



Support vector regression and extended nearest neighbor for video object retrieval

C. A. Ghuge¹ · Sachin D. Ruikar² · V. Chandra Prakash¹

Received: 22 May 2018 / Revised: 3 September 2018 / Accepted: 18 September 2018 / Published online: 28 September 2018
© The Author(s) 2018

Abstract

Video retrieval is one of the emerging areas in video capturing that gained various technical advances, increasing the availability of a huge mass of videos. For the text or the image query given, retrieving the relevant videos and the objects from the videos is not always an easy task. A hybrid model was developed in the previous work using the Nearest Search Algorithm (NSA) and exponential weighted moving average (EWMA), for the video object retrieval. In NSA + EWMA, the object trajectories are retrieved based on the query specific distance. This work extends the previous work by developing a novel path equalization scheme for equalizing the path length of the query and the tracked object. Initially, a hybrid model based on Support Vector Regression and NSA tracks the position of the object in the video. The proposed density measure scheme equalizes the path length of the query and the object. Then, the identified path length related to the query is given to extended nearest neighbor classifier for retrieving the video. From the simulation results, it is evident that the proposed video retrieval scheme achieved high values of 0.901, 0.860, 0.849, and 0.922 for precision, recall, F-measure, and multiple object tracking precision, respectively.

Keywords Video retrieval · Support vector regression · Path length equalization · Extended nearest neighbor · CAVIAR database

1 Introduction

In recent years, video surveillance has seen a progressive development in various fields, such as driving assistance, human–computer interaction, augmented reality, and so on [1]. In the computer vision applications, numerous data is collected for detecting the moving objects in the video frame. Surveillance cameras serve as a key link towards the security system, and they provide massive videos. For efficient implementation of the security system, it is necessary to track and retrieve objects and their corresponding trajectory path [2]. Video retrieval algorithms find more suitability in important security applications as they identify the track of moving objects from frame to frame [3]. Tracking

becomes difficult as the video contains a large number of rigid objects. For achieving improved tracking performance, it is necessary to improve the robustness of visual tracking [4]. In sports videos, abrupt video cut may lead to abrupt motion, which leads to a sudden change in position, speed, and direction of the target object [5]. The results achieved by detection schemes affect the performance of the tracking mechanism as both the detection and the tracking are inter-related [6]. Video retrieval strategy can be categorized as text-based retrieval and object-based retrieval. The first technique fails in scenarios when there is a no-textual description of target object [7].

Object tracking is the process of determining the positions and other significant information of moving objects in image sequences. Object retrieval from videos undergoes two major processes, listed as (1) detecting and tracking the position of objects and (2) defining object descriptors for feature extraction and retrieving the objects [2]. In [8], contour based schemes are developed for tracking the position of the object. Pattern classification technique is commonly employed for detecting the position of objects. It employs a classifier for differentiating the position of objects

✉ C. A. Ghuge
caghugekl@gmail.com

¹ Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India

² Department of Electronics Engineering, Walchand College of Engineering, Sangli, Maharashtra, India

in their respective frames. Commonly used classifiers for object detection and recognition are support vector machine (SVM), AdaBoost classifiers, etc. Classifiers defined for object detection should be designed with good classification accuracy, and it must consider the time constraint. For dealing with the non-rigid structures in the video, Gaussian mixture model (GMM) has been utilized as it provides good applicability. Besides various advantages, the GMM model fails to model the videos with the noisy and non-stationary background. Object detection phase concentrates on detecting the exact location of the object in the video frame. Detection of trajectory path through object detection gets affected due to various noise factors, such as illumination, slow-moving objects, shadows and other phenomena. Object detection mainly carries out noise reduction such that the retrieval process can be made more efficient [9]. One of the commonly used tracking mechanisms is the region-based target tracking. Region-based tracking separates the foreground and the background and treats the path of target especially. Besides, it models the background and detects the motion of the object [4]. SIFT-based features for tracking the objects in the video have achieved significant results. The pixel-based matching and feature-based matching achieved high speed and accuracy during the object tracking [10].

Researchers involved in literature can be categorized as video retrieval, tracking objects in the video, trajectory matching, object appearance matching and so on [7]. In [1], visual tracking scheme has been used for tracking the objects. Visual tracking scheme [1] finds the position of the target object in consecutive frames of the object and finds its trajectory or moving path. One of the major challenges involved in visual tracking scheme is detecting the target object in the video prevailed with long-term occlusions. Further, during the video processing schemes, the video may lose quality due to the rotation, and geometric deformation [1]. Techniques, such as region-based [11], feature-based [12], model-based [13], and active contour-based techniques [14] achieved significant results in visual object tracking [1]. Several works have opted for multi-object tracking to estimate the position of the object in the video frame during tracking visual impairments, such as position, size, identification, etc. [15]. Other key problem faced by video retrieval mechanism through detection by tracking, is the occurrence of association between the visual measurements and multiple objects in the video [15]. Some techniques make use of appearance-based features for video retrieval. The presence of background clutters in the video affect the learning of appearance-based features, and thus, reduces the performance of the tracking mechanism [16]. In [17], object tracking is done by defining the optical flow, and temporal-spatial context. In [18], object tracking was done with the help of daubechis complex wavelet transform and Zernike moment. Fusing several tracking algorithms also provides significant

results. In [19], an online fusion tracking method has been employed for single object tracking. Discriminative trackers, such as SVM and boosting classifiers distinguish the target object as the binary classification task [20]. Optimization algorithms are used for object retrieval to obtain the best performance [21].

This work extends the previous work [22] for object retrieval and tracking from the videos. Previously, video retrieval was performed by defining the hybrid model based on EWMA and the NSA [23] model. Also, the path length between the query and the length of the target object is equalized with the help of novel QSD. As an extension of this method, in the second work, the video object retrieval is performed using SVR [24], along with a classifier. The proposed work is developed in three stages, namely object tracking, path length equalization, and retrieval. Initially, object tracking is performed to find the path of the object by combining spatial and visual tracking approaches. The spatial tracking is done based on SVR, whereas the visual tracking follows the NSA model. In the path length equalization, the path length of the query and that of the tracked object are equalized by sample selection based on the density measure. Finally, the video is retrieved in the retrieval phase, using ENN classifier [25].

The major contribution of this paper is the development of a novel path equalization scheme based on the density measure. The proposed path equalization scheme refines the retrieval process, by equalizing the path length of the query and thereby, tracks the object.

The structure of the paper is organized as follows: Sect. 1 deals with the video retrieval process, and techniques involved in developing the video retrieval system. Various works contributed towards the video retrieval strategy has been discussed in Sect. 2. Section 3 deals with the proposed path length equalization scheme and the ENN based retrieval strategy. Section 4 depicts the results achieved by the proposed density measure scheme for achieving path length equalization, and the simulation results by taking the video clips from CAVIAR database. Section 5 concludes this paperwork.

2 Motivation

2.1 Literature survey

This section presents eight literary works dealing with object tracking and retrieval from the videos.

Lai and Yang [7] presented the video retrieval system for reducing the complexity issues prevailing in tracking the location of the object. The video retrieval is done by considering 3D graphical user interface and trajectory of the objects in the frame. The proposed scheme considered

more intuitive interactions between the frames for retrieving suitable video contents. Wang et al. [2], proposed the video retrieval scheme for indexing and retrieving objects. The scheme concentrated on retrieving the video contents of possible interest from large-scale video surveillance system. The scheme improved the retrieval performance by encoding the deep features into short binary codes. Since the scheme used multi deep features for learning, the video retrieval is done significantly.

Ding et al. [26] proposed the Surveillance surfing (Surv-Surf) technique, for retrieving the moving objects in the video frame. The technique exploited the characteristics of big data for video retrieval, and hence, used big data processing tools. The authors have proposed motion information for segmenting the video contents. Further, the MapReduce framework eliminates the challenges posed by big data. This framework concentrated on specified areas of the frame rather than the entire frame, and thus, achieved improved video retrieval. Zhang and Jeong [3] proposed the video retrieval scheme for surveillance of airport applications. A retrieval algorithm was designed for detecting the moving objects in the video frame, by adopting the Harr face cascade classifier for the classification. The model was suitable for implementation in a cloud platform.

Kanagamalliga and Vasuki [8] presented the retrieval algorithm with the help of the Optical Flow and Gabor Features Based Contour Model. The model performed both the object detection and motion analysis through the algorithm. The model utilized the AdaBoost classifier for the classification. The model failed to estimate the non-rigid objects in the video frame.

Joy and Peter [1] introduced the new color-independent tracking approach, the contributions of which are threefold. First, the illumination level of the sequences was maintained constant using fast discrete curvelet transform. Then, Fisher information metric was calculated based on a cumulative score by comparing the template patches with a reference template at different timeframes. This metric was used for quantifying distances between the consecutive frame histogram distributions. Then, an iterative algorithm, called conditionally adaptive multiple template update, was proposed to regulate the object templates for handling dynamic occlusions effectively.

Li Liang-qun, [15] presented the video retrieval strategy with the fuzzy logic data association algorithm. The algorithm had the characteristics of a fuzzy inference system and thus, allowed multi-object tracking. Moreover, it tackled the threats posed by long-term occlusions, by using track-to-track association approach. Bency et al. [16] developed a methodology for retrieving the video contents, and the retrieval strategy depends on the leverage human knowledge. For the retrieval, the model generates a document representing the motion information prevailing in

the video. Then, the document was listed against the videos in library content, and the necessary video files were extracted. The scheme does not require the object detectors for locating the position of objects in the video.

2.2 Challenges

Since the era of video technology, various challenges prevailing in the video object tracking makes the retrieval process to be difficult. Developing an environment for retrieving the movement of objects from one frame to other faces several challenges and they are listed below.

- Retrieving objects from videos consumes more time, and faces overhead space issues since the surveillance camera produces high resolution videos [26].
- Also, operations, such as object retrieval and action classification tend to consume more time as video sequences are mostly noisy, non-segmented and multi-dimensional [26].
- Even though several algorithms have been developed in the literature, a problem occurring due to the presence of non-rigid structures in the video has been still unaddressed [8].
- Noise reduction during the object detection model significantly reduces the visualization effects of video [9].
- Some of the particle challenges faced during object tracking are, object to object occlusion, object to scene occlusion, abrupt object motion, various lighting conditions, etc. [15].
- The appearance of the object in the frame to its successive frames shows negligible changes, and thus, tracking the object of the path may become difficult. Also, concentrating only on features of objects during video retrieval, may produce adverse effects in tracks, as some algorithms do not segment the object from its background. Thus, for improved tracking performance, considering both the object and the background features make the algorithm efficient [17].

The proposed method is designed to solve the above challenges in video object retrieval. The proposed method utilizes SVR and NSA to track the position of the object. SVR acknowledges the occurrence of non-linearity in the data and offers an excellent prediction model. It utilizes the regularization parameter, which avoids the object occlusion and it is more robust. Also, this paper introduces a novel density measure technique for equalizing the path length of both the query and the objects in the video frame. Making the path length of the object related to query simplifies the video retrieval process and consumes less time.

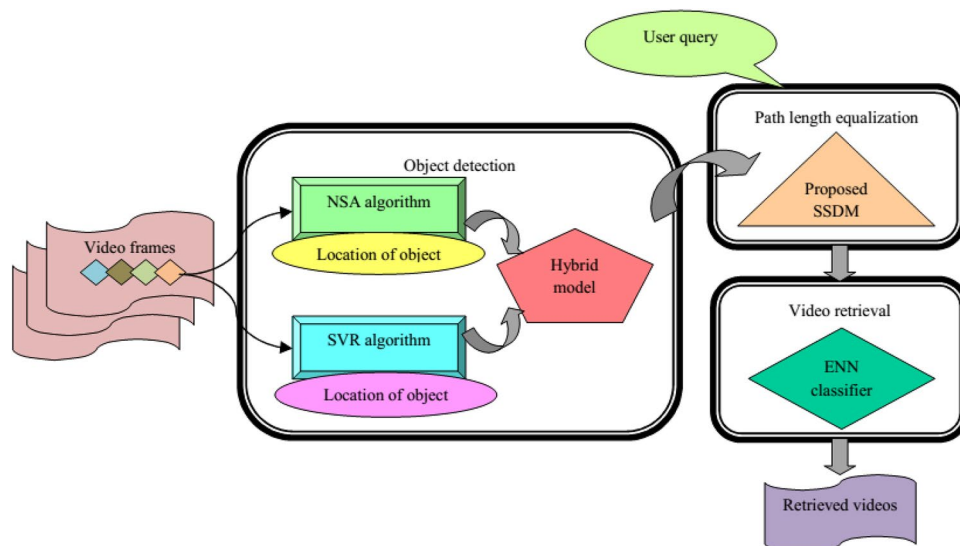
3 Video retrieval with the proposed SSDM based path equalization and ENN based retrieval strategy

This section presents the proposed video retrieval strategy and the design procedure for the newly developed Sample Selection based Density Measure (SSDM) technique in video retrieval scheme and the architecture is depicted in Fig. 1. As given in Fig. 1, the target objects within the frame are subjected to path tracking process. For retrieving suitable video frame contents, the path tracking is necessary, and this paper uses a hybrid model for path tracking. The hybrid model comprises of NSA algorithm and SVR approach. After tracking the path of objects, the length of path tracked by the tracking mechanism and query need to be matched. Path length equalization acts as a major step towards video retrieval since the video contents more related to user query need to be retrieved. This work introduces the SSDM technique for path length equalization. After obtaining the equalized path related to query, ENN classifier retrieves the suitable video frames.

The entire operation involved in the proposed video retrieval scheme is presented here. Initially, the video is subjected to framing, and the keyframes in the video are extracted. Then, the hybrid model with the NSA and the SVR identifies the position of the object in the keyframe. The tracked path from the video frame is provided to the ENN module for video retrieval. Designing video retrieval related to user query can be done through three important steps,

- Object Path tracking
- Path length equalization
- Video retrieval related to the user query

Fig. 1 The architecture of video retrieval scheme with the proposed SSDM technique and ENN classifier



Consider the video R with Z frames for the analysis. For retrieving the videos related to the query, it is necessary to extract the keyframes from the video. Video considered for the analysis is represented as follows,

$$R = \{U_i; 1 \leq i \leq Z\}, \quad (1)$$

where, U_i refers to the i th frame present in the video, and the value of i varies between 1 and Z .

3.1 Object tracking using the hybrid model for tracking the position of the object

The initial step involved in video retrieval is tracking the movement of the object from one frame to another. For the object detection purpose, this paper uses the hybrid tracking model. The hybrid tracking model uses techniques, such as NSA [23] and SVR [24] and hybridizes the results achieved by both the algorithms for object tracking. NSA is one of the commonly used techniques for tracking the position of the object, due to its simple characteristic, and stability. NSA tracks the path between the objects through the Euclidean distance measure. NSA fails to identify the unknown regression prevailing in the video, and it can be tracked with the help of SVR approach. SVR identifies a nonlinear estimate, by integrating the linear estimate with the nonlinear function.

3.1.1 Tracking the position of the object based on the NSA model

The first algorithm used for the hybrid approach is the NSA. The model identifies the trajectory path of the target object by detecting the best location in the current frame concerning the location of the object in the next reference. Here, the position calculation depends on the minimal

Euclidean distance between the successive frames. The steps involved in identifying the exact location of the object in video frame through the NSA algorithm are given in the following steps:

3.1.1.1 Localization of target object NSA approach identifies the exact location of the object by initially localizing various objects in the frame. Localization refers to the identification of the key points of the object, which helps in tracking the position of the object in successive frames. As expressed in Eq. (1), the video has several frames, and the frames can be individually represented by the following equation,

$$U = \{U_1, U_2, \dots, U_i, \dots, U_Z\} \tag{2}$$

where, U_i refers to the i th frame in the video and the value of i extends up to Z . The frame has the collection of objects located at a different position. Consider the i th video frame that has A number of objects. The objects in the frame have different positions, and also follow different paths in a successive frame. Thus, the trajectory path followed by the object can be represented as a vector of dimension $(2 \times A)$. The position of the o th object in the frame U_i is represented as follows,

$$G_o^i = (u_o^i, v_o^i), \tag{3}$$

where, G_o^i indicates the position of o th object in the frame U_i , and the term (u_o^i, v_o^i) indicates the x th and y th elements representing the position G_o^i .

3.1.1.2 Determining the position of the object in the next frame In this step, the location of the successive frame is identified by fixing the rectangular window. The rectangular window helps in identifying the location of the object in the next frame by fixing an extensive parameter α . Based on the extensive parameter, the key position of the object is extended in every possible direction. After locating the position of the object, the distance between the position of the object in the current frame and successive frame is measured. Here, the value of the extensive parameter providing the minimum distance is considered for position determination. Following are the steps briefed for identifying the position of the object in the next frame.

(a) *Fixing the rectangular bound window*: The key position of the object is determined by fixing the rectangular window, which again depends on an extensive parameter α . For the random value of α , the rectangular window is extended in every possible direction. Then, the distance is calculated with the use of the extensive parameter. Then, the value providing minimal distance is fixed as an extensive parameter, and the rectangular window is constructed from newly found extensive parameter.

(b) *Determining the position of the object in the successive frame*: The position of the object in the successive frame is identified once a minimum value is fixed for the threshold α . Based on the fixed value of the extensive parameter α , the new position of the object in successive frames is identified, and it is represented as follows,

$$G_{o+1}^i = (u_o^i \pm \alpha, v_o^i \pm \alpha), \tag{4}$$

where, α indicates the extension parameter, which is a constant value.

3.1.1.3 Calculation of distance between the objects in the successive frame The next step in NSA is determining the distance between the objects in the successive frame. The distance calculation is done based on the position of the object in the current and next frame. The expression for the distance between the positions of the object in the successive frame is represented as follows,

$$\text{Dist}(G_o^i, G_{o+1}^i) = \text{Dist} \sqrt{(G_o^i, G_{o+1}^i)}. \tag{5}$$

It is necessary to fix the value of α , which provides minimal distance.

3.1.1.4 Determining the position of objects with NSA In the final step, the steps two and three are successively repeated until positions of all objects in the video frame is calculated. The position of the object as retrieved by NSA scheme is expressed as,

$$P_{NSA}^o = \{G_o^i, G_{o+1}^{i+1}, \dots, G_{o+A}^{i+Z}\}, \tag{6}$$

where, $G_o^i, G_{o+1}^{i+1}, \dots, G_{o+A}^{i+Z}$ indicate the location of the objects identified through NSA scheme.

3.1.2 Tracking the position of the object based on Support vector regression

Including the SVR for tracking the object’s location in the video frame improves the quality of the tracking process. SVR technique uses the mapping function, which maps the linear estimate with the nonlinear function and thereby, finds the unknown regression.

Thus, for finding the accurate position of the object in the video frame, the SVR generates the training set containing the input–output pair given as,

$$\zeta = \{(p_1, q_1), (p_2, q_2), \dots, (p_Z, q_Z)\}, \tag{7}$$

where, ζ indicates the input and output space and (p_1, q_1) indicates the input–output pair. For generating the suitable input–output pair, SVR uses the optimization problem. The SVR model generates the inputs through a training set, and

the selection acts as an optimization problem. The condition for the optimization criteria is suggested below:

$$\min_{n,b,\xi+/-} \left(\frac{1}{2} \|n\| + Q \sum_{i=1}^z (\xi_i^+ + \xi_i^-) \right), \tag{8}$$

where, Q indicates the cost and z indicates the total number of input and output pair. The optimization problem can also be expressed as dual maximization problem, which is indicated as,

$$\max_{\beta^+,\beta^-} -\frac{1}{2} \sum_{i,o} \{ (\beta_i^+ + \beta_i^-) (\beta_o^+ + \beta_o^-) Y(p_i.p_j) \} - \delta \sum_i (\beta_i^+ + \beta_i^-) + \sum_i q_i (\beta_i^+ + \beta_i^-). \tag{9}$$

The regression estimate for the dual maximization problem is expressed as follows,

$$w(x) = \sum_i (\beta_i^+ + \beta_i^-) Y(p.p_i) + c, \tag{10}$$

where, c indicates the deviation parameter. The final path estimated based on SVR is expressed as,

$$P_{SVR}^o = \{ G_o^i, G_{o+1}^{i+1}, \dots, G_{o+A}^{i+Z} \}. \tag{11}$$

3.1.3 Hybrid model for object tracking

Here, the hybrid model utilizes the advantages of both the SVR and the NSA schemes and thereby, achieves object tracking. The object tracking algorithms discussed in the literature have their pros and cons, and hence, for leveling effects of one algorithm, hybridizing the results of one improves the overall performance. In the hybrid model, the results obtained with NSA and SVR schemes are averaged, and the final path of the object is tracked. The expression for the path tracked by the hybrid model is expressed as,

$$T_o = [\{ P_{NSA}^o \} + \{ P_{SVR}^o \}] \quad \forall o, \tag{12}$$

where, P_{NSA}^o , and P_{SVR}^o indicate the path tracked by NSA and the path tracked by SVR approach, respectively. Hybridization of results achieved from both NSA and SVR improves the accuracy of tracking performance, and it further refines the search process.

3.2 Path length equalization using the proposed SSDM technique

Next major process in video retrieval is the path length equalization. The paths tracked by NSA + SVR hybridization approach may differ from the path provided by the user query. For making the video retrieval process more efficient, it is necessary to identify the tracks matched with the user query. For identifying the suitable tracks related to the query,

path length equalization is necessary. One of the important strategies involved in video retrieval is the path length equalization. The query given by the user to the video retrieval system may have a different size than the path tracked from the video. Hence, for retrieving the actual video contents related to the query, it is necessary to equalize the path of the query and the tracked object. While the query arrives at the video retrieval system, it is necessary to retrieve the tracks from the video equivalent to the query such that the accuracy of the retrieval system can be improved. For this purpose,

this paper introduces a novel density measure technique for equalizing the path length of both the query and the objects in the video frame. Figure 2 shows the block diagram of the proposed SSDM scheme for path length equalization.

As depicted in the figure, the length of the query path and the object is subject to various modifications for achieving the equalized length. Making the path length of the object related to query simplifies the video retrieval process, as path length equalization alters the grid size of the objects similar to the user query.

While the query arrives in the video retrieval system, the paths identified by the hybrid tracking model are compared with the paths of the query. Then, the matching process is subjected to neighborhood calculation. Here, both the query M_{uv} and tracked paths T_{rv}^o have different vector sizes, and it is represented as follows,

$$M_{uv} \Rightarrow \left\{ \begin{array}{l} 1 \leq u \leq k \\ 1 \leq v \leq 2 \end{array} \right\}; \quad T_{rv}^o \Rightarrow \left\{ \begin{array}{l} 1 \leq r \leq l \\ 1 \leq v \leq 2 \end{array} \right\}, \tag{13}$$

where, k and l indicate the respective size of the query and the tracked object vector. The vector of both the query and the tracked path takes the two dimensional format. Based on the query and the tracked path, neighborhood calculation J_{rv}^o is done. The expression for the neighborhood calculation is given as follows,

$$J_{uv}^o = \text{Arg Min}_{r,v} \left(\sqrt{\sum_{v=1}^2 (T_{rv}^o - M_{uv})^2} \right) \quad \forall \{1 \leq r \leq F\}. \tag{14}$$

The neighborhood calculation vector is expressed as follows,

$$J_{uv} \Rightarrow \left\{ \begin{array}{l} 1 \leq u \leq l \\ 1 \leq v \leq 2 \end{array} \right\}. \tag{15}$$

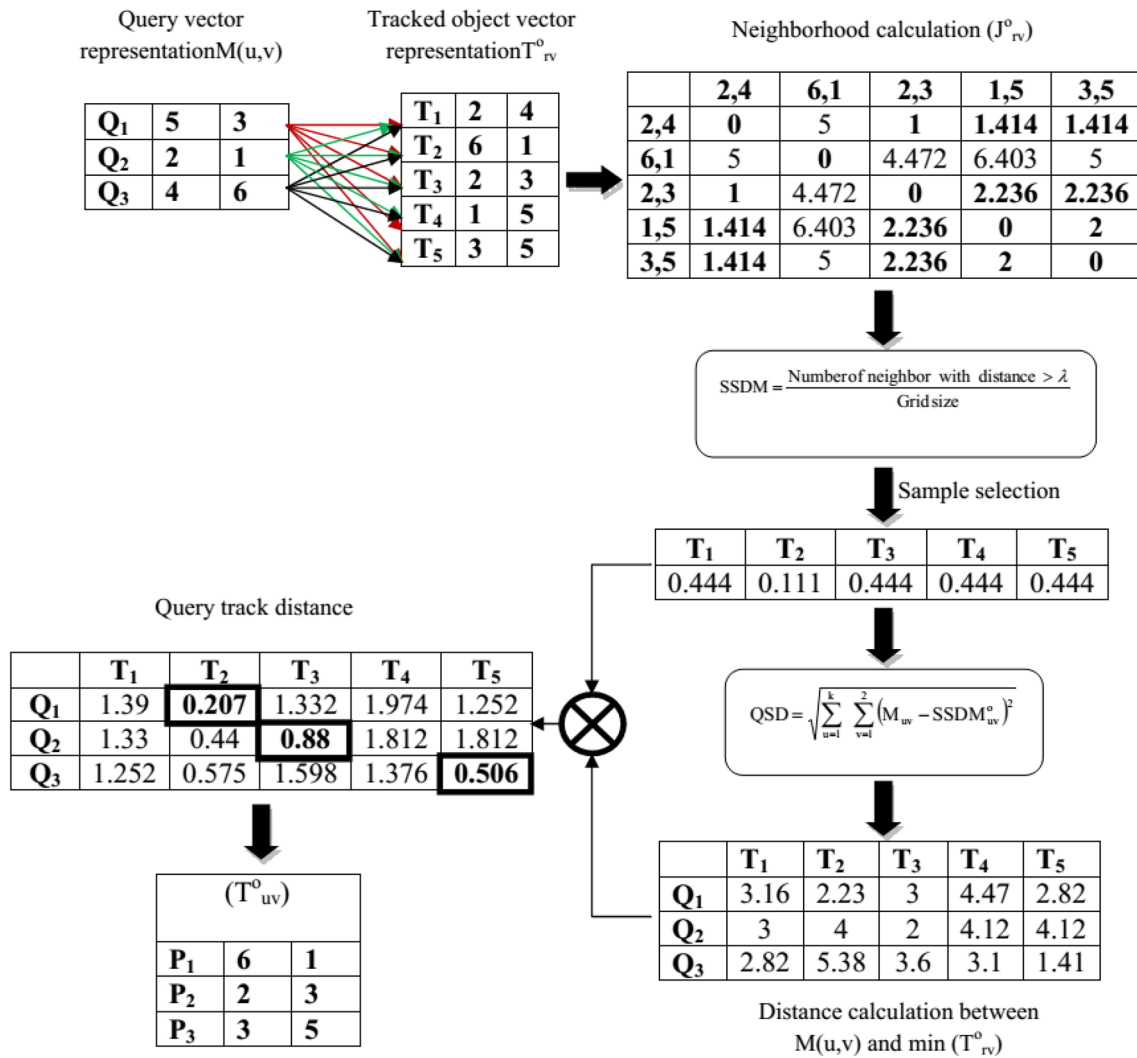


Fig. 2 Proposed SSDM for path length equalization

After the neighborhood calculation J_{rv}^o , the proposed SSDM measure is calculated. The proposed SSDM measure finds the suitable samples appropriate for video tracking by fixing a minimum threshold value λ . The expression for the proposed SSDM measure is expressed as follows,

$$SSDM = \frac{\text{Number of neighbor with distance } > \lambda}{\text{Grid size}} \quad (16)$$

where, λ indicates the minimal threshold set for sample selection. As mentioned in the above expression, the SSDM creates a corresponding value for each element by identifying the neighbor with a minimum threshold λ to the grid size. Here, grid size is chosen to be 3×3 . After identifying the SSDM measure, QSD is calculated. QSD defined in previous work identifies the distance between the query and the SSDM elements. The expression for the QSD

measure between the query and SSDM measure is indicated as follows,

$$QSD = \sqrt{\sum_{u=1}^k \sum_{v=1}^2 (M_{uv} - SSDM_{uv}^o)^2}, \quad (17)$$

where $SSDM_{uv}^o$ indicates the SSDM measure between the elements in the SSDM vector. After the QSD and SSDM vector calculation, both the vectors are multiplied to obtain query track distance mentioned,

$$H = SSDM \times QSD. \quad (18)$$

From the SSDM measure H, the minimum values are identified, and their corresponding path vector is identified. The retained path has the same length as the query. The retained path having the path length same as the query is

expressed as T_{uv}^o and it is provided to the classifier for video retrieval.

3.3 Video retrieval related to user query: Existing ENN classifier

This work uses the ENN classifier [27] for retrieving video sections related to the user query. The ENN classification scheme is the extension of the K-Nearest neighbor algorithm, which helps in classifying the data into their respective classes. The ENN scheme utilizes maximum gain related to intraclass coherence for identifying the suitable class information. Mathematical expression behind ENN for the classification is given as follows,

Consider the input database and the class label as E . The corresponding class label for each data can be obtained as,

$$E = \text{Arg} \left\{ \max_{l=1}^D (E_l) \right\}. \quad (19)$$

The class label l has the maximum value of the output sample, and it is obtained through the neighbor vector. ENN performs the data classification by matching data sample based on a Euclidean distance measure. Here, suitable v samples providing the minimum distance is matched with other data samples, and again, a new set of v samples is selected. Finally, the class belonging to the maximum number of samples is identified, and it is mentioned as,

$$E_l = \frac{1}{X_{lv}} \sum_{p \in m_l} \sum_{gg=1}^v S_{rr}(p,m), \quad (20)$$

where, X_l indicates the data samples in l th class, and rr refers to the neighbors. The class information $S_{rr}(p,m)$ for each neighbor is expressed as follows,

$$S_{rr}(p,m) = \begin{cases} 1; & \text{if } p \notin m_i \& X_{rr}(M, R \in m_i) \\ 0; & \text{otherwise} \end{cases} \quad (21)$$

The above equation expresses the output class information while the query arrives at the video retrieval system.

4 Results and discussion

The simulation results of video retrieval strategy with proposed density measure based path equalization scheme are presented here. The experimentation of the proposed scheme is done by taking the videos from standard CAVIAR database and compared with the metrics, such as MOTP, precision, recall, and F-measure.

4.1 Experimental setup

Experimentation of the proposed video retrieval is implemented in MATLAB tool. The PC used for experimentation requires the Windows 10 OS, 4 GB RAM, and Intel I3 processor. The proposed video retrieval scheme uses various video clips for the analyzing the performance.

4.1.1 Database description

For the experimentation of the proposed video retrieval scheme, required videos are utilized from standard CAVIAR database [28]. CAVIAR database contains the collection of video clips recorded under various scenarios under walk, browse, slump, left the object, fight, window shop, etc. For the comparative analysis, this work utilized five video clips from the CAVIAR database. The directions of movement of objects have been marked in ground information.

4.1.2 Evaluation metrics

Various evaluation metrics used for analyzing the performance of the proposed SSDM + ENN technique in video retrieval strategy are explained below:

Precision: It refers to the fraction of most relevant instances among the retrieved instances. A good precision mechanism depends on the most retrieved relevant track of the object in the video, and the expression for precision is given as follows,

$$\text{Precision} = \frac{\text{Relevant instances} \cap \text{Retrieved instances}}{\text{Retrieved instances}}. \quad (22)$$

Recall: It refers to the fraction of finding the full set of relevant videos in a huge result set by effective mining techniques, and the expression is given as follows,

$$\text{Recall} = \frac{\text{Relevant instances} \cap \text{Retrieved instances}}{\text{Relevant instances}}. \quad (23)$$

F-measure: It is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test.

$$F\text{-measure} = \frac{2 * \text{Precision} * \text{recall}}{\text{Precision} + \text{recall}}. \quad (24)$$

MOTP: The ability of the tracker to track the position of the objects accurately in the video refers to MOTP, and it is expressed as follows,

$$\text{MOTP} = \frac{\sum_{i,t} d_t^k}{\sum_t M_t}, \quad (25)$$

where d_t^k refers to a number of matched objects identified at the time t and the term M_t indicates the number of matches at a time t .

4.1.3 Comparative techniques

The comparative technique includes the NSA [23], Exponential Weighted Moving Average (EWMA) [29], NSA + EWMA, NSA + NARX. The results of the proposed SSDM + ENN are compared with the other existing techniques to highlight the dominance of the techniques. Description of the various comparative techniques is presented below:

NSA: Here, the NSA scheme was utilized for tracking the position of the objects in the video frames.

EWMA: EWMA scheme finds the weighted function and exponential function for identifying the position of the target object and tracks the path of the object in video frames.

NSA + EWMA: This work makes use of both NSA and EWMA schemes for video retrieval. NSA and EWMA algorithms are hybridized for tracking the exact position of the object in the video frame.

NSA + NARX: Similar to NSA + EWMA model, the existing NSA + NARX model uses both the NARX and NSA for object tracking.

4.2 Experimental results

This section deals with the performance analysis of five videos and their experimental results. The results are taken from five videos obtained from the CAVIAR dataset, and their performances are evaluated. Objects are tracked and the paths detected are stored in the database.

The retrieved results on a user query are provided in the form of trajectory images with the similar trajectory as shown in Fig. 3. It shows the tracked path according to the specified user query. Figure 3 depicts the tracked path along with the user query obtained for the videos 1–5 respectively. As shown in Fig. 3, the actual trajectory path in the video frame, user query, and the retrieved trajectory path are presented. Figure 3a, b, c provide the video samples, query and retrieved trajectory for video sample 1. Trajectory retrieved from video sample 2 and 3, is depicted in Fig. 3f, i. Similarly, the retrieved trajectory path based on a query for video sample 4 and 5 is given in Fig. 3l, o, respectively.

4.3 Comparative analysis

The comparative analysis of the proposed SSDM + ENN concerning the existing techniques is analyzed here. The results shown in the proposed model are compared with the

existing models regarding Precision, Recall, F-measure, and MOTP for varying video samples.

4.3.1 Comparative analysis using video 1

This subsection covers the comparative analysis of the proposed method for Video 1. Figure 4a, b show the comparative analysis with the following parameters, such as MOTP, precision, recall and the F-measure. While increasing the number of objects, the value of MOTP decreases. When the number of objects is 8, the MOTP obtained by the proposed SSDM + ENN method is at a rate of 0.764, whereas the MOTP obtained using the other existing methods, like NSA, EWMA, NSA + EWMA, and NSA + NARX is at a rate of 0.726, 0.700, 0.725 and 0.755, respectively. As compared to other existing techniques, the value of MOTP is greater for the proposed SSDM + ENN technique. The MOTP gradually decreases while moving the number of objects from 2 to 8.

Figure 4b shows the comparative analysis implemented regarding Precision, Recall, and F-measure. The precision obtained for the proposed SSDM + ENN method is higher as compared to the existing methods. The proposed SSDM + ENN method attains a precision value at a rate of 0.822, whereas the existing methods, such as NSA, EWMA, NSA + EWMA, and NSA + NARX attain a precision of 0.721, 0.717, 0.746, and 0.795, respectively. The proposed SSDM + ENN method attains a recall value at a rate of 0.847, whereas the existing NSA, EWMA, NSA + EWMA, and NSA + NARX attain a value of 0.759, 0.734, 0.776, and 0.828. The proposed SSDM + ENN method shows the highest recall value as compared to the existing methods, NSA, EWMA, NSA + EWMA, and NSA + NARX. The proposed SSDM + ENN method attains an F-measure value at a rate of 0.806, whereas the existing NSA, EWMA, NSA + EWMA, and NSA + NARX attain an F-measure value of 0.7451, 0.7542, 0.7680, and 0.7948.

4.3.2 Comparative analysis using video 2

This subsection deals with the comparative analysis of the proposed method for Video 2. From Fig. 5a, it is observed that as MOTP increases the number of the object to increases. The MOTP acquired by the proposed model is greater than the existing methods. The MOTP values acquired by NSA, EWMA, NSA + EWMA, and NSA + NARX are 0.73, 0.71, 0.74, and 0.77 when a number of the object is two. The comparative analysis of techniques in terms of precision, Recall and F-measure is shown in Fig. 5b. The proposed SSDM + ENN method attains a precision value of 0.859, whereas the existing NSA, EWMA, NSA + EWMA, and NSA + NARX attain a precision of 0.759, 0.768, 0.793, and 0.815, respectively. The proposed SSDM + ENN method attains a recall value


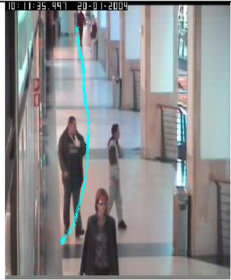
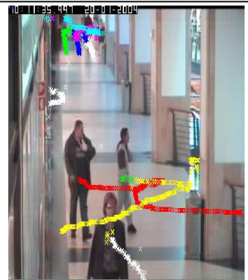
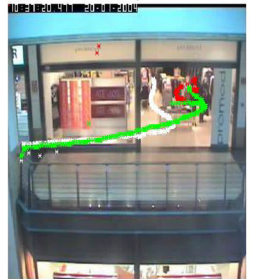


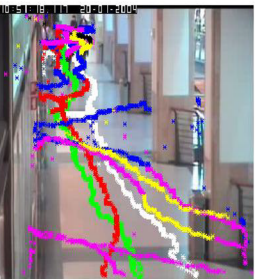

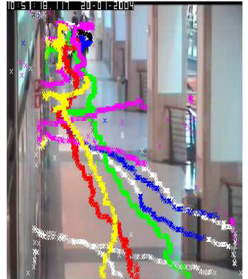



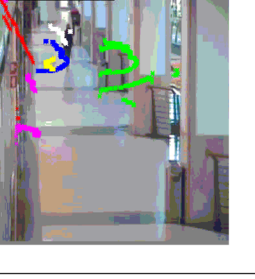
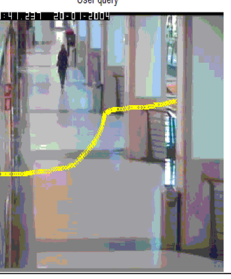

	The trajectory of the objects	User Query	Retrieved trajectory
Video 1			
	a	b	c
Video 2			
	d	e	f
Video 3			
	g	h	i
Video 4			
	j	k	l
Video 5			
	m	n	o

Fig. 3 The trajectory of the objects retrieved from the videos 1–5

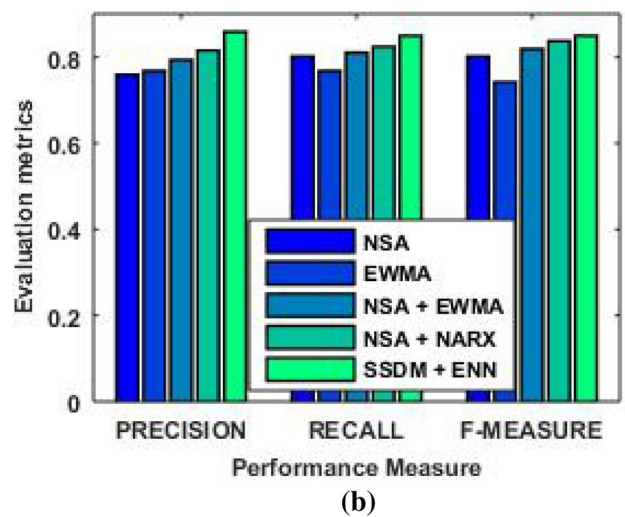
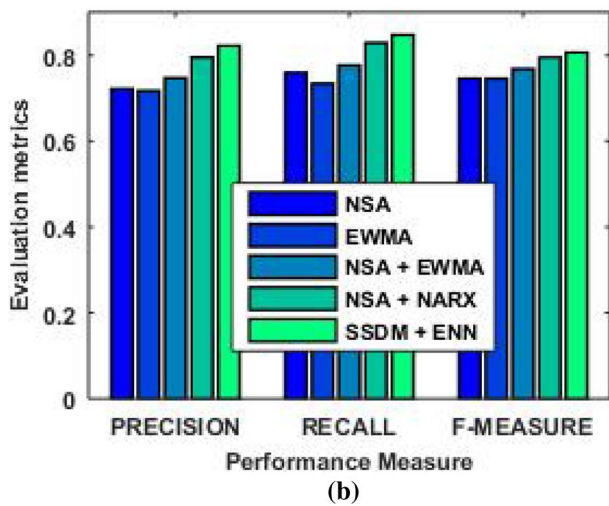
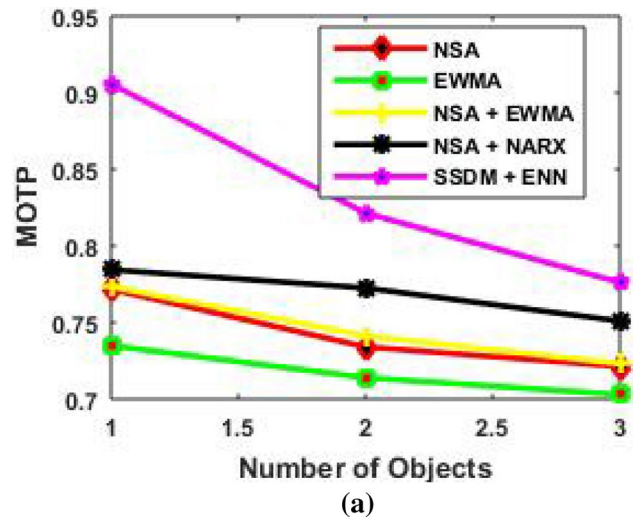
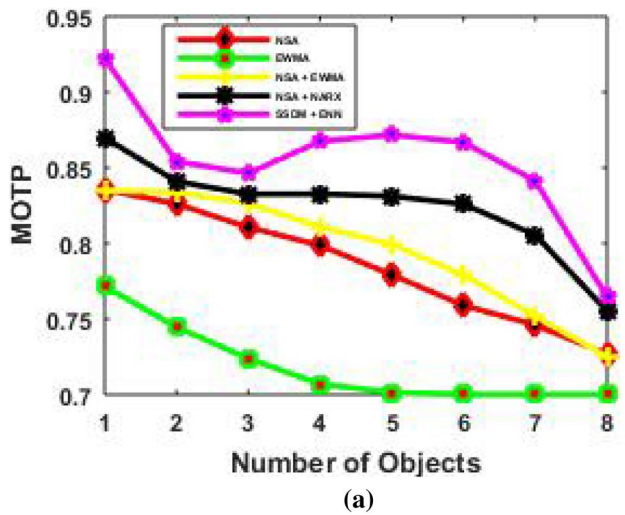


Fig. 4 Comparative analysis using the video one based on **a** MOTP. **b** Evaluation metrics

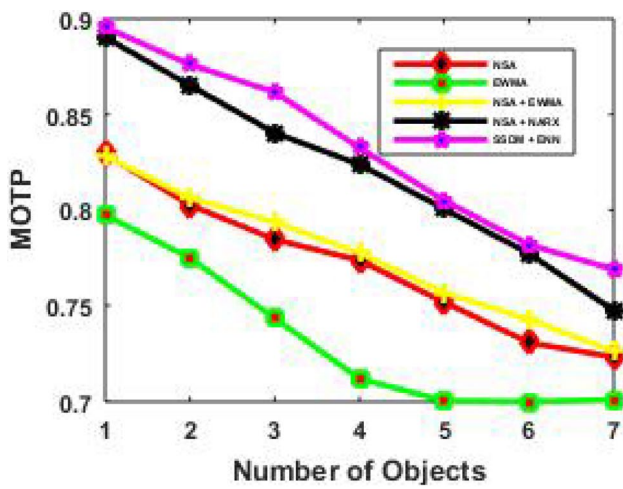
Fig. 5 Comparative analysis using the video two based on **a** MOTP. **b** Evaluation metrics

at a rate of 0.849, whereas the existing methods, such as NSA, EWMA, NSA + EWMA, and NSA + NARX attain a value of 0.802, 0.767, 0.810, and 0.824. The proposed SSDM + ENN method shows the highest recall value as compared to existing NSA, EWMA, NSA + EWMA, and NSA + NARX. The proposed SSDM + ENN method attains an F-measure value of 0.849, whereas the existing NSA, EWMA, NSA + EWMA, and NSA + NARX attain a value of 0.802, 0.742, 0.818, and 0.837, respectively.

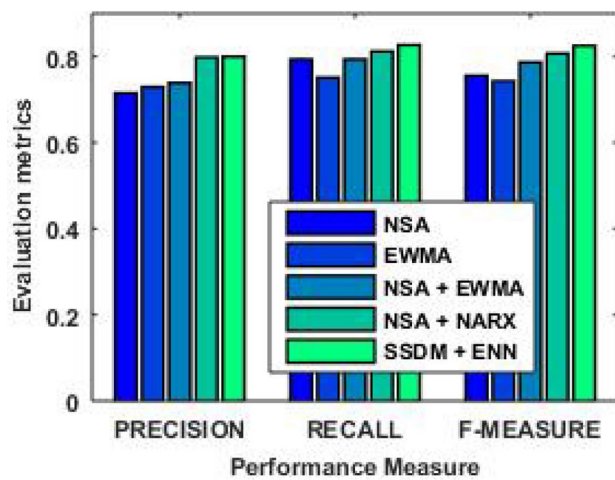
4.3.3 Comparative analysis using video 3

In this subsection, the comparative analysis of the proposed method for Video 3 has been described. From Fig. 6a, it is depicted that MOTP increases with the increasing number of the objects. The MOTP acquired by the proposed SSDM + ENN model is greater than the

existing methods. The MOTP values acquired by NSA, EWMA, NSA + EWMA, and NSA + NARX are 0.751, 0.700, 0.756, and 0.800 when the number of the object is five. The combined comparative analysis of the proposed SSDM + ENN and other existing techniques in terms of precision, Recall and F-measure is shown in Fig. 6b. The proposed SSDM + ENN method attains a precision value of 0.799, whereas the existing methods, such as NSA, EWMA, NSA + EWMA, and NSA + NARX attain a value of 0.714, 0.729, 0.738, and 0.798, respectively. The proposed SSDM + ENN method attains a recall value of 0.826, whereas the existing NSA, EWMA, NSA + EWMA, and NSA + NARX attain a recall value of 0.793, 0.750, 0.793, and 0.811. The proposed SSDM + ENN method shows the highest recall value as compared to the existing methods, like NSA, EWMA, NSA + EWMA, and NSA + NARX. The proposed SSDM + ENN method attains an F-measure



(a)



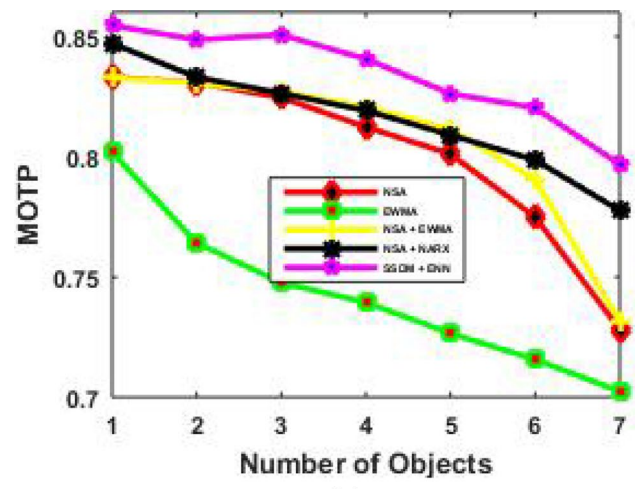
(b)

Fig. 6 Comparative analysis using the video three based on **a** MOTP. **b** Evaluation metrics

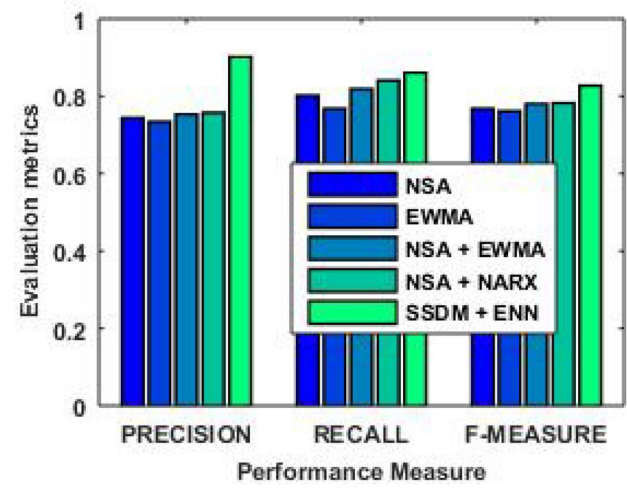
value of 0.824, whereas NSA, EWMA, NSA + EWMA, and NSA + NARX had a value of 0.755, 0.743, 0.786, and 0.807.

4.3.4 Comparative analysis using video 4

This subsection deals with the comparative analysis of the proposed method for Video 4. From Fig. 7a, it is observed that when the number of the object increases, MOTP also increases. The MOTP acquired by the proposed SSDM + ENN model is greater than that of existing methods. The MOTP values acquired by NSA, EWMA, NSA + EWMA, and NSA + NARX are 0.77, 0.71, 0.790, and 0.798 when the number of objects is six. The comparative analysis of the proposed SSDM + ENN regarding precision, Recall and F-measure is shown in Fig. 7b. The proposed SSDM + ENN method attains a precision of 0.901, whereas the existing methods, such as NSA,



(a)



(b)

Fig. 7 Comparative analysis using the video four based on **a** MOTP. **b** Evaluation metrics

EWMA, NSA + EWMA, and NSA + NARX had attained a value of 0.743, 0.734, 0.753, and 0.757, respectively. The proposed SSDM + ENN method attains a recall value of 0.860, whereas the existing methods, like NSA, EWMA, NSA + EWMA, and NSA + NARX had a value of 0.802, 0.767, 0.819, and 0.840. The proposed SSDM + ENN method shows the highest recall value as compared to the existing NSA, EWMA, NSA + EWMA, and NSA + NARX. The proposed SSDM + ENN method attains F-measure of 0.827, whereas the existing methods, such as NSA, EWMA, NSA + EWMA, and NSA + NARX attain an F-measure of 0.768, 0.761, 0.780, and 0.782.

4.3.5 Comparative analysis using video 5

In this subsection, the comparative analysis of the proposed method for Video 5 is shown. Figure 8a presents the performance of comparative models for video five based on

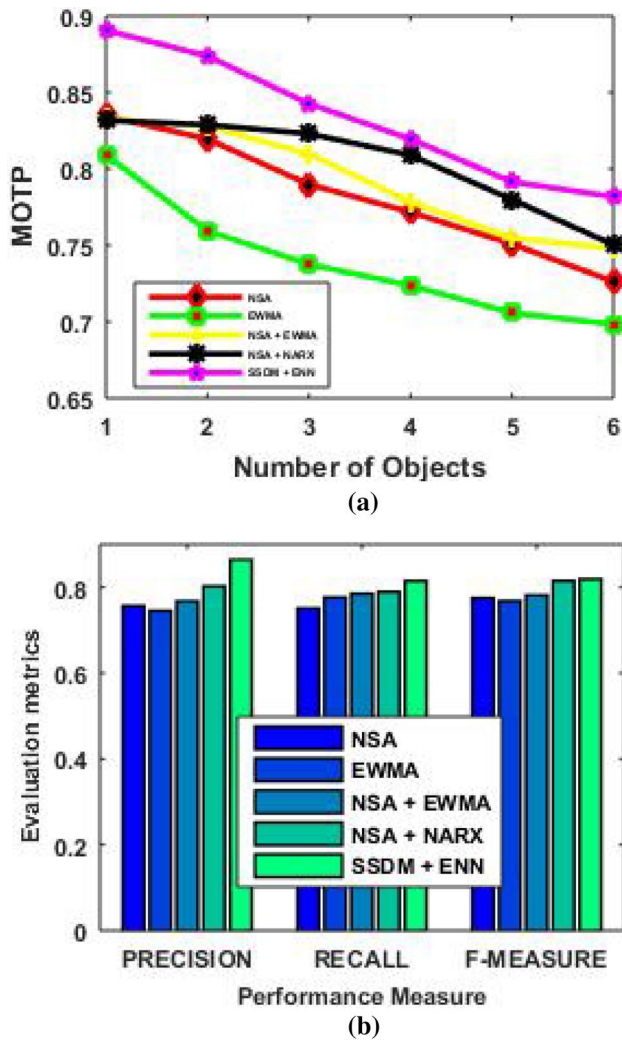


Fig. 8 Comparative analysis using the video five **a** MOTP. **b** Evaluation metrics

MOTP measure. As the number of objects increases, MOTP increases. Hence, the MOTP acquired by the proposed model is greater than the existing methods. The MOTP values acquired by the NSA, EWMA, NSA + EWMA, and NSA + NARX are 0.771, 0.723, 0.778, and 0.808 when the number of the object is four. The combined performance

analysis of techniques in terms of precision, Recall and F-measure is shown in Fig. 8b. The proposed SSDM + ENN method attains a precision value of 0.864, whereas the existing methods, NSA, EWMA, NSA + EWMA, NSA + NARX attain of 0.756, 0.745, 0.767, and 0.802, respectively. The proposed SSDM + ENN method attains a recall of 0.814, whereas the existing NSA, EWMA, NSA + EWMA, and NSA + NARX attain a value of 0.750, 0.776, 0.785, and 0.789. The proposed SSDM + ENN method shows the highest recall value as compared to the existing NSA, EWMA, NSA + EWMA, and NSA + NARX. The proposed SSDM + ENN method attains F-measure value at a rate of 0.819, whereas the existing methods, like NSA, EWMA, NSA + EWMA, and NSA + NARX had a value of 0.774, 0.768, 0.781, and 0.815.

4.4 Comparative discussion

Table 1 highlights the performance metrics of the proposed method to show its precedence among the existing methods. The MOTP rates attained by NSA + NARX and proposed SSDM + ENN are 0.869 and 0.922. Thus, the proposed SSDM + ENN has higher tracking value than the existing methods. The precision rate attained by NSA, EWMA, and Colour feature-based model is 0.7413, whereas that of the proposed SSDM + ENN is 0.901. Hence, the precision for tracking the object is higher in the proposed SSDM + ENN. The Recall rates attained by NSA, EWMA, Colour feature-based model, Trajectory clustering-based model, NSA + EWMA, NSA + NARX, and proposed SSDM + ENN are 0.7732, 0.7748, 0.774, 0.7927, 0.8106, 0.840, and 0.860. Hence, the fraction of the successfully retrieving trajectory relevant to the query trajectory is higher in the proposed method. The F-measure rates attained by NSA, EWMA, Colour feature-based model, Trajectory clustering-based model, NSA + EWMA, NSA + NARX, and proposed SSDM + ENN are 0.7434, 0.7581, 0.7507, 0.7625, 0.7670, 0.837, and 0.849, respectively. From the table, it is evident that proposed SSDM + ENN technique achieved better video retrieval performance with the values of 0.901, 0.860, 0.849, and 0.922 for precision, recall, F-measure, and MOTP, respectively.

Table 1 Comparative discussion

Methods	MOTP	Precision	Recall	F-measure
NSA	–	0.7413	0.7732	0.7434
EWMA	–	0.7413	0.7748	0.7581
Colour feature-based model [27]	–	0.7413	0.7740	0.7507
Trajectory clustering-based model [7]	–	0.7465	0.7927	0.7625
NSA + EWMA [22]	–	0.7517	0.8106	0.7670
NSA + NARX	0.869	0.815	0.840	0.837
SSDM + ENN	0.922	0.901	0.860	0.849

5 Conclusion

Video retrieval strategy has gained more interest in recent years as the technique can be applied in various applications. The proposed scheme develops an extensive approach to the previously developed video retrieval strategy. Here, video retrieval is done in three phases, (1) object detection, (2) path length equalization, and (3) video retrieval. Initially, the trajectory path of the target object gets detected with the help of the hybrid model, having the NSA and SVR scheme. Path length equalization retrieves the path having the same length as the query from a user and the proposed SSDM path length equalization scheme finds the suitable paths related to query. After that, video retrieval is done using the ENN classifier. Simulation of the proposed SSDM + ENN video retrieval scheme is done by obtaining videos from CAVIAR database, and the implementation is done in MATLAB tool. The evaluation metrics, such as MOTP, precision, recall, and F-measure evaluates the performance of the proposed SSDM + ENN technique with other comparative techniques. From the simulation results, it is evident that the proposed SSDM + ENN technique achieved better video retrieval performance with the values of 0.901, 0.860, 0.849, and 0.922 for precision, recall, F-measure, and MOTP, respectively.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Joy E, Peter JD (2018) Visual tracking with conditionally adaptive multiple template update scheme for intricate videos. *Multimedia Syst* 24(2):175–194
- Wang J, Lu H, Guo H (2016) Multiple deep features learning for object retrieval in surveillance videos. *IET Comput Vision* 10(4):268–272
- Zhang N, Jeong HY (2017) A retrieval algorithm for specific face images in airport surveillance multimedia videos on cloud computing platform. *Multimedia Tools Appl* 76(16):17129–17143
- Yang X, Zhou Y, Zhou D, Hu Y (2017) Image segmentation and proto-objects detection based visual tracking. *Optik Int J Light Electron Opt* 131:1085–1094
- Morimitsu H, Bloch I, Cesar-Jr RM (2017) Exploring structure for long-term tracking of multiple objects in sports videos. *Comput Vis Image Underst* 159:89–104
- Zhu Z, Ren X, Chen Z (2017) Integrated detection and tracking of workforce and equipment from construction jobsite videos. *Autom Construct* 81:161–171
- Lai YH, Yang CK (2015) Video object retrieval by trajectory and appearance. *IEEE Trans Circuits Syst Video Technol* 25(6):1026–1037
- Kanagamalliga S, Vasuki S (2018) Contour-based object tracking in video scenes through optical flow and gabor features. *Optik Int J Light and Electron Opt* 157:787–797
- Mahalingam T, Subramoniam M (2018) A robust single and multiple moving object detection, tracking and classification. *Appl Comput Inform*. <https://doi.org/10.1016/j.aci.2018.01.001>
- Jin R, Kim J (2017) Tracking feature extraction techniques with improved SIFT for video identification. *Multimedia Tools Appl* 76(4):5927–5936
- Elgammal A, Harwood D, Davis L (2000) Non-parametric model for background subtraction. In: *Proceedings of the European conference on computer vision*, Springer, Berlin, Heidelberg, pp. 751–767
- Ren T, Qiu Z, Liu Y, Yu T, Bei J (2015) Soft-assigned bag of features for object tracking. *Multimedia Syst* 21(2):189–205
- Ross DA, Lim J, Lin RS, Yang MH (2008) Incremental learning for robust visual tracking. *Int J Comput Vis* 77(1–3):125–141
- Sugandi B, Kim H, Tan JK, Ishikawa S (2010) A color-based particle filter for multiple object tracking in an outdoor environment. *Artif Life Robot* 15(1):41–47
- Liang-qun L, Xi-yang Z, Zong-xiang L, Wei-xin X, (2018) Fuzzy logic approach to visual multi-object tracking. *Neurocomputing* 281:139–15115
- Bency AJ, Karthikeyan S, De Leo C, Sunderrajan S, Manjunath BS (2017) Search tracker: human-derived object tracking in the wild through large-scale search and retrieval. *IEEE Trans Circuits Syst Video Technol* 27(8):1803–1814
- Ma Y (2017) An object tracking algorithm based on optical flow and temporal–spatial context. *Cluster Comput* 1–9
- Khare M, Srivastava RK, Khare A (2017) Object tracking using combination of daubechies complex wavelet transform and Zernike moment. *Multimedia Tools Appl* 76(1):1247–1290
- Leang I, Herbin S, Girard B, Droulez J (2018) On-line fusion of trackers for single-object tracking. *Pattern Recogn* 74:459–473
- Yang H, Qu S, Zhu F, Zheng Z (2018) Robust objectness tracking with weighted multiple instance learning algorithm. *Neurocomputing* 288:43–532
- Ratre A, Pankajakshan V (2017) Tucker visual search-based hybrid tracking model and fractional Kohonen self-organizing map for anomaly localization and detection in surveillance videos. *Imag Sci J* 66(4):195–210
- Ghuge CA, Ruikar DS, Prakash VC (2016) Query-specific distance and hybrid tracking model for video object retrieval. *J Intell Syst* 27(2):195–212
- Turaga P, Chellappa R (2010) Nearest-neighbor search algorithms on non-Euclidean manifolds for computer vision applications. In: *Proceedings of the seventh Indian conference on computer vision, graphics and image processing*, pp. 282–289,
- Ni KS, Nguyen TQ (2007) Image superresolution using support vector regression. *IEEE Trans Image Process* 16(6):1596–1610
- Tang B, He H (2015) ENN: Extended nearest neighbor method for pattern recognition [research frontier]. *IEEE Comput Intell Mag* 10(3):52–60
- Ding S, Li G, Li Y, Li X, Zhai Q, Champion AC, Zhu J, Xuan D, Zheng YF (2017) SurvSurf: human retrieval on large surveillance video data. *Multimedia Tools Appl* 76(5):6521–6549
- Cai Z, Liang Y, Hu H, Luo W (2016) Offline video object retrieval method based on color features. In: Li K, Li J, Liu Y, Castiglione A (eds) *Computational intelligence and intelligent systems*. ISICA 2015. Communications in computer and information science, vol 575. Springer, Singapore, pp 495–505
- CAVIAR database (2018) <https://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/>. Accessed Feb 2018
- Kui Liu B, Liu E, Blasch D, Shen Z, Wang H, Ling G, Chen (2015) A cloud infrastructure for target detection and tracking using audio and video fusion. In: *Proceedings of IEEE conference on computer vision and pattern recognition workshops (CVPRW)*, pp 74–81

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.