Received: November 23, 2020. Revised: March 2, 2021.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

An Integrated Approach Using Optimized Naive Bayes Classifier and Optical Flow Orientation for Video Object Retrieval

Chandrashekhar Ghuge[1]* **Vudatha Chandra Prakash**[1] **Sachin Ruikar**[2]

[1]*Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur district, AP, India*
[2]*Department of Electronics Engineering, Walchand College of Engineering, Sangli, Maharashtra, India*
* Corresponding author's Email: caghugeklu@gmail.com

Abstract: The popularity of video recordings either on mobile devices or video surveillance has contributed to the demand for video data applications. As a result, the management of video has become significant in the object retrieval process. However, the video object retrieval is considered as a major issue in the video objects management. This paper proposes an integrated optimization-based classifier for video object retrieval. Here, the input video is subjected to optical flow estimation using a spatio-temporal feature descriptor, named Histograms of Optical Flow Orientation and Magnitude (HOFM) for extracting the events and establishing the HOFM indexed database using the detected objects. Hence, object tracking is a major step, which detects and tracks the video objects using a hybrid model named Nearest Search Algorithm-Support Vector Regression (NSA-SVR). For the video retrieval, the training of Naive Bayes (NB) classifier is performed using the proposed Lion-Salp Swarm Algorithm (LSSA) on the indexed database of the tracked objects. Then, the recognized events interrelated to the query are subjected to the Naive Bayes classifier to retrieve the required video. The performance of the proposed integrated LSSA-Naive Bayes Classifier is found to be better than the existing methods with maximal precision of 82.985 percent, recall of 87.451 percent, and F-measure of 87.847 percent.

Keywords: Video surveillance, HOFM, Naive bayes classifier, Video retrieval, Optical flow.

1. Introduction

Innovations in Web and multimedia technology are growing tremendously, attracting the researchers to multimedia applications such as video archives, entertainment, and news video broadcasting. It is therefore essential to index video for successful retrieval from large video datasets. However, the retrieval of videos from a huge database needs knowledge about videos. These textual queries are effective, if the users pose precise knowledge for structuring databases in order to manually search the videos using the index. However, the conventional approaches cannot work when the user searches for the required videos from a web database or to locate a video without having any knowledge about those videos [1]. The systems based on videos can capture huge amounts of required information and are relatively cost-effective due to simpler camera installation and maintenance. A huge number of video cameras are installed all over the place for security. Due to this, there is an imperative need for developing techniques using videos, which can substitute the human operators for monitoring the area under observation. A robust tracking system recognizes and tracks the moving objects in intelligent video systems [2]. Some of the studies of vision-based object recognition are mainly focused on detecting objects [3]. The multi-class image segmentation and object detection are the two tasks, which are used to solve the issues of object detection. [4]. several techniques are devised using the learned local representations and can be considered as the basis for texture detectors. A conditional random field model is devised due to its local nature, for enforcing spatial consistency [5]. Weakly supervised object detection is considered as a major challenge, due to complex training procedure, the location of objects and the dynamic model appearance in each image. The conventional approach for solving the tracking issues is to consider the object location, which is of great interest for the static images, and the approaches provided a way to minimize the loss produced by such dependent variables while learning [6].

Moreover, the appearance of learning and its localization are the two interrelated tasks; the appearance is not symmetric, and the method is trapped in a local minimum; that is, the algorithm misses the objects for some images [7]. However, there have been significant developments over many years, specifically in the detection of objects using static images. Due to the success of these systems, the current work on the recognition of actions has extended several triumphant facts into the Spatio-temporal realms [8].

The automated video surveillance system needs proficient video analysis and video retrieval and indexing techniques, considering event and object levels for filling the semantic gap that persists between the high-level semantic content and low-level features of videos. The enhanced video indexing and retrieval techniques offer adequate facilities to access the contents of video surveillance [9]. The progress in Machine Learning has led to the usage of Deep Learning techniques in many domains [10]. In object detection-oriented clustering method is used for indexing video semantics. On the other hand, the object retrieval system resolved the problems of several state-of-the-arts object detectors such as You Live Only Once (YOLO) [11]. And Faster region-based convolutional neural network (FRCNN) [12, 13]. In [14] query examples with FRCNN are utilized in the region as a feature detector. The cosine similarities between the deep features query and object proposals are utilized for ranking [15]. The deep learning techniques are efficient in determining the objects, a broad range of techniques and classifiers were devised, which includes Bayesian localization, possibility theory [16], neural networks [17] Bayesian inference [18] and fuzzy logic [19], which is utilized for object localization.

The proposed method is devised in three stages, namely object tracking, event extraction, and retrieval. Here, the video object location is determined by employing a hybrid model named NSA-SVR. The spatial tracking is done using SVR, whereas visual tracking is done by the NSA. The events are extracted by applying the HOFM on each tracked object. The events are further employed for training the NB classifier for determining the video objects. In the video object retrieval, the query is processed by the classifier for retrieving the highly relevant object trajectories and the events associated with it. The efficiency of the object retrieval is based on the proposed LSSA algorithm, which is developed using the standard Lion Optimization Algorithm and Salp Swarm Algorithm.

The paper is structured as follows: Section 1 provides the introductory part related to video retrieval. Section 2 elaborates on different techniques based on video retrieval. Section 3 elaborates the proposed technique for training the NB classifier. Section 4 demonstrates the results of the methods for the video object retrieval, and section 5 provides the conclusion.

2. Related work

This section discusses significant video object retrieval methods, their benefit and drawbacks and addresses the papers research gaps.

Lin [20] developed a learning framework to retrieve the desired frame from the sequence of input videos. A query-adaptive Multiple Instance Learning (q-MIL) algorithm was devised by exploiting the information of visual appearance using the query based on the video frames. The derived learning model adapted additional discriminating abilities while retrieving the relevant instances. Nguyen [21] developed a fusion method, object detectors were utilized based on a denser feature to determine the similarity score and object bounding box. Then, the spatial relationship of the visual object and the object proposal was utilized for determining the query. The method was flexible, which helped to enable any object detectors without altering the system structure. Durand [22], developed retrieval based on video segmentation, which helped to shorten the time taken for retrieval. The long videos were detected using a Spatio-Temporal Interest Points (STIPs) detection algorithm. Then, the super frame segmentation of the refined, long video was performed for gaining an interesting clip for huge videos. For selecting the keyframes, the Region Of Interest (ROI) was created based on STIP, and the saliency detection of the ROI was used for screening out the video keyframes. Rashad [23] developed a method for overcoming the issues of the indexing and retrieval techniques by adapting deep learning methodologies. The method utilized stage object detector stage like YOLOv2, which can be utilized as an effective tool for detecting the events. The idea behind the method was that the pixel values were determined by a deep detector to categorize the objects in the image by removing the background information. The method offered a solution for solving the issues of video quality using the light convolutional network for efficient object retrieval. However, the method failed to use an ensemble strategy to improve the object recognition performance. Lou [24] designed an object retrieval approach using cameras and Inertial Measurement Unit (IMU) sensors to retrieve 3D objects. The method utilized fast and compacted image descriptors and the absolute orientation for building multi-view centred retrieval object models. Furthermore, a Hough transformation paradigm was utilized for evaluating the query results using several video frames. Lin [25] developed deep-learning features and utilized in Compact Descriptor for Video Analysis (CDVA) evaluation framework for studying the effectiveness of video analysis. Garcia [26] developed a feature fusion method using multimodal graph learning for retrieving the 3D objects. Different graphs were devised on the basis of different features for learning optimized weights for fusing the features. However, the method failed to use different types of features and distance functions for generating geo-location-based applications. The main drawback associated with detecting the objects suffer from lack of the history of the objects, occlusions, small size of objects, confusion regarding the posture, movement in objects.

To overcome the aforementioned drawbacks the effective video object retrieval method is developed using the proposed LSSA-NB classifier.

3. Proposed LSSA-NB classifier for video object retrieval

The proposed method uses the training phase for learning the indexed database and testing phase for retrieving the relevant objects based on a user query. In the training phase, the keyframes in the video are chosen considering the frames, for which initially the object detection of the individual frames is determined using a hybrid model named NSA-SVR. The detected object undergoes optical flow extraction using a spatio-temporal feature descriptor, named HOFM [27] for extracting the HOFM features of each object. The key points are obtained from the detected objects from the selected keyframes using the HOFM descriptor, and the indexed database is developed using the HOFM descriptor of the video objects in such a way that the proposed LSSA-NB classifier is trained using the indexed database. The proposed LSSA is newly developed by modifying the update process of the LA [28] with the update process of SSA [29]. In the testing phase, the query is given as an input for which the HOFM is applied for extracting the object location. After tracking the object's location, the events tracked by the tracking method and query need to be matched based on the probability function in LSSA-NB classifier. The event analysis plays a major role in video retrieval as the video contents related to the user query are retrieved. Hence, the NB classifier is employed to retrieve suitable video frames. If the equivalent event is computed for the input query, then the suitable video frames are retrieved shown in Fig. 1.

Consider a database Dcontaining gnumber of videos and is represented as,

$$D = \{V_j \ ; \ 1 \leq j \leq g\} \tag{1}$$

Where, V_j jrepresents j^{th} video, and g is the total number of videos. Every video includes $d$number of frames, which are expressed as,

$$V = f_m \ ; 1 \leq m \leq d \tag{2}$$

Where, f_m is the m^{th} frame, and drefers the total number of frames.

3.1 Hybrid NSA-SVR for object detection and tracking

This subsection illustrates the process for tracking the object movement considering different frames. In order to determine the object, utilizes the hybrid NSA-SVR model. The NSA [30] and SVR integrate the results attained by both for tracking objects.

3.2 Object tracking based on NSA

The location of the targeted objects recognizes by determining the best location in the current frame. The steps contained in recognizing the accurate object location in the video frame using the NSA are elaborated below:

1) Detection of target object

The NSA identifies the exact location of object by locating different objects present in the image. As per equation (2), the video has several frames which is expressed as,

$$f = \{f_1, f_2, \dots, f_m, \dots, f_d\} \tag{3}$$

where,$f_m$refers to the m^{th}frame, and total frames present in the video is given as d.

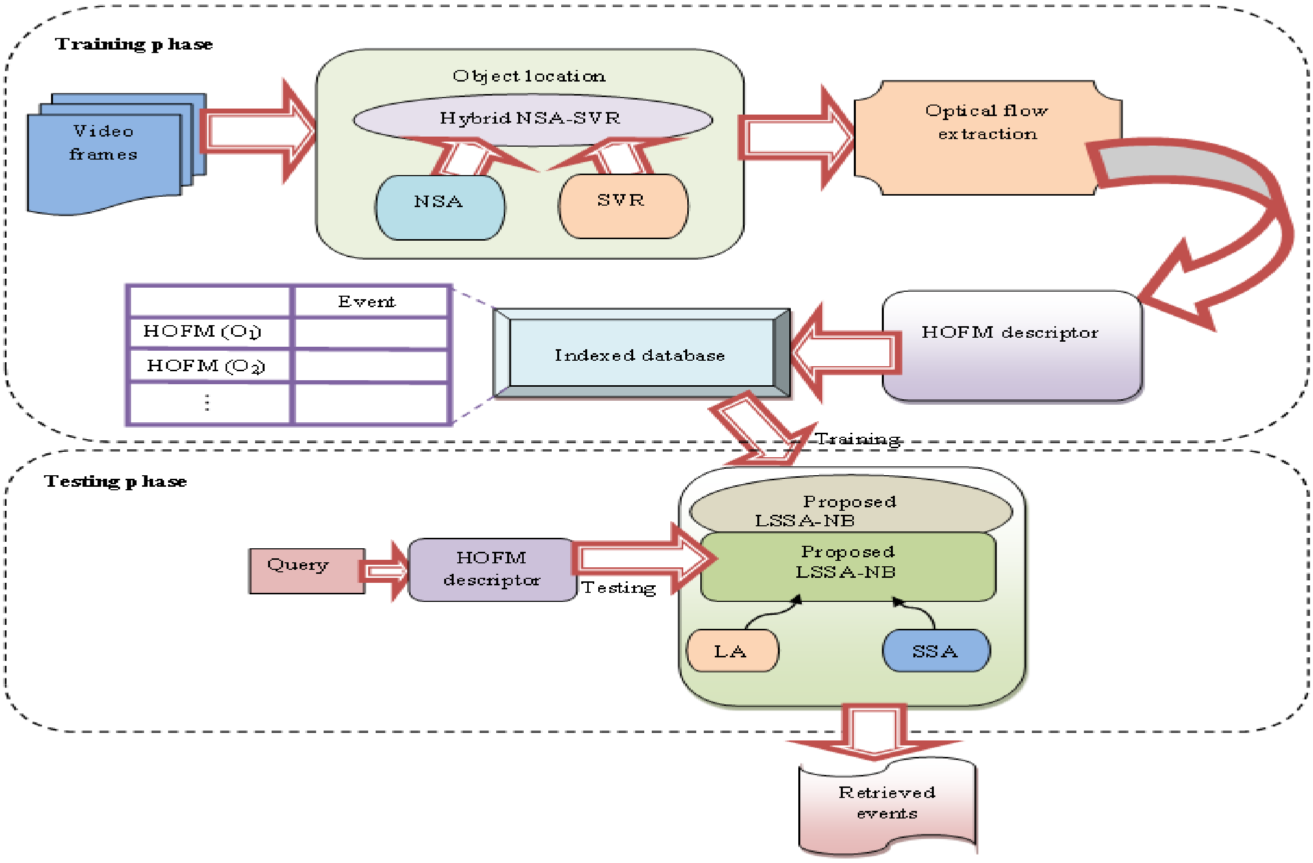


Figure. 1. Schematic view of video object retrieval framework using proposed LSSA-NB classifier

Each frame contains different objects positioned at different locations. $Assume m^{th}$ video frame has T the number of objects. The position of the z^{th} object in the $frame f_m$ is given as,

$$P_z^m = (X_z^m, Y_z^m) \qquad (4)$$

where, P_z^m indicates the position of z^{th} object in the frame f_m, and the term (X_z^m, Y_z^m) specifies the X^{th} and the Y^{th} elements specifying the position P_z^m.

2) Determination of object position based on the next frame

This step utilizes the location of the next frames by setting the rectangular window. The rectangular window is utilized for identifying the object location by setting a parameter β. The object position in the consecutive frame is determined using a threshold. The new position of the object in successive frames is given as,

$$P_{z+1}^m = (X_z^m \pm \beta,\ Y_z^m \pm \beta) \qquad (5)$$

where, β denotes the extension parameter, which is constant.

3) Computation of distance between objects

The subsequent step in the NSA is to calculate the distance between the object positions of current and the next frame is given as,

$$dist(P_z^m,\ P_{z+1}^m) = dist\sqrt{(P_z^m,\ P_{z+1}^m)} \qquad (6)$$

4) Determination of objects position using NSA

Finally, the second and third steps are repeated until the positions of all objects in the video frame are calculated. The position of objects retrieved by NSA is given as,

$$N^z = \{P_z^m, P_{z+1}^{m+1}, \dots, P_{z+T}^{m+d}\} \qquad (7)$$

Where, $P_z^m, P_{z+1}^{m+1}, \dots, P_{z+T}^{m+d}$ specifies the position of the objects obtained by NSA.

3.2.1. Object tracking based on SVR

SVR [31] is utilized to track the object location in the video frame in order to enhance the tracking process quality. The SVR uses a mapping function that maps a linear estimate with the non-linear function for determine an unknown regression. Thus, to determine the exact location of the object in the video frame, the SVR produces the training set with input-output pair represented as,

$$G = \{(k_1, l_1), (k_2, l_2), \ldots, (k_d, l_d)\} \tag{8}$$

where, G specifies the input and output space and (k_1, l_1) represents the input-output pair. Where d is the total input and output pair. For producing the appropriate input-output pair, SVR utilizes the optimization issues. The optimization issue can be modelled as a dual maximization problem. The regression estimate for the dual maximization problem is represented as,

$$H(e) = \Sigma(\gamma_m^+ + \gamma_m^-) M(k.k_m) + f \tag{9}$$

where, f refers to a deviation parameter. The final object location evaluated using SVR is formulated as,

$$S^z = \left\{P_z^m, P_{z+1}^{m+1}, \ldots, P_{z+T}^{m+d}\right\} \tag{10}$$

3.2.2. Object tracking based on hybrid NSA-SVR

Finally, the hybrid NSA-SVR uses the merits of both NSA and SVR to accomplish object tracking. In the hybrid model, the result obtained by average of NSA and SVR, and the final location of the object is tracked. The object position determined is given as,

$$J_z = [\{N^z\} + \{S^z\}] \forall z \tag{11}$$

where,N^z refers to the position tracked by NSA andS^z indicates the position tracked by the SVR, respectively. The tracked objects are represented as,

$$O = \{O_1, O_2, \ldots O_z, \ldots O_T\} \tag{12}$$

3.3 Optical flow extraction using HOFM descriptor

HOFM is the highly desired descriptor and is widely utilized in object detection. The HOFM feature descriptor produces interest points, which is considered as a global descriptor. This technique computes horizontal and vertical gradients, orientation and magnitude of gradients. In the training phase, the spatiotemporal feature descriptor, HOFM, is utilized for capturing the moving patterns from non-overlapping regions from the videos, whereas in the testing phase, the incoming patterns for each region are evaluated with the stored patterns. At first, the Histograms of Oriented Optical Flow (HOOF) and the sampling approach are used for evaluating the optical flow for computing location of the pixel. The HOFM features are established for all the detected objects and are stored in the indexed database such that when the object query is given, the events are retrieved.

i) HOOF

The histograms are obtained using the HOOF, which is represented as $h_{\vartheta,v} = [h_{\vartheta,1}, h_{\vartheta,2}, \cdots, h_{\vartheta,v}]$ at each time instant l and for each block ϑ in the frame. Here, each flow vector is binned based on angle and weighted based on the magnitude.

ii) Optical flow extraction

For computing optical flow, a binary mask is created using image subtraction between $f_m$andf_{m+v}frame. For a specific threshold, Uif the resulting difference is less than Uthen, the pixel is discarded or else set for moving pixels.

iii) HOFM

The HOFM utilizes the optical flow information, which is the orientation and magnitude for constructing the feature vector. For feature vector, a matrix is constructed$B_{\tau \times v}$, where τ is number of orientation ranges and v is the number of HOOF magnitude ranges. Thus, the matrix is formed based on the orientation of vector and magnitude of vector. Thus, for given pixels $p(a, b, c)$ and$p'(a, b, c)$, the vector filed belongs to magnitude $F$and orientationθ. Thus, the feature matrix $B$is considered at each time interval$l$and is given as,

$$B(i, o) = \sum_{\vec{v}} \begin{cases} 1 if (\eta = mod(F, F')) and (\rho = mod(\theta, v) \\ 0 Otherwise \end{cases} \tag{13}$$

Where, $i$represents the orientation and $o$indicates the magnitude. After applying HOFM to each detected objects, the output obtained is given as the events which is represented as,

$$E = \{E_1, E_2, \ldots E_Z, \ldots E_T\} \tag{14}$$

Where, T represents the total events and $E_Z$indicates the Z^{th}event.

The obtained events are subjected to NB classifier for obtaining the retrieved events, which is elaborated in the next section.

3.4 Retrieval of objects using the Naive Bayes classifier

The NB classifier is well-known for its speed, as it has the capability for making real-time predictions. Moreover, the NB classifier can predict the posterior probability of different classes in order to produce a higher success rate. The NB classifier is simple and needs less data for the classification. The NB classifier is described as a probabilistic classifier, which is derived on the basis Bayes theorem using specific features. Here, the proposed LSSA method is utilized for training the NB classifier. The proposed LSSA is designed by integrating SSA with LA for generating optimal probabilistic parameters. The NB classifier calculates the mean and variance of each sample and then calculates the posterior function [32]. Then, the output obtained is the sample, which has a higher probability value as the output. Here, the NB classifier is used for retrieving the events.

The equivalent class label $\kappa$for each event can be generated using the NB classifier by defining the relevant objects from the classified events is given as,

$$\kappa = \overset{Z\underset{T}{=}1}{argmax} post(E_Z|T) \tag{15}$$

Where, $post(E_Z|T)$ indicates the posterior probability of the events of theZ^{th}event of the test data based on class and Tindicates the number of events.

$$post(E_Z|T) = p(\kappa_T)\prod_{Z=1}^{T} p\left(E_Z^T\middle|\kappa_T\right) \tag{16}$$

$$p\left(E_Z^T\middle|\kappa_T\right) = \frac{1}{\sqrt{2\pi\sigma_T^2}} \times e^{\left(-\frac{\left(E^T-\mu_T\right)^2}{2\sigma_T^2}\right)} \tag{17}$$

Where, μ_T represents the mean value and σ_T indicates the value of variance, μ_T and σ_T are calculated during the training phase. $p(\kappa_T)$ refer probability of the class.

3.4.1. Optimization of naive bayes parameters

The NB classifier assumes strong independent distributions in accordance with the HOFM features. The NB classifier employs Bayes theorem to determine the probability of data fitting into a particular class with given events. The NB classifier is trained using the proposed LSSA algorithm. The solution encoding, fitness function, and the proposed LSSA algorithm are elaborated as follows.

i) Solution encoding and evaluation of fitness

The solution encoding provides a symbolic solution representation for optimizing the tuning parameters of the NB classifier using the proposed LSSA algorithm. Assume Tbe the total number of classes, which represents total events. Moreover, the number of features considered in the NB classifier includes mean and variance. The number of classes represents the number of events. Each event Tpossess the dimension of$[1 \times (\kappa \times \ell)]$, wherein κrepresents the number of classes based on the number of features given asℓ.

The fitness function used to select the best objects is based on accuracy, and the fitness function is considered to be the maximization function, and an equation for fitness computing is shown as,

$$Fitness = \frac{T_p+T_n}{T_p+T_n+F_p+F_n} \tag{18}$$

Where, $T_p$is true positive, $F_p$denote false positive, $T_n$represents true negative, and $F_n$is the false positive.

iii) Proposed integrated LSSA algorithm

This section elaborates on the proposed LSSA for training the NB classifier in order to determine the video objects. The SSA is inspired from the salps swarming behaviour while foraging and navigating in the oceans. This algorithm enhances the random solution in an effective manner and produces optimal solutions with high coverage. Addresses the problems of real-world having complex unknown search spaces and it are effective in determining the accurate solution by solving the issues of multi-objective with unknown search spaces. The LA is inspired from the social behaviours of lion's. Moreover, the LA is considered as the best solution for solving the optimization issues. Hence, the integration of the LA and SSA obtained global optimal solution with effective calculation and approximation.

1. Initialization

The first step is the initialization of the solutions using the SSA and is given as,

$$Q = \{Q_1, Q_2, \ldots, Q_x, \ldots, Q_z\} \tag{19}$$

Where, Q_x represents x^{th} solution, and z is total solutions.

2. Evaluation of fitness function

The fitness of each individual solution is computed shown in equation (18), and the solution that has the maximum fitness value is selected as the optimal solution.

3. Determination of update equation

The SSA is very competent and provides optimum performance while calculating the issues, and there is less parameter needed for fine-tuning and the SSA is recognised for its simpler algorithmic structure. As per the SSA algorithm, the updated solution is given as,

$$Q_o^A = \frac{1}{2}(Q_o^A + Q_o^{A-1}) \tag{20}$$

$$Q_o^A(\upsilon + 1) = \frac{1}{2}\left(Q_o^A(\upsilon) + Q_o^{A-1}(\upsilon)\right) \tag{21}$$

Where, $Q_o^A(\upsilon + 1)$ represents the update solution, $Q_o^A(\upsilon)$ denotes the current solution and $Q_o^{A-1}(\upsilon)$ is the previous solution. The LA is recommended by the researchers for solving the problems generated from the open-source optimization tools. Thus, the update equation of the LA is given by,

$$Q_o^A[(\upsilon)] = Q_\gamma^w(\upsilon) + (0.1\lambda_2 - 0.05)\left(Q_\gamma^\rho(\upsilon) - \lambda_1\left(Q_\gamma^w(\upsilon)\right)\right) \tag{22}$$

Where, $Q_\gamma^w(\upsilon)$ and $Q_\gamma^\rho(\upsilon)$ indicates the w^{th} and the ρ^{th} vector elements, and λ_1 λ_2 represents the random integers ranging in between [0, 1] and γ is the random integer obtained at interval υ. The final update equation is obtained by integrating the update position of SSA with the update position of LA. $Q_o^A(\upsilon + 1)$ is represents the update solution, and $Q_o^{A-1}(\upsilon)$ is the previous solution.

4. Terminate

The optimal solutions are derived iteratively until the maximum number of iterations. Thus, whenever there is a query for video object retrieval, the features of the input video frames are refined and fed to the NB classifier that processes the HOFM features of the detected objects from the input query with respect to the original indexed database and derives the relevant object and events for the concerned input video.

4. Results and Discussion

This section presents the results and discussion in terms of the performance metrics, and compares the results of proposed method with the existing methods to demonstrate the better performance of the proposed method.

4.1 Database description

The analysis of the proposed method is carried out with five videos taken from the CAVIAR dataset [33]. The CAVIAR dataset includes a collection of different videos containing different scenarios, such as one shop, walk by shop, walking, three persons walking, shopping and so on.

4.2 Experimental setup

The performance is evaluated using the MATLAB environment. Experimentation is done on a personal computer with Intel Core i-5 processor 4GB RAM and a 64-bit operating system.

4.3 Performance metrics

Precision: Precision in terms of the video retrieval system refers to the fraction of the most relevant trajectory of the object retrieved as compared to the search. A good retrieval mechanism depends on the retrieved relevant trajectory of the object in the video.

Recall: Recall refers to the fraction of the successfully retrieved trajectory that is relevant to the query trajectory. A good retrieval mechanism retrieves all possible trajectories relevant to the query.

F-measure: Fmeasure is defined as the harmonic mean of precision and recall.

$$F = \frac{2(PR)}{P+R} \tag{23}$$

Where P and R is the precision and recall.

4.4 Methods taken for comparison

The analysis is performed using a comparative method such as NSA+EWMA, Exponential weighted moving average (EWMA), Nearest Search Algorithm-based nonlinear autoregressive exogenous neural network (NSA+NARX).

NSA: NSA finds the position of the object by an extension parameter, which determines the location of the object in all the consecutive frames depends on the minimum distance value.

EWMA: In EWMA, the object is tracked depends on the location. EWMA determines the weighted function and exponential function for finding the position of the target object and tracks the path of the object in video frames.

NSA+EWMA: This work uses both NSA and EWMA for retrieving the video object. NSA and EWMA algorithms are hybridized for tracking the exact position of the object in the video frame [30].

NSA+NARX: This work uses both NSA and NARX to retrieve the video object and is averaged to determine the object tracking [34].

4.5 Experimental results

This section elaborates on the performance analysis using videos taken from the CAVIAR dataset. In this, the objects are tracked, and the relevant objects of the video are stored in the database. Fig. 2 represents input frame and a retrieved object based on HOFM.

4.6 Performance analysis

Fig. 3 shows the performance analysis of the proposed method based on Recall, F-measure, and Precision. The results of proposed LSSA-NB classifier method are compared with existing techniques like NSA, EWMA, NSA+EWMA and NSA+NARX.

The analysis of methods based on recall, F-measure, and precision is illustrated in Fig. 3. The analysis of methods in terms of precision parameter is demonstrated in Fig. 3a. When the retrieval objects are 2, the corresponding precision values computed by existing NSA is 78.134%, EWMA is 82.436%, NSA+EWMA is 82.451%, NSA+NARX is 84.434%, and proposed LSSA-NB is 82.985%, respectively. Likewise, when the retrieval objects are 4, the corresponding precision values computed by existing NSA is 72.956%, EWMA is 72.106%, NSA+EWMA is 77.145%, NSA+NARX is 77.985%, and proposed LSSA-NB is 81.985%, respectively.

The analysis of methods based on the recall parameter is depicted in Fig. 3b. When the retrieval objects are 2, the corresponding recall values computed by existing NSA is 82.853%, EWMA is 77.847%, NSA+EWMA is 83.091%, NSA+NARX is 86.324%, and proposed LSSA-NB is 86.451%, respectively. Likewise, when the retrieval objects are 4, the corresponding recall values computed by existing NSA is 82.853%, EWMA is 77.847%, NSA+EWMA is 83.091%, NSA+NARX is 86.324%, and proposed LSSA-NB is 87.451%, respectively.

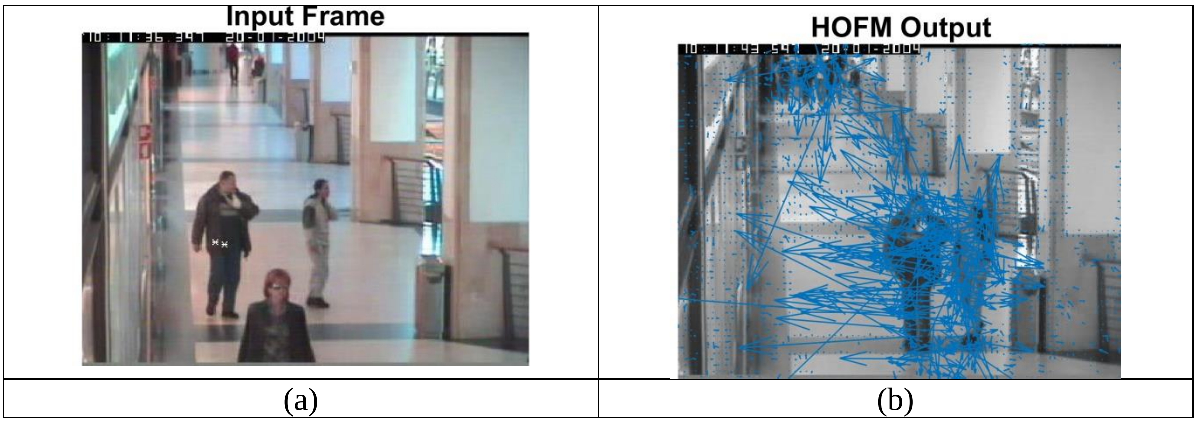
The analysis of methods based on the F-measure parameter is depicted in Fig. 3c. When the retrieval objects are 2, the corresponding F-measure values computed by existing NSA is 82.775%, EWMA is 77.160%, NSA+EWMA is 82.935%, NSA+NARX is 84.399%, and proposed LSSA-NB is 87.847%, respectively. Likewise, when the retrieval objects are 4, the corresponding F-measure values computed by existing NSA is 77.598%, EWMA is 72.136%, NSA+EWMA is 80.487%, NSA+NARX is 82.847%, and proposed LSSA-NB is 86.847%, respectively.

4.7 Comparative discussion

Table 1 discusses the comparative analysis of the existing and proposed methods based on maximal performance using recall, F-measure, and precision values. The analysis is done by analyzing five videos using performance measures. The highest precision values achieved by existing NSA is 72.956%, EWMA is 72.106%, NSA+EWMA is 77.145%, NSA+NARX is 77.985%, whereas the proposed LSSA-NB is 82.985%, respectively. The highest recall values achieved by existing NSA is 77.623%, EWMA is 75.347%, NSA+EWMA is 80.310%, NSA+NARX is 82.451%, whereas the proposed LSSA-NB is 87.451%, respectively. The highest F-measure values achieved by existing NSA is 77.598%, EWMA is 72.136%, NSA+EWMA is 80.487%, NSA+NARX is 82.847%, whereas the proposed LSSA-NB is 87.847%, respectively.

5. Conclusion

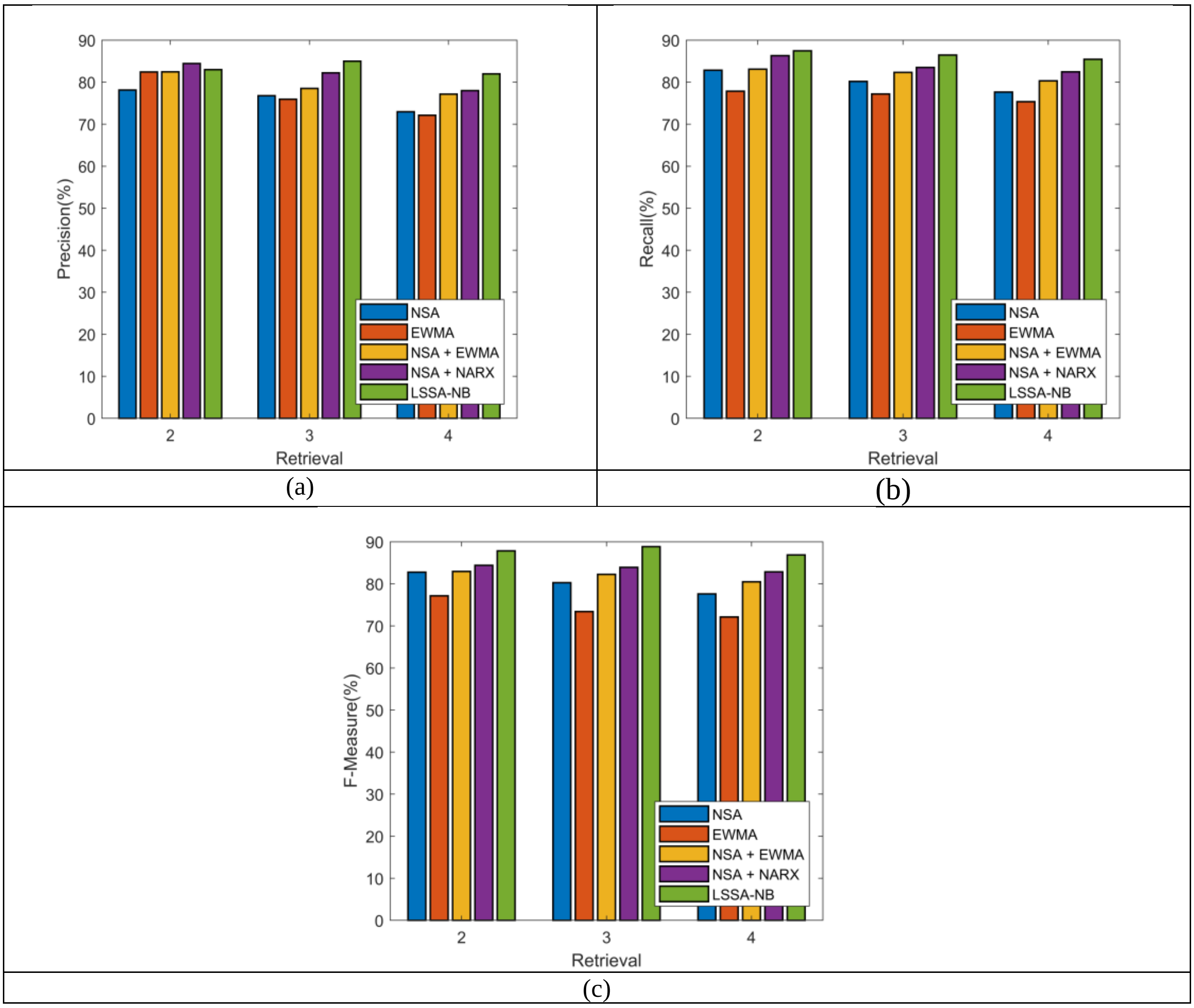
Video object retrieval is proposed for tracking the objects from the videos. The proposed method is devised in three phases, namely object tracking, event extraction, and retrieval. In object tracking, the location of the video object is detected by employing a hybrid model named NSA-SVR; the spatial tracking is done using SVR, whereas visual tracking done for retrieving the relevant objects. The tracked objects are indexed in the database by applying the HOFM to each tracked object in such a way that the Naive Bayes classifier is trained using an indexed database. In the video object retrieval, the query of the tracked object is given to the classifier, wherein the matching is done for retrieving the relevant objects. Then, the identified events related to the query are retrieved using the proposed LSSA-NB classifier. The results of the proposed method prove the effectiveness of the method. This proposed method is better with respect to the metrics with greater values of precision, recall and F-measure as 82.985%, 87.451%, and 87.847%, respectively.



(a) (b)

Figure. 2**.** Experimental results of proposed LSSA-NB classifier using

(a) input frame, and (b) retrieved object based on HOFM output



(a) (b)

(c)

Figure. 3 Analysis of methods using: (a) precision, (b) recall, and (c) f-measure

Received: November 23, 2020. Revised: March 2, 2021.

Table 1. Comparative analysis

Videos	Metrics	NSA	EWMA	NSA+EWMA	NSA+NARX	Proposed LSSA-NB
Using video 1	**Precision**	71.367	71.155	72.685	79.008	81.008
	Recall	74.903	72.318	75.122	79.793	82.793
	F-measure	71.863	73.565	75.082	77.142	79.808
Using video 2	**Precision**	72.956	72.106	77.145	77.985	82.985
	Recall	77.571	72.140	77.996	81.662	84.662
	F-measure	77.598	72.136	80.487	82.847	87.847
Using video 3	**Precision**	71.442	72.885	72.650	78.402	80.688
	Recall	77.145	72.565	77.145	77.751	80.751
	F-measure	73.685	71.723	77.498	79.082	81.991
Using video 4	**Precision**	72.902	72.771	74.448	75.470	77.755
	Recall	77.623	75.347	80.310	82.451	87.451
	F-measure	75.662	73.572	77.285	77.362	80.271
Using video 5	**Precision**	73.281	71.439	73.842	79.190	81.857
	Recall	72.565	75.079	77.360	77.602	80.602
	F-measure	74.227	74.376	74.545	79.755	82.955

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st author. The supervision and project administration have been done by 2nd and 3rd author.

References

[1] S. Hou, S. Zhou, and M. A. Siddique, "A compressed sensing approach for query by example video retrieval", Multimedia tools and applications, Vol. 72, No. 3, pp. 3031-3044, 2014.

[2] H. Y. Cheng, and J. N. Hwang, "Integrated video object tracking with applications in trajectory-based event detection", Journal of Visual Communication and Image Representation, Vol. 22, No. 7, pp. 673-685, 2011.

[3] J. Gong, and C. H. Caldas, "An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations", Automation in Construction, Vol. 20, No. 8, pp. 1211-1226, 2011.

[4] S. Gould, T. Gao, and D. Koller, "Region-based segmentation and object detection", Advances in Neural Information Processing Systems, Vol. 22, pp. 655-663, 2009.

[5] N. Zhang, and H. Y. Jeong, "A retrieval algorithm for specific face images in airport surveillance multimedia videos on cloud computing platform", Multimedia Tools and Applications, Vol. 76, pp. 17129-17143, 2017.

[6] H. Bilen, M. Pedersoli, and T. Tuytelaars, "Weakly supervised object detection with convex clustering", In: Proc. of the IEEE Conf. On Computer Vision and Pattern Recognition, Boston, MA, USA, pp. 1081-1089, 2015.

[7] S. D. Pande, M. S. R. Chetty, "Position Invariant Spline Curve Based Image Retrieval Using Control Points", International Journal of Intelligent Engineering and Systems, Vol. 12, No. 4, 177-191, 2019.

[8] F. F. Chamasemani, L. S. Affendey, N. Mustapha and F. Khalid, "Surveillance Video Retrieval Using Effective Matching Techniques", Fourth International Conf. On Information Retrieval and Knowledge Management, Kota Kinabalu, pp. 1-5, 2018.

[9] J. Gall, A. Yao, N. Razavi, L. Van Gool, and V. Lempitsky, "Hough forests for object detection, tracking, and action recognition", IEEE transactions on pattern analysis and machine

intelligence, Vol. 33, No. 11, pp. 2188-2202, 2011.

[10] B. Yellapragada, P. Rajaram, V. P. Sriram, S. Sengen, and B. P. Kolla, “Effective Handwritten Digit Recognition using Deep Convolution Neural Network”, International Journal of Advanced Trends in Computer Science and Engineering, Vol. 9, No. 2, pp. 1335-1339, 2020.

[11] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," In: Proc. IEEE Conf. On Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 6517-6525, 2017.

[12] G. Ning, Z. Zhang, C. Huang, X. Ren, H. Wang, C. Cai, and Z. He, “Spatially supervised recurrent convolutional neural networks for visual object tracking", In: Proc. of IEEE International Symposium on Circuits and Systems, Baltimore, MD, pp. 1-4, 2017.

[13] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real time object detection with region proposal networks”, Advances in Neural Information Processing Systems, Vol. 28, pp. 91-99, 2015.

[14] A. Salvador, X. Giró-i-Nieto, F. Marqués and S. Satoh,“Faster R-CNN Features for Instance Search”, In: Proc. IEEE Conf. On Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, pp. 394-401, 2016.

[15] G. Awate, S. Bangare, G. Pradeepini, and S. Patil, “Detection of Alzheimers Disease from MRI using Convolutional Neural Network with Tensorflow”, arXiv: 1806.10170, 2019.

[16] E. Jaynes, “Bayesian Methods: General Background Maximum Entropy and Bayesian Methods in Applied Statistics”, In: Proc. of the Fourth Maximum Entropy Workshop University of Calgary, Cambridge University Press, pp. 1-25,1986.

[17] M. Ismail, V. Vardhan, V. Mounika, and K. Padmini, “An Effective Heart Disease Prediction Method using Artificial Neural Network”, International Journal of Innovative Technology and Exploring Engineering, Vol. 8, pp. 1529-1532, 2019.

[18] S. Hussain, and Y. Prasanth, “A Panoramic Bayesian Analogy Based Method for Software Project Cost Estimation”, Asian Journal of Information Technology, Vol. 15, No. 3, pp. 647-650, 2016.

[19] K. V. Rajkumar, Y. Adimulam, and K. Subrahmanyam, “Fuzzy clustering and Fuzzy C-Means partition cluster analysis and validation studies on a subset of CiteScore dataset”, International Journal of Electrical and Computer Engineering, Vol. 9, No. 4, pp. 2760-2770, 2019.

[20] T. Lin, M. Yang, C. Tsai, and Y. F. Wang, “Query-adaptive multiple instance learning for video instance retrieval”, IEEE Transactions on Image Processing, Vol.24, No.4, pp.1330-1340, 2015.

[21] V. T. Nguyen, D. D. Le, M. T. Tran, T. V. Nguyen, T. D. Ngo, S. I. Satoh, and D. A. Duong, “Video instance search via spatial fusion of visual words and object proposals”, International Journal of Multimedia Information Retrieval, Vol. 8, pp. 181-192, 2019.

[22] T. Durand, X. He, I. Pop, and L. Robinault, “Utilizing Deep Object Detector for Video Surveillance Indexing and Retrieval”, In: Proc. of International Conf. On Multimedia Modeling, Springer, Cham, pp. 506-518, 2019.

[23] L. Czúni, and M. Rashad, “The use of IMUs for Video object retrieval in lightweight devices”, Journal of Visual Communication and Image Representation, Vol. 48, pp. 30-42, 2017.

[24] Y. Lou, Y. Bai, J. Lin, S. Wang, J. Chen, V. Chandrasekhar, L. Y. Duan, T. Huang, A. C. Kot, and W. Gao, “Compact deep invariant descriptors for video retrieval”, In: Proc. of Data Compression
Conf., Snowbird, UT, USA, pp. 420-429, 2017.

[25] J. Lin, L. Duan, S. Wang, Y. Bai, Y. Lou, V. Chandrasekhar, T. Huang, A. Kot, and W. Gao, “HNIP: Compact deep invariant representations for video matching, localization, and retrieval”, IEEE Transactions on Multimedia, Vol. 19, No. 9, pp. 1968-1983, 2017.

[26] N. Garcia, “Temporal aggregation of visual features for large-scale image-to-video retrieval”, In: Proc. of the ACM on International Conf. On Multimedia Retrieval, pp. 489-492, 2018.

[27] R. V. H. M. Colque, C. Caetano, M. T. L. de Andrade and W. R. Schwartz, "Histograms of Optical Flow Orientation and Magnitude and Entropy to Detect Anomalous Events in Videos", in IEEE Transactions on Circuits and Systems for Video Technology, Vol. 27, No. 3, pp. 673-682, 2017.

[28] R. Boothalingam, “Optimization using lion algorithm: a biological inspiration from lion’s social behavior”, Evolutionary Intelligence, Vol. 11, No. 2, pp. 31-52, 2018.

[29] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, “Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems”, Advances in

Engineering Software, Vol. 114, pp. 163-191, 2017.

[30] C. A. Ghuge, S. D. Ruikar, and V. C. Prakash, “Query–Specific Distance and Hybrid Tracking Model for Video Object Retrieval”, Journal of Intelligent Systems, Vol. 27, No. 2, pp. 195-212, 2018.

[31] K. S. Ni, and T. Q. Nguyen, “Image Superresolution Using Support Vector Regression”, in IEEE Transactions on Image Processing, Vol. 16, No. 6, pp. 1596-1610, 2007.

[32] K. Suppala, and N. Rao, “Sentiment Analysis Using Naïve Bayes Classifier”, International Journal of Innovative Technology and Exploring Engineering, Vol. 8, No. 8, pp. 264-269, 2019.

[33] CAVIAR database, https://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1

[34] C. A. Ghuge, V C. Prakash, S. D. Ruikar, “Weighed query-specific distance and hybrid NARX neural network for video object retrieval”, The Computer Journal, Vol. 63, No. 11, pp. 1738-1755, 2020