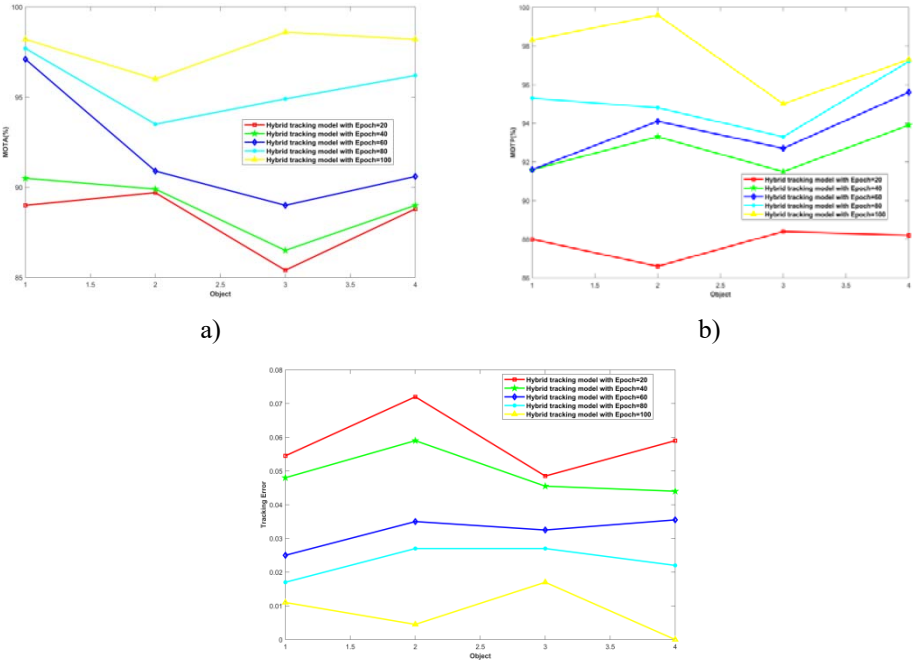
Tracking error: Tracking Error is defined as the difference between the actual position of the target and the estimated position of the target.

$$Track_{error} = Ac_p - Es_p \tag{36}$$

where, Ac_p is the actual position and Es_p is the estimated position

4.4Performance evaluation

The MOTA, MOTP, and tracking error are considered as the key parameters to execute the performance evaluation of the proposed hybrid tracking model and the discussion is elucidated in this section below. Figure 6 represents the performance evaluation of the proposed Hybrid tracking model, where a video frame contains up-to four objects. The performance evaluation in terms of MOTA with respect to the epoch is depicted in the Figure 6 a). The MOTA obtained by the proposed hybrid tracking model with four objects are 88.8 %, 89%, 90.6%, 96.2% and 98.2%, respectively in accordance with the epoch values of 20, 40, 60, 80 and 100.Thus, the maximum value of MOTA is obtained at the epoch of 100. From the figure, it is manifested that the MOTA increases with increase in the epoch. Figure 6 b) demonstrate the performance evaluation of the MOTP with respect to the epoch. When the epoch is set to be 20, the MOTP obtained by the proposed Hybrid tracking model is found to be 88%, 86.6%, 88.4% and 88.2 % for the number of objects 1,2, 3 and 4, respectively. The performance analysis of proposed Hybrid tracking model with respect to the tracking error is demonstrated in the Figure 6c). When the number of object is one, thetracking error obtained using the proposed Hybrid tracking model is 0.0545, 0.048, 0.025, 0.017 and 0.011. It is significantly reduced to 0.059, 0.044, 0.0355, 0.022 and 0 with 4 objects in the video frame.Thus, from the figure 6, it is proved that the performance of the proposed model is enhanced with respect to the epoch.



a) b)

c)

Figure 6. Performance evaluation of proposed tracking model with maximal of four objects in the video frame, a) MOTA, b)MOTP and c) Tracking error

The performance evaluation of the proposed hybrid tracking model with maximal of six objects in the video frame is demonstrated in the figure 7. The figure 7a) demonstrates the performance analysis of the proposed hybrid tracking model in terms of MOTA. When epoch is 100, the maximum MOTA for the proposed hybrid tracking model is 98.8% with 6 objects in the video frame. Figure 7b) illustrates the performance evaluation of the proposed tracking model in terms of MOTP. With an object and when the epoch is 100, the maximal MOTP is99.1% while fixing the objects in the video frame at 6. Figure 7 c) shows that the analysis of the proposed hybrid tracking model based on varying the epoch.The analysis is progressed through fixing the number of objects in the video frame to be 6. Figure 7 c) visualizes that the tracking error reduces with the increasing epochs and at the maximal epoch, the tracking error of the hybrid tracking model is 0.005 that is less when compared with the tracking error values of the epochs 20, 40, 60, and 80, respectively. Moreover, it is clear that the hybrid tracking model tracks the trajectories of the objects in the video-frames irrespective of the total objects in the video and locating the objects in the multi-frames is done using the deep-LSTM-based object prediction model.

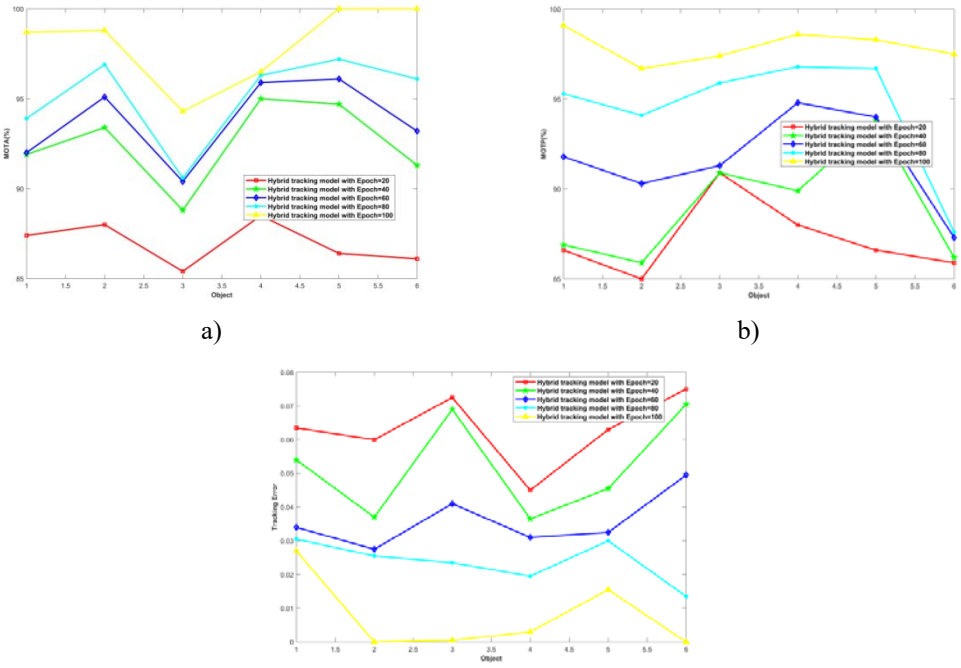


Figure 7. Performance evaluation of proposed tracking model with the maximal of 6 objects in the video frame, of a) MOTA, b)MOTP and c) Tracking error.

In figure 6 and figure 7, the analysis of the proposed hybrid tracking model with 4 and 6 persons in the multi-view images is demonstrated. The analysis is processed with respect to the epoch values based on the performance metrics. It is worth interesting to note that the performance of the hybrid tracking model is boosted with the higher number of epochs in proposed deep LSTM. The better performance is highlighted through the minimal value of the tracking error and maximal value of MOTA and MOTP with respect to the maximal epoch of 100, where the classifier trains effectively using the training data and renders an optimum result for the given test data within the minimal number of iterations.

4.5Competentmethods

The convenient methods utilized in the research to find the effectiveness of the proposed Hybrid tracking methods are Visual tracking(Yuan, et al., 2020), Deep LSTM(Farazi & Behnke, 2017), Ant colony optimization (ACO) (Dorigo, et al., 2006), Deep LSTM(Farazi & Behnke, 2017), Deep Sort(Wojke, et al., 2017) , and artificial bee colony optimization (ABC) (Karaboga & Ozturk, 2011) .

4.6 Comparative analysis of the tracking models

The comparative analysis enables to manifest the efficacy of the proposed Hybrid tracking model. For the comparative analysis, the existing methods, such as Visual tracking, Deep LSTM, ACO-Deep LSTM, Deep Sort ABC-Deep LSTM is utilized so that the efficacy of the proposed hybrid tracking model is revealed in terms of the performance measures. The comparative analysis of the hybrid tracking model with maximal offourobjects in the video frame is illustrated in the Figure 8.

Figure 8 a) illustrates the comparative analysis in terms of MOTA and from the figure, it is clear that the maximum MOTA of 98.5% is obtained by the proposed Hybrid tracking model when the number of the objects in the video-frame is considered to be 2. At the same time, the MOTA percentage obtained by the competent methods, such as Visual tracking, Deep LSTM, ACO-Deep LSTM, Deep Sort and ABC-Deep LSTM are 72.1%, 76.3%, 85.5%, 91.9%, and 98.5%, respectively with the same two objects in the video-frame. The comparative analysis in terms of MOTP is illustrated in the Figure 8 b).The maximum percentage of MOTP obtained using the proposed hybrid tracking model is 98.1% withmaximal of fourobjects in the video-frame, which is better when compared with the MOTP percentage corresponding to the existing tracking models, like Visual tracking, Deep LSTM, ACO-Deep LSTM, Deep Sort and ABC-Deep LSTM. The comparative analysis in terms of the Tracking error is depicted in the Figure8c).The minimum tracking error of 0.005 is obtained by the proposed Hybrid tracking model with four objects, while the tracking error obtained by the Deep sort algorithm is 0.043. Thus, it is clear that the proposed

hybrid tracking model provides a better tracking performance when compared to the existing models, likeVisual tracking, Deep LSTM, ACO-Deep LSTM, Deep Sort and ABC-Deep LSTM.

It is accepted that the better trade-off between the exploration and exploitation phases of the proposed search source-based deep LSTM and modified deep sort algorithm contributes a lot in rendering an effective tracking performance along with the visual tracking model. Moreover, the objects in the video-frames are accurately localized using the prediction based on deep LSTM, which optimally locates the objects in the frame and this criterion further boosts the accuracy of tracking.

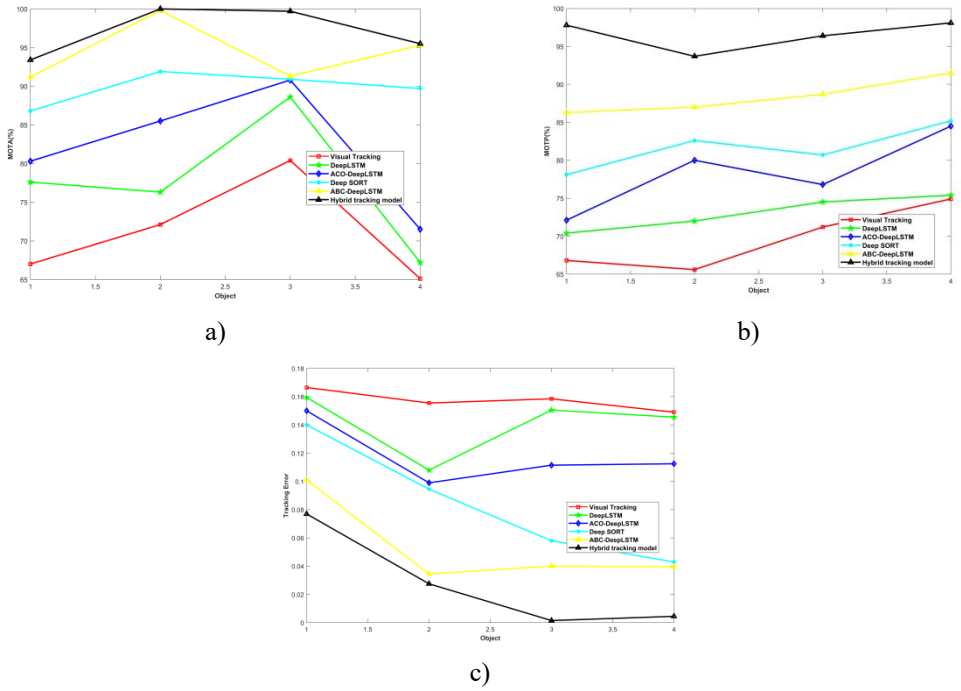


Figure 8. Comparative analysis of proposed tracking model with the maximal of four objects in the video frame,a) MOTA, b)MOTP and c) Tracking error.

The comparative analysis of the hybrid tracking model using the video-frames with six persons is illustrated in the figure 9. Figure 9 a) illustrates the comparative analysis in terms of MOTA and from the figure, it is clear that the maximum MOTA value of 98.5% is obtained by the proposed Hybrid tracking model with an object. At the same time, the MOTA obtained by the competent methods, like Visual tracking, Deep LSTM, ACO-Deep LSTM, Deep Sort and ABC-Deep LSTMwith an object is 68%, 73.7 %, 93.4%, 95%, and 98% respectively. It is peculiar to note that the proposed hybrid tracking model attains the performance improvement of 44.85 %, 33.64%, 5.46%, 3.68% and 0.51%, respectively when compared with the existing methods, such as Visual tracking, Deep LSTM, ACO-Deep LSTM, Deep Sort and ABC-Deep LSTM.The comparative analysis in terms of MOTP is illustrated in the Figure 9 b).The maximum value of MOTP obtained using the proposed hybrid tracking model is 98.7% with maximal of 6objects in the multi-view camera views. The proposed hybrid tracking model reports the percentage improvement of 2% with that of the conventionalDeep Sort and ABC-Deep LSTM models. The comparative analysis in terms of the Tracking error is depicted in the figure 9c).The minimum error of 0.0135 is obtained by the proposed Hybrid techniquewith an object, while the tracking error evaluated using the deep sort algorithm is 0.0675.

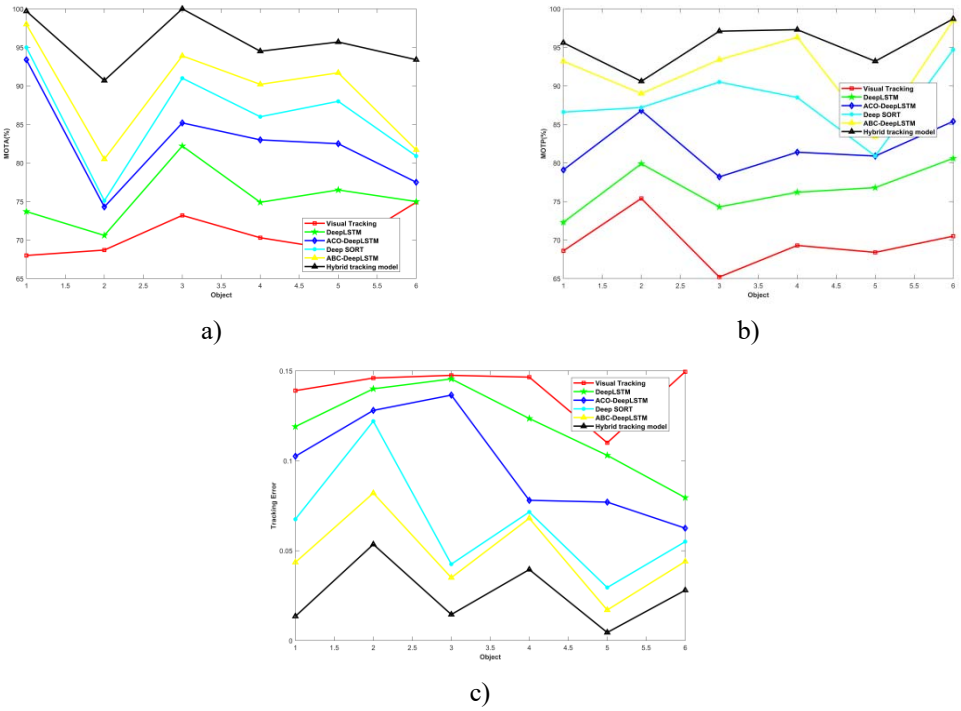


Figure 9. Comparative analysis of proposed tracking model with the maximal of 6 objects in the video frame, a) MOTA, b)MOTP and c) Tracking error.

4.7Comparative discussion

In this section, the analysis of the tracking methods based on the performance measures with two criterions of data. Table 1 depicts the comparative discussion of the tracking models with the four and six objects in the video frame. It is well-known from the contribution that the proposed hybrid tracking model is the combination of three tracking approaches, like visual tracking approach, deep learning-based approach, and deep sort-based approach, which renders an effective tracking experience due to the efficient way of the handling the motion of the video objects, effective diversification and intensification phases of proposed deep LSTM, and deep feature and similarity-based tracking experience of modified deep sort algorithm. The combined effect of these tracking models render the effective localization of the objects in the video-frames, and permit the accurate tracking of their trajectories. This phenomenon is explained through the comparative discussion. In table 1, the tracking performance of the methods when the video frame holds a maximal of four and six video objects is demonstrated. Table 1show that the proposed hybrid tracking model outperforms all the other conventional techniques in terms of MOTA, MOTP and Tracking error. The effective performance of the method is understood through the maximal values of MOTA and MOTP, while the tracking error to be minimal.

While considering the maximum number of object in the frame as 4, the optimal MOTA value obtained by the proposed Hybrid-tracking model is 98.5%. The optimal MOTA value obtained by the competent methods, such as Visual tracking, Deep LSTM, ACO-Deep LSTM, Deep Sort, ABC-Deep LSTM are 80.4%, 88.6%, 90.8%, 90.9% and 91.3%, respectively. Thus, the proposed hybrid model shows the performance improvement of 22.51%, 11.17%, 8.48%, 8.36% and 7.88% with respect to the comparative methods. The value of MOTP obtained by the proposed Hybrid tracking model for the maximum four objects in the frame is found to be 98.7%. At the same time, the MOTP obtained by the conventional methods, such as Visual tracking, Deep LSTM, ACO-Deep LSTM, Deep Sort, ABC-Deep LSTM are 74.9%, 75.4%, 85.2%, and 91.5%, respectively. The minimum tracking error obtained by the conventional methods, such as Visual tracking, Deep LSTM, ACO-Deep LSTM, Deep Sort, ABC-Deep LSTM for maximum four numbers of objects in the frames are 0.149, 0.1445, 0.1125, 0.043, and 0.0395, respectively. At the same time, the error obtained through the proposed Hybrid tracking model is 0.005, which is found to be the minimal tracking error.

On the other hand, when considering the maximum number of object in the frame as 6, the optimal MOTA value obtained by the proposed Hybrid-tracking model is 98.5%. The optimal MOTA value obtained by the convenient methods, such as Visual tracking, Deep LSTM, ACO-Deep LSTM, Deep Sort, ABC-Deep LSTM are 73.2%, 82.2%, 85.2%, 91.2%, and 93.9%, respectively. Thus, the proposed hybrid model shows the performance improvement of 34.5%, 19.82%, 15.61%, 8.24% and 4.89% with respect to the comparative methods. The values of MOTP obtained by the proposed Hybrid tracking model for the maximum of six objects in the frame are found to be 98.7%. At the same time, the MOTP obtained by the conventional methods, such as Visual tracking, Deep

LSTM, ACO-Deep LSTM, Deep Sort, ABC-Deep LSTM are 70.5%, 80.6%, 85.4%, and 94.7%, respectively. The minimum tracking error obtained by the conventional methods, such as Visual tracking, Deep LSTM, ACO-Deep LSTM, Deep Sort, ABC-Deep LSTM for maximum six objects in the frames are 0.11, 0.103, 0.077, 0.0295 and 0.017, respectively. At the same time, the error obtained through the proposed Hybrid tracking model is 0.005.

PTable 1.Comparative discussion of the tracking models

Methods	**Maximal of four objects in the Video frames**			**Maximal of six objects in the Video frames**		
	MOTA	**MOTP**	**Tracking error**	**MOTA**	**MOTP**	**Tracking error**
Visual tracking	80.4%	74.9%	0.149	73.2%	70.5%	0.11
Deep LSTM	88.6%	75.4%	0.1445	82.2%	80.6%	0.103
ACO-Deep LSTM	90.8%	84.5%	0.1125	85.2%	85.4%	0.077
Deep Sort	90.9%	85.2%	0.043	91%	94.7%	0.0295
ABC-Deep LSTM	91.3%	91.5%	0.0395	93.9%	98.5%	0.017
Proposed Hybrid model	**98.5%**	**98.1%**	**0.005**	**98.5%**	**98.7%**	**0.005**

5. Conclusion

In this research, a hybrid object tracking model is proposed to restrain the issues, like illumination and occlusions for effective recognition of objects in the video surveillance. In this proposed hybrid model, the object in the frames are tracked using the hybrid tracking model based on the rectangular bounding box, proposed search source-based Deep LSTM, and modified deep SORT algorithm such that the accuracy of tracking is enhanced. The proposed search space algorithm is utilized in this hybrid model to train the Deep LSTM and in turn, establishes an effective trade-off between the exploration and exploitation phases. The modified deep sort algorithm is used to solve the assignment problem to incorporate the motion information, which in turn increases the tracking efficiency. The performance evaluation and comparative analysis are executed to demonstrate the effectiveness of the proposed Hybrid model. The proposed Hybrid model shows impressive results in terms of MOTA, MOTP and the tracking error and the obtained values are 100%, 98.1% and 0.0135 respectively, which exceeds all the other conventional methods. The future dimensions will be based on the effective object tracking based on purely meta-heuristic search methods, which may impact much of the performance of tracking.

References

[1] Akhlaq, M., Sheltami, T.R., Helgeson, B., Shakshuki, E.M. (2012). Designing an integrated driver assistance system using image sensors. *Journal of Intelligent Manufacturing*, **23**(6), pp.2109–2132.
[2] CAVIAR4REID dataset, "https://lorisbaz.github.io/caviar4reid.html", Accessed on November 2020.
[3] Chen, D.Y., Hsieh, P.C. (2012). Face-based gender recognition using compressive sensing. *in International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS), IEEE ,* pp.157–161.
[4] Chen, X., Wu, S and Yu, Z. (2020). Self-Enhanced R-CNNs for Human Detection with Semi-Supervised Assumptions, *IEEE Access,* **8**, pp.15132-15143.
[5] Chowdhry, D., Paranjape, R., Laforge, P. (2015). Smart home automation system for intrusion detection. *In14th IEEE Canadian Workshop on Information Theory (CWIT),* pp.75–78.
[6] Damotharasamy, S. (2020). Approach to Model Human Appearance Based on Sparse Representation for Human Tracking in Surveillance. *IET Image Processing,* **14**(11), pp.2383-2394.
[7] Dilawari, A., Khan, M.U.G., ur Rehman, Z., Awan, K.M., Mehmood, I and Rho, S. (2020). Toward generating human-centered video annotations. *Circuits, Systems, and Signal Processing,* **39**(2), pp.857-883.
[8] Dorigo, M., Birattari, M and Stutzle, T. (2006). Ant colony optimization. *IEEE computational intelligence magazine,* **1**(4), pp.28-39.
[9] Eshel, R., Moses, Y. (2008). Homography based multiple camera detection and tracking of people in a dense crowd. *in IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp.1–8.
[10] Farazi, H and Behnke, S. (2017). Online visual robot tracking and identification using deep LSTM networks. *In proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* pp.6118-6125.
[11] Gajjar, V., Gurnani, A and Khandhediya, Y. (2017). Human detection and tracking for video surveillance: A cognitive science approach. *In Proceedings of the IEEE International Conference on Computer Vision Workshops,* pp.2805-2809.
[12] Gamage, G., Sudasingha, I., Perera, I and Meedeniya, D. (2018). Reinstating Dlib Correlation Human Trackers under Occlusions in Human Detection based Tracking. *In 18th IEEE International Conference on Advances in ICT for Emerging Regions (ICTer),* pp.92-98.
[13] Karaboga, D and Ozturk, C. (2011). A novel clustering approach: Artificial Bee Colony (ABC) algorithm. *Applied soft computing,* **11**(1), pp.652-657.
[14] Karpagavalli, P and Ramprasad, A.V. (2020). Automatic multiple human tracking using an adaptive hybrid GMM based detection in a crowd. *Multimedia Tools and Applications*, **79**(39), pp.28993-2901.
[15] Lai, H.E., Lin, C.Y., Chen, M.K., Kang, L.W., Yeh, C.H. (2013). Moving objects detection based on hysteresis thresholding. *in Advances in Intelligent Systems and Applications*, **2**, pp.289–298.

[16] Li, T., Chen, H., Sun, S & Corchado, J. M. (2017). Joint smoothing, tracking, and forecasting based on continuous-time target trajectory fitting. *IEEE Transactions on Automation Science and Engineering*, **6**(1), pp.1-16.
[17] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Doll´ar, P and L. Zitnick, C. (2014). Microsoft coco: Common objects in context. *In Proceedings of the European Conference on Computer Vision,* pp.740–755.
[18] Liu, J., Hao, K., Ding, Y. (2017). Moving human tracking across multi-camera based on artificial immune random forest and improved color-texture feature fusion. *The Imaging Science Journal,* **65**(4), pp.239–251.
[19] Liu, L., Jiao, Y and Meng, F. (2020). Key Algorithm for Human Motion Recognition in Virtual Reality Video Sequences Based on Hidden Markov Model. *IEEE Access,* **8**, pp.159705-159717.
[20] Mozerov, M., Amato, A., Roca, X and Gonzàlez, J. (2009). Solving the multi object occlusion problem in a multiple camera tracking system. *Pattern Recognition and Image Analysis,* **19**(1), pp.165-171.
[21] Multi-view Multi-class Detection dataset, "https://www.epfl.ch/labs/cvlab/data/data-multiclass/", Accessed on November 2020.
[22] Nieto, R.G and Restrepo, H.D.B. (2020). Quality aware feature selection for video object tracking. *Electronic Imaging,* **2020**(9), pp.169-1-169.
[23] Ran, Y., Zheng, Q., Chellappa, R., Strat, T.M. (2010). Applications of a simple characterization of human gait in surveillance. *IEEE Transactions on Systems, Man, and Cybernetics—Part B,* **40**(4), pp.1009–1020.
[24] Reddy, K.K., Shah, M. (2013). Recognizing 50 human action categories of web videos, *Machine Vision and Applications,* **24**(5), pp.971–981.
[25] Shao, S., Zhao, Z., Li, B., Xiao, T., Yu, G., Zhang, X and Sun, J. Crowdhuman. (2018). A benchmark for detecting human in a crowd. :arXiv preprint arXiv:1805.00123.
[26] Thome, N., Miguet, S., Ambellouis, S. (2008). A real-time, multi-view fall detection system": a LHMM-based approach. *IEEE Transactions on Circuits and Systems for Video Technology,* **18**(11), pp.1522–1532.
[27] Walia, G.S., Kumar, A., Saxena, A., Sharma, K and Singh, K. (2020). Robust object tracking with crow search optimized multi-cue particle filter. *Pattern Analysis and Applications,* **23**(3), pp.1439-1455.
[28] Wang, X., Shen, C., Li, H and Xu, S. (2019). Human Detection Aided by Deeply Learned Semantic Masks. *IEEE Transactions on Circuits and Systems for Video Technology,* **8**(8), pp.2663-2673.
[29] Wang, Y., Choi, J., Zhang, K., Huang, Q., Chen, Y., Lee, M.-S and JayKuoa, C.-C. (2020). Video object tracking and segmentation with box annotation. *Signal Processing: Image Communication,* **85**, pp.115858.
[30] Wojke, N., Bewley, A and Paulus, D. (2017). Simple online and real time tracking with a deep association metric. *In proceedings of IEEE international conference on image processing (ICIP)*, pp.3645-3649.
[31] Wu, B and Nevatia, R. (2007). Detection and tracking of multiple, partially occluded humans by Bayesian combination of Edgelet based part detectors. *International Journal of Computer Vision,* **75**(2), pp.247-266.
[32] Yuan, D., Chang, X and He, Z. (2020). Accurate bounding-box regression with distance-IoU loss for visual tracking.
[33] Zhang, Y. (2019). Detection and Tracking of Human Motion Targets in Video Images Based on Cam shift Algorithms, *IEEE Sensors Journal,* **20**(20), pp.1187-11893.
[34] Zhang, Y., Zhang, M., Cui, Y and Zhang, D. (2020). Detection and tracking of human track and field motion targets based on deep learning. *Multimedia Tools and Applications,* **79**(13), pp.9543-9563.