




An improved surveillance video synopsis framework: a HSATLBO optimization approach

Subhankar Ghatak¹  · Suvendu Rup¹ · Banshidhar Majhi² · M. N. S. Swamy³

Received: 31 August 2018 / Revised: 8 February 2019 / Accepted: 18 February 2019 /

Published online: 2 March 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Video surveillance cameras capture huge amount of data 24 hours a day. However, most of these videos contain redundant data which make the process difficult for browsing and analysis. A significant amount of research findings have been made in summarization of recorded video, but such schemes do not have much impact on video surveillance applications. On the contrary, video synopsis is a smart technology that preserves all the activities of every single object and projects them concurrently in a condensed time. The energy minimization module in video synopsis framework plays a vital role, which in turn minimizes the activity loss, number of collision and temporal consistency cost. In most of the reported schemes, Simulated Annealing (SA) algorithm is employed to solve the energy minimization problem. However, it suffers from slow convergence rate resulting in a high computational load to the system. In order to mitigate this issue, this article presents an improved energy minimization scheme using hybridization of SA and Teaching Learning based Optimization (TLBO) algorithms. The suggested framework for static surveillance video synopsis generation consists of four computational modules, namely, Object detection and segmentation, Tube formation, Optimization, and finally Stitching and the central focus is on the optimization module. Thus, the present work deals with an improved hybrid energy minimization problem to achieve global optimal solution with reduced computational time. The motivation behind hybridization (HSATLBO) is that TLBO algorithm has the ability to search rigorously, ensuring to reach the optimum solution with less computation. On the contrary, SA reaches the global optimum solution, but it may get disarrayed and miss some critical search points. Exhaustive experiments are carried out and results compared with that of benchmark schemes in terms of minimizing the activity, collision and temporal consistency costs. All the experiments are conducted on five widely used videos taken from standard surveillance video data set (PETS 2001, MIT Surveillance Dataset, ChangeDetection.Net, PETS 2006 and UMN Dataset) as well as one real generated surveillance video from the IIIT Bhubaneswar Surveillance Dataset. To make a fair comparison, additionally, performance of the proposed hybrid scheme to solve video synopsis optimization problem is also compared with that of the other benchmark functions. Experimental

✉ Subhankar Ghatak
subhankar@iiit-bh.ac.in

evaluation and analysis confirm that the proposed scheme outperforms other state-of-the-art approaches. Finally, the suggested scheme can be easily and reliably deployed in the off-line video synopsis generation.

Keywords Energy minimization · Optimization · Simulated annealing (SA) · Teaching learning based optimization (TLBO) · Video surveillance · Video synopsis

1 Introduction

Towards worldwide application of video surveillance solutions, the existence of CCTV camera is obvious. These enormous amount of cameras are working 24 hours a day and generate a gigantic amount of surveillance video data. With the advancement of camera technology, the quality and the use of surveillance cameras is increasing exponentially. To locate an event of interest manually is time consuming and strenuous because of the redundant nature of the surveillance video. So, there is a growing need to develop a smart technology, which in turn forms compact representation of the original video by preserving all the essential activities. Video synopsis is such an efficient and intelligent technology to generate shorter version of the original surveillance video through the projection of all important activities concurrently. Video synopsis also serves as an efficient index to the original video, thus making video browsing easy and simultaneously resolving the issue of storage scarcity.

Due to the absence of a standard to measure, evaluation of the quality and performance of a synopsis video is highly dependent on the application. Like other reported literature [16] in this field, the proposed work has also adopted the following qualitative measurement for the performance evaluation of a synopsis video:

- Preservation of maximum energy in a minimum time period.
- A minimum number of artifacts like collision among objects.
- Good-conservation of the original order of objects' appearance and their spatial positions.

In 2006, the authors in [32] introduced the concept of video synopsis and stated two different directions, namely, low-level video synopsis generation and object-based video synopsis generation. Gradually the efficiency and intellectuality of the object based approach to generate a video synopsis made itself the mainstream. The optimization module is considered to be the central block among the other building blocks (object detection and segmentation, tube generation, and stitching) of object-based video synopsis framework, since this module supervises the final length, activity preservation and visual quality of the synopsis video. However, the complexity of the optimization problem is not only high, but also quite extensive due to gigantic amount of combinations. For feasibility purpose, authors in [32], restricted the solution to two class, namely, lossy and lossless object-based video synopsis. These restricted solutions are then solved using SA [12]. Similarly, authors in [27], solved the optimization problem using greedy approach.

From quality enhancement point of view in video synopsis, many notable findings have been carried out in this field. One of the considered performance measures, namely, collision reduction, is addressed in [24] through the spatio-temporal rearrangement of objects. The same issue is also addressed through another innovative idea to reduce the size of objects in [16]. Similarly, [20, 50] focused on the issue of improved tube formation, while [36] focused on tracking strategies. One major limitation of video synopsis technology stated in [27] is that it is not applicable to crowded video, as the synopsis video is itself crowded than the original. To limit this crowded appearance of the synopsis, an abnormal activity

based synopsis generation is depicted in [18] and an event based one in [44]. Returning to the solution to the key module, energy minimization in video synopsis framework, a global energy function in [24] is solved by Graph Cut [2]-based technique. Blob Trajectory Optimization is employed in [18] to address the energy minimization problem and [15] employed a greedy approach. Maximum a Posteriori Probability Estimation is used for real-time synopsis generation in [9, 44]. However, SA is widely employed for most of the off-line synopsis generation schemes as in [16, 50] to solve the objective function. The iterative computational nature of SA makes the process lengthy and time-consuming to produce better results.

1.1 Motivation & overview

The video synopsis preparation is formulated as an energy minimization problem, which minimizes collision among objects, minimizes temporal consistency cost to maintain uniformity among objects' tracks, and minimizes activity loss with the inclusion of maximum objects' activities. The aforementioned costs are regulated by certain parameters like number of objects, object size, spatial distance between objects and tube length as per original video. These parameters are not consistent from one video to another, which makes the whole process highly arbitrary and minimizing the energy function complex. Hence, SA is an ultimate choice in most of the video synopsis approaches to address the above issues through its probabilistic selection process and its ability to reach a global optima in the presence of several local optima regardless of the objective function. Thus, the application of SA is very prominent to minimize the energy in the field of video synopsis. The nature of probabilistic selection of SA to reach the global optimal solution sometimes may get disarrayed and hence, miss some crucial search points. Various control parameters like population size, number of generations, and elite size, are required in all the probability grounded evolutionary and swarm intelligence-based algorithms. Along with these parameters, certain algorithms require other algorithm-specific parameters for regulating the optimization process. The effectiveness of these algorithms is highly dependent on the algorithm-specific control parameters and if not tuned properly, it may lead to increased computational time or a local optimal solution. Thus, TLBO [31] is a better choice for this. The purpose of using TLBO is that it has fewer number of adjusting parameters in comparison to traditional meta-heuristic optimization algorithms like Genetic Algorithm, Differential Evolution, and Particle Swarm Optimization etc. Further, TLBO shows faster convergence and is easier to implement.

Based on the above observations, authors are motivated to propose a hybrid optimization algorithm using SA and TLBO (HSATLBO) for solving the global energy function in object-based video synopsis framework. Usually, meta-heuristic techniques for any given optimization problem produce random set of solutions in a predefined range. These techniques are able to achieve the objectives, but they suffer from certain issues like getting stuck in a local minimum and/or time to reach the optimum solution. So, a hybrid technique is essential in this regard to combine and enhance the properties of the parent techniques.

Motivation for the hybridization is as follows:

- TLBO is one of the recent meta-heuristic algorithms which has been successfully applied to distinct optimization problems in various research areas of electrical engineering, mechanical design, thermal engineering, manufacturing engineering, civil engineering, structural engineering, computer engineering, electronics engineering, physics, chemistry, biotechnology and economics [30].

- The algorithm consists of two phases (i.e. Teacher and Learner Phase), which enhance the ability of the technique to search rigorously within the search space before moving onto the next population of students, thus ensuring to reach the optimum solution with less computations.
- The Simulated Annealing (SA) on the other hand undergoes a probabilistic selection depending on the temperature scheduled.
- Although the probabilistic selection empowers the technique to reach the global optimal solution but, it may get disarrayed and miss some critical search points. Here, TLBO technique further assists SA to perform a rigorous search within a certain search space prior to probabilistic selection. This reduces the possibility of missing an optimum point due to probabilistic selection and increases the effectiveness of the technique in achieving a global optimum solution.

1.2 Contributions

Video synopsis not only requires all activity by an object to be preserved within the shortest possible time length but also to satisfy the cost constraints. The contributions of the present work are as follows:

- An object-based video synopsis framework is formulated for static surveillance videos without missing any activity.
- A hybrid optimization scheme, namely HSATLBO, is formulated by combining the properties of both SA and TLBO to minimize the energy function in terms of activity, collision and temporal consistency costs.

1.3 Organizations

The remaining part of the article is organized as follows: Section 2 elaborates on the related literature study. Some preliminaries about SA and TLBO are summarized in Section 3. The suggested surveillance video synopsis framework is presented and elaborately discussed in Section 4. Section 5 critically discusses the experimental results and comparisons are made with that of the benchmark schemes. Finally, concluding remarks and scope for the future work are outlined in Section 6.

2 Related work

As mentioned in the previous section, optimization module is the key as it affects the overall efficacy of video synopsis framework. Hence, this section deals with some of the notable findings in the context of optimization module in video synopsis. Initially [32] proposed a dynamic video synopsis scheme in terms of low-level and object-based approach for video synopsis generation. In low-level approach, the optimization framework is formulated as a 3D Markov random field (MRF) [13] and minimized by graph cuts algorithm. For the case of object-based video synopsis, the moving objects are extracted from the video and then they are rearranged without maintaining the chronological order of the objects' appearance to create a short and seamless video synopsis. In object-based video synopsis, the problem of energy minimization is made feasible through two restricted solutions, one being video synopsis with a pre-determined length and the other being the lossless video synopsis. In both the cases, SA is employed. Authors in [26] proposed an object-based video synopsis methodology by expanding the work of [32] for the endless video stream captured

by webcam. In this method, the energy minimization is done through SA. A detail presentation of the work of [32] and [26] is presented in [27], where the energy minimization is done through greedy approach.

To further enhance the subjective performance of the synopsis video, authors in [24] proposed a video synopsis generation procedure through global spatial and temporal optimization, where the alpha-beta swap graph cut method is used for energy minimization. Similar to [24], authors in [16] proposed a novel scheme of object-based video synopsis, where the subjective performance of the synopsis video is enhanced via scaling down the object size. This methodology is an addition to pixel domain optimization and solved using SA.

Deviating from the aforementioned optimization techniques, blob trajectory optimization is used for synopsis generation in [18]. The problem of tube rearrangement in [44] is formulated as a maximum a posteriori probability estimation. In [15], seam carving is proposed for video synopsis generation process and optimized through a greedy algorithm. Dynamic programming is used in [48] and [39] for energy minimization. Though video synopsis technology is superior in summarizing, the shorter version is a bit confusing as it projects multiple activities simultaneously. To reduce the confusion level of the synopsis video, authors in [28] proposed a scheme to generate synopsis video through the grouping of similar activities and their projection. In this methodology, the problem of tube rearrangement is mapped through k-nearest neighbours algorithm and kd-tree implementation. In the work proposed by authors in [45], an object-based video synopsis is generated considering the objects as a spatio-temporal collection and minimization is done by mean shift algorithm.

An updated version of the energy function proposed in [26] is reflected in [40] by the addition of foreground importance to achieve multi-scale scalable browsing. Here, SA is used to optimize the function. Similarly, SA is also employed to minimize the energy function in [51] and [41]. Emphasizing on the issue of memory requirement and time cost behind tube generation, authors in [17] proposed an algorithm based on 3D graph cuts and realized synopsis video through blank frame deletion. Recently, the problem of synopsis generation has also been formulated in terms of scheduling problem and solved through trajectory scheduling algorithm [3] and graph coloring algorithm [7]. Genetic algorithm [47] and [37] is also explored in this field for tube rearrangement. Table 1 presents a comparison chart among several state-of-the-art techniques in the field of surveillance video synopsis with a central focus towards energy minimization, methodology used, and potential applications.

From the related works, it is observed that the outcome of video synopsis technology greatly depends on the central module, the so called optimization module. To improve the overall efficacy of the video synopsis generation, several attempts have been made in this direction. Hence, keeping this in mind, this article presents a hybrid optimization algorithm (HSATLBO) for object-based video synopsis to improve the overall performance of video synopsis generation as well as to minimize the activity, collision, and temporal consistency costs.

3 Preliminaries

3.1 Simulated annealing

Inspired from metallurgy, in 1953, authors in [21] proposed an algorithm for efficient molding of atoms in symmetry at a certain fixed temperature. A new state is designated by the change of a randomly nominated particle from an initial state with energy E_0 . Let, E be the

Table 1 Comparison of surveillance video synopsis techniques

Existing works	Methods	Energy minimization	Applications
Rav-Acha et al. [32]	Object-based: extracted moving objects rearranged without maintaining the chronological order of appearance	Simulated Annealing	Object based Video Synopsis Generation
Pritch et al. [26]	Object-based video synopsis generation based on user query for endless videos generated by webcams or surveillance cameras	Simulated Annealing	Query based WebcamVideo Synopsis Generation for Endless Videos
Pritch et al. [27]	Object detection, tracking, tube generation, and temporal shifting of object tubes to generate synopsis video	Greedy Algorithm	Non-chronological Video Synopsis and Video Indexing
Yildiz et al. [48]	Non-linear Image Resizing, Volume Projections and Surface Discarding and Synopsis Generation	Dynamic Programming	Real-Time Video Synopsis
Xu et al. [45]	Object start-time programming using set theory to maximize visual content in the synopsis video	Mean Shift Algorithm	Video Synopsis
Pritch et al. [28]	Clustering similar trajectories and produce short synopsis by temporal shifting of objects	K-Nearest Neighbors Algorithm	Non-chronological Video Synopsis
Vural and Akgul. [39]	Frequency-based background subtraction with eye gaze base tracking and synopsis generation	Dynamic Programming	Real-time Video Synopsis
Wang et al. [40]	Detailed based synopsis generation with storing and browsing of synopsis video based on object flags	Simulated Annealing	Video Synopsis and Scalable Browsing
Zhu et al. [51]	Video Synopsis generation based on key observation selection	Simulated Annealing	Video Synopsis
Wang et al. [41]	Object flag-based video coding and temporal shifting of objects for compressed video synopsis	Simulated Annealing	Video Synopsis and Indexing

Table 1 (continued)

Existing works	Methods	Energy minimization	Applications
Nie et al. [24]	Objects shifting in both spatial and temporal domain to generate compact video synopsis	Alpha-Beta Cut Swap Graph	Compact Video Synopsis
Yao et al. [47]	Object detection, tracking, tube generation, and temporal shifting of object tubes to generate synopsis video	Genetic Algorithm	Video Synopsis
Lin et al. [18]	Local-patch-learning-based abnormality detection, blob sequence optimization, and type-based surveillance videos synopsis generation	Blob Trajectory Optimization	Activity-based Video Synopsis
Chou et al. [3]	Object detection and tracking, grouping, trajectory clustering based synopsis generation	Trajectory Scheduling Algorithm	Event-based Video Synopsis
Li et al. [15]	Tube generation and Seam carving method for video condensation	Greedy Algorithm	Effective Video Synopsis
Li et al. [16]	Object resizing and shifting in temporal domain through pixel domain optimization to produce synopsis	Simulated Annealing	Surveillance Video Synopsis
Tian et al. [37]	Trajectory mapping model is utilized to keep important relationship between objects in video synopsis generation	Genetic Algorithm	Video Synopsis
Wang et al. [44]	Tube rearrangement based on kinematic events of trajectory and their clustering using affinity propagation	Maximum A Posteriori Probability Estimation	Event-based Video Synopsis
Liao et al. [17]	Synopsis analysis method based on blank frame deletion in the compressed domain	3D Graph Cuts Algorithm	Synopsis analysis and video browsing
He et al. [7]	Graph coloring-based scheduling problem to generate optimized video synopsis	Graph Coloring Algorithm	Video synopsis and condensation

energy of the current state and for $E_0 > E$, E is being nominated along with a new next state generation. Otherwise if, $E_0 \leq E$, the likelihood to persist in this new state is given by:

$$\exp(-(E - E_0)/k_b T) \quad (1)$$

where T is the current temperature and k_b is Boltzmann Constant. This probabilistic acceptance strategy is called Metropolis criterion. The major limitation of the Metropolis approach is that it is based on a single fixed temperature. To eliminate this, authors in [12] globalized it by integrating an annealing schedule to decrease the temperature. Beginning with a high initial temperature, it drops according to the annealing schedule and the step iterates until the system freezes. The system finally freezes in a state of global minimum energy.

3.2 TLBO algorithm

Authors in [31] have proposed a meta-heuristic optimization based on the process of teaching and learning. It considers the significance of a teacher's presence in the class and the teacher's ability to influence the students to improve in the respective subjects. The individual factors to be optimized are considered as subjects and the result obtained in each subject is the optimized parameter value. The optimization process undergoes two phase, namely, Teacher phase and Learner phase.

This technique operates with a population size P_{max} (i.e. apparent optimum state, $k = 1, 2, \dots, P_{max}$) and N number of decision variables $j = 1, 2, \dots, N$ for every individual of the population. After evaluating the initial set of population, the set with the best solution is designated as the teacher solution T_k^{best} . The mean M_j is obtained for all students in each subject. For any i^{th} iteration, the difference between the mean and the teacher solution is calculated by (2) and added to the individual student solution to update them with respect to the teacher solution given as in (3).

$$DM_{j,k,i} = r_i(T_{j,k,i}^{best} - T_f M_{j,i}) \quad (2)$$

$$X_{j,k,i}^{mod} = X_{j,k,i} + DM_{j,k,i} \quad (3)$$

Here, r_i is a random value in the interval of $[0, 1]$ and T_f is considered either 1 or 2 based on (4).

$$T_f = \text{round}[1 + \text{rand}(0, 1)\{2 - 1\}] \quad (4)$$

On completion of teacher phase the updated solution is passed on to the student phase, where two individual learners m, n are chosen and a self-learning process is undertaken based on the equation

$$X_{j,m,i}^{learner} = X_{j,m,i}^{mod} + r_i |X_{j,n,i}^{mod} - X_{j,m,i}^{mod}| \quad (5)$$

Thus, an optimum result is obtained at the end of the iteration which pushes the subsequent population of next iteration to an even improved solution until the termination criterion is fulfilled.

4 Surveillance video synopsis framework

Initially the input video is preprocessed for the purpose of detection and segmentation of object activities. Subsequently, tubes are formulated for each object as per the tracked information. Further, the objective function is formulated to evaluate the various penalty costs involved in synopsis generation. Next, the HSATLBO scheme is proposed to simplify the objective function, which is the weighted sum in terms of activity, collision, and temporal consistency costs. Finally, the optimized object tubes and the background are blended together to produce the final synopsis video. The flow diagram of the proposed framework is depicted in Fig. 1.

4.1 Object detection and segmentation

Generally, surveillance videos are of long duration and consists of either static or dynamic backgrounds. The proposed approach concentrates on the surveillance videos with static background. In such surveillance videos, the frames usually change due to the movements of the observed objects and illumination. Taking these facts into account along with the view point of video synopsis generation, the problem of object detection and segmentation focused on the extraction of moving foregrounds like other object based video synopsis generation approaches such as [16, 27], and [24]. Similar to [16], the proposed scheme adopted the qualitative criteria that are exploited for the preservation of major activities in video synopsis field and hence several state-of-the-art [43, 49], and [19] in the field of object detection and segmentation are not suitable for the proposed framework. In the suggested scheme, a robust multi-layer background subtraction algorithm [46] is employed to extract the moving objects along with a modeling of the background. The multi-layer background subtraction algorithm efficiently extracts foreground through the power of local binary pattern (LBP) features and photometric invariant color measurements in RGB color

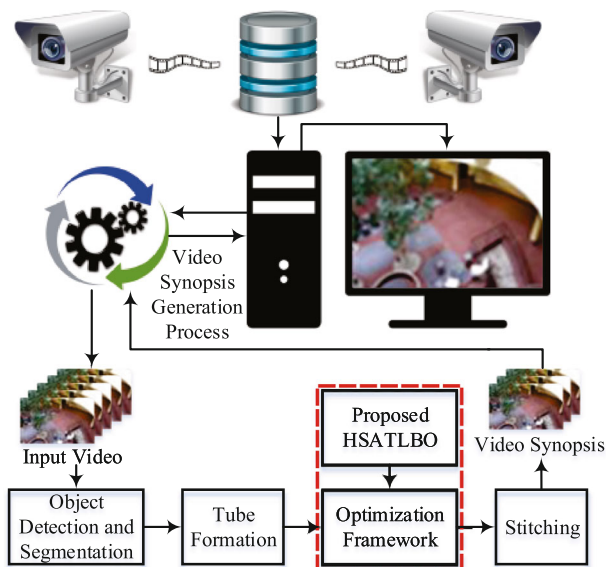


Fig. 1 Flow diagram of the proposed surveillance video synopsis framework

space. Further, the morphological dilation and erosion are applied on the resulted foreground frames to reduce any noise, if present. The final resultant foreground masks are passed to blob analyzer for the detection of moving objects. In this work, the original video is mathematically represented in terms of a space-time volume, denoted as $V(x, y, t)$, where t is the temporal and x, y are the spatial indices of each frame. The resulting background model is denoted as $B(x, y, t')$, where $t' \leq t$ and the detected and segmented moving objects are represented through precise and independent bounding boxes.

4.2 Tube formation

Tube formation is another important preprocessing module of the video synopsis generation framework. Efficient track information of the individual moving object is called a tube. The quality of the synopsis video not only depends on efficient object detection and segmentation, but also on efficient tracking. In the proposed methodology, similar to [16], Kalman Filter [1] is employed for multi-object tracking. The track or tube is composed of the set of all connected bounding boxes in the time axis. The next module, i.e. the optimization module works on the tubes and decides the temporal shift applied to each tube for the minimization of energy.

4.3 Proposed optimization framework

The proposed optimization framework deals with multiple objectives that are associated with the video synopsis technology. The objectives are activity cost, collision cost, and temporal consistency cost, which are discussed below. The temporal shift for each tube is decided by the optimization module and it guarantees the minimal cost for each of the objectives considered. To solve the formulated optimization problem, HSATLBO, that produces an optimized time location for each object tube to start, is proposed so as to reduce the cost asserted. Finally, stitching of the objects as per the time location completes the synopsis process. f_μ , a mapping of the original video to the synopsis video is executed and E is defined by the following equation, which represents the total penalty cost incurred for the process.

$$E(f_\mu) = E_a(f_\mu) + \omega_0 E_c(f_\mu) + \omega_1 E_t(f_\mu) \quad (6)$$

The weights ω_0 and ω_1 determine the precedence of the individual cost functions while formulating a single objective function. A higher value would signify a high precedence of the cost over other penalties. The following subsection elaborates the mathematical formulation of the various energy costs involved.

4.3.1 The energy costs

Activity cost $E_a(f_\mu)$ is the penalty imposed for dropping any activity from the original video.

$$E_a(f_\mu) = \sum_{o \in O} E_a(o^s) \quad (7)$$

Where, O is a set comprising of all the objects' tubes, whereas o represents an individual tube present in the original video. Here, o^s is a mapping of o pertaining to the time position obtained for formulating the synopsis video. The term $E_a(o^s)$ is defined as:

$$\sum_{o \in O} E_a(o^s) = \sum_{o \in O} \sum_{\hat{o} \in o^s \wedge \hat{o} \notin synopsis} absDiff(\hat{o}) \quad (8)$$

$$absDiff(\hat{o}) = \sum_{(x,y) \in bbox(\hat{o})} ||V(x, y, t_{\hat{o}}) - B(x, y, t_{\hat{o}})|| \tag{9}$$

Where, $absDiff()$ is a function for evaluating the absolute difference between the element \hat{o} and its corresponding background and \hat{o} a part of o^s . In (9), V and B denote the original video frames and the corresponding background model, respectively and $bbox()$ stands for bounding box function. Equation (8) summarizes all the active pixels not included in the synopsis video. Activity cost would be zero for a scenario, where all events are conserved in the synopsis video (i.e. lossless synopsis).

Collision cost Considering the size of the synopsis video as compared to the original video size, collisions among object tubes are obvious. $E_c(f_{\mu})$ is the penalty imposed for suffering any collision among the objects. Minimizing the collision count guarantees enhanced synopsis quality. Let o_m^s & o_n^s be the mapping to the synopsis video from the original video elements, o_m & o_n ; the collision between them is defined as

$$E_c(f_{\mu}) = \sum_{o_m, o_n \in O} E_c(o_m^s, o_n^s) \tag{10}$$

$$E_c(o_m^s, o_n^s) = \sum_{\hat{o}_m \in o_m^s, \hat{o}_n \in o_n^s} Area(bbox(\hat{o}_m) \cap bbox(\hat{o}_n)) \tag{11}$$

Where, the $Area()$ function calculates the overlap region between the bounding box of \hat{o}_m and \hat{o}_n . If a video produced has no collisions, the cost is considered to be zero.

Temporal consistency cost The procedure for the generation of synopsis video may cause violation of the temporal associations among captured objects due to the time-based alteration. $E_t(f_{\mu})$ is the penalty for mislaying the chronology and relative appearance among objects with respect to the original video. The order of appearance of the objects are mapped from the source video and the penalty aims at minimizing the deviation from the original order. To realize the penalty cost for temporal inconsistency, this work follows the following two situations. In the first situation, any two objects in the original video share some common frames. In the second situation, two objects are not sharing any common frames. The second situation can be further sub-classified into two classes. In the first class, the consistency with respect to chronology as well as entry difference is violated. The other class deals with the case where the same is preserved. These situations can efficiently address the issue of preservation of interaction among objects in the synopsis if it exists in the original video.

Spatial relationships (Δ) between two moving objects can be utilized to measure their probability of interaction. This measurement represents the first considered situation. The spatial relationships between two moving objects, say o_m and o_n , is established as:

$$\Delta(o_m, o_n) = \exp(-min_{t \in t_{o_m} \cap t_{o_n}} (\delta(o_m, o_n, t))/\sigma) \tag{12}$$

Here, the Euclidean distance between the objects o_m and o_n from the original video in t^{th} frame is represented as $\delta(o_m, o_n, t)$, where, σ is used to tune the level of space communication between o_m and o_n .

A high value of σ may lead to a situation where the higher cost will be assigned for the violation of temporal consistency. Hence, the optimization process attempts to minimize the cost, even though there may be very little possibility of interaction between the objects. Thus, in this scenario the cost allocation is ambiguous due the high value of σ . On the other hand, a low value of σ drives the situation to assign a low temporal consistency cost which is insignificant in the minimization process. But, in this case, there might be a high possibility

of interaction between objects which may be neglected. Henceforth, in our experiment, we have assumed 40 as an intermediate value of σ .

Likewise, the temporal relationships (τ) between objects can be expressed as:

$$\tau(o_m, o_n) = \|(\vec{t}_{o_m} - \vec{t}_{o_n}) - (\vec{t}_{o_m^s} - \vec{t}_{o_n^s})\| \tag{13}$$

Here, $\vec{t}_{o_m^s}$ and $\vec{t}_{o_n^s}$ are the entry frame indices in the synopsis video and \vec{t}_{o_m} and \vec{t}_{o_n} are that of o_m and o_n , respectively, in the original video.

In second the situation, the measurement of the temporal violation ($\hat{\tau}$) between o_m and o_n is defined as:

$$\hat{\tau}(o_m, o_n) = (\vec{t}_{o_m} - \vec{t}_{o_n}) \times (\vec{t}_{o_m^s} - \vec{t}_{o_n^s}) \tag{14}$$

The positive value of $\hat{\tau}$ implies successful preservation of temporal consistency between o_m and o_n in the synopsis incurring zero penalty cost. Thus, based on the considered situations, the complete representation of the temporal consistency cost is represented as:

$$E_t(f_\mu) = \sum_{o_m, o_n \in O} E_t(o_m^s, o_n^s) \tag{15}$$

$$E_t(o_m^s, o_n^s) = \begin{cases} \Delta(o_m, o_n) \times \tau(o_m, o_n) & \text{if } (t_{o_m} \cap t_{o_n}) \neq \phi \\ \exp(\|(\vec{t}_{o_m^s} - \vec{t}_{o_n^s})\|/\gamma) & \text{if } \hat{\tau}(o_m, o_n) \leq 0 \\ 0 & \text{Otherwise} \end{cases} \tag{16}$$

Here, the time of interaction of the objects, o_m and o_n , in the original video is denoted by $t_{o_m} \cap t_{o_n}$. Now, there can be two cases, first, $t_{o_m} \cap t_{o_n}$ is not an empty set, i.e. the objects have shared some common frames in the original video. In this case, $\Delta(o_m, o_n)$ and $\tau(o_m, o_n)$ are utilized to preserve the temporal consistency. The equal value of the individual time of intervals $(\vec{t}_{o_m} - \vec{t}_{o_n})$ and $(\vec{t}_{o_m^s} - \vec{t}_{o_n^s})$ makes $\tau(o_m, o_n)$ zero, thereby resulting in the first term, $\Delta(o_m, o_n) \times \tau(o_m, o_n)$, to be zero which implies that the temporal consistency is well preserved. On the other hand, for the non-zero values of $\tau(o_m, o_n)$, $\Delta(o_m, o_n)$ is used to regulate the temporal consistency measure. Violation of temporal relationship is crucial for close distant objects, resulting in high temporal consistency cost.

Under the second case, $t_{o_m} \cap t_{o_n}$ is an empty set, there can two sub-cases. First, if $\hat{\tau}(o_m, o_n) \leq 0$, the temporal consistency violation is proportional to the exponential growth of penalty cost in terms of the absolute difference of the temporal violation, normalized by the factor γ , with respect to the synopsis video. γ is the normalization factor, defining the time interval in which events still have temporal communication. Based on the nature of the considered videos for experiments, authors have assumed the value of γ as 20. Second, for $\hat{\tau}(o_m, o_n) > 0$, the temporal consistency cost is zero as the positive value of $\hat{\tau}(o_m, o_n)$ implies successful preservation of temporal consistency in the synopsis video.

4.3.2 Energy minimization

Minimization of the energy function in (6) is a tedious job as it searches through an enormous amount of odds. In [16], authors have addressed this issue with the application of Simulated Annealing, while authors in [47] have used Genetic Algorithm for the same. The base line for any optimization technique remains the same (namely, to generate a solution with least possible cost and in the smallest possible time for computation). The technique must be able to verify all possible mapping μ , and produce a temporal shift which is capable

of optimizing all the individual objectives in (6). To initiate the process it is assumed that all the objects start from the first frame and the cost function value is evaluated. The upper bound is decided by the difference in length of the tube for a certain object and the synopsis length. It can be user defined, but to study a lossless structure it is set to be a fixed value (i.e. length of the longest object tube). To meet the requirement of the complex objective function and to justify the search space, an efficient algorithm is suggested for solving (6) by the hybridization of Simulated Annealing and Teaching Learning Based Optimization (HSATLBO).

4.3.3 Proposed HSATLBO

The proposed hybrid technique is presented in Algorithm 1. In the initialization step, all the required parameters are initiated as described in Step-1. A set of Initial population μ^T is produced for each object which is being shifted within a range $[0, (maxlen - objlen)]$ as described in Step-2, where $f_{\mu k}$ represents each population as a row vector of μ^T . The decision variables, x_j , are the elements of $f_{\mu k}$. Finally, the activity, collision & temporal consistency costs, as well as the corresponding objective function values E_1, E_2, \dots, E_k for each k^{th} population is evaluated initially according to (6), as given in Step-3. As per the TLBO technique, for the solution with best fitness (E^{best}), the corresponding decision variable value (T^{best}) is designated as the Teacher. The rest of the population is termed as student who would learn and update themselves in regard to teacher solution as described in the teacher phase. The mean M_j is calculated as stated in Step-4 and the teacher phase is initiated as depicted in Steps 6-16. For each k^{th} population, teaching factor T_f is calculated as shown in Step-7. Each element of μ^T is updated as depicted in Steps 8 and 9 to produce $\mu_{j,k}^{new}$ and the new set of decision variables are applied. From (6), E_k^{new} is evaluated as described in Step-10. Now the process for selection or rejection of the updated solution is done in the selection phase, which is represented in Steps 11-16. For the purpose of achieving a global optimum solution, selection steps of simulated annealing is applied instead of the conventional technique of TLBO. After selection step, the obtained E^{best} and μ^T are passed to the student phase for further processing.

In the student phase, two students $f_{\mu m}$ and $f_{\mu n}$ are selected such that $f_{\mu m} \neq f_{\mu n}$. $f_{\mu m}^{new}$ is generated in Steps 17-23 to accomplish a mutual learning process. The selection at the end of the student phase is similar to that of the teacher phase, where simulated annealing is applied for that purpose. The teacher phase and the student phase are repeated until 'Gen' and 'T' values are within the specified limits.

To make a fair comparison, the efficacy of the proposed optimization algorithm, HSATLBO, is tested on standard benchmark functions as listed in Table 2 and the analysis results in terms of Best, Worst, Mean and Standard Deviation based on 10 independent runs with 100 iterations and population size 10 for unconstrained as well as constrained single objective functions are displayed in Tables 3, 4, and 5, respectively. From Tables 3 and 4, it can be seen that the proposed hybrid optimization algorithm, HSATLBO, yields better results compared to that of Grey Wolf Optimizer (GWO) [22], NSGA II [4], JAYA [29], SA, and TLBO to solve unconstrained single objective benchmark functions. Similarly, from Table 5, it is evident that the proposed HSATLBO outperforms GWO, NSGA II, JAYA, SA, and TLBO to address constrained single objective benchmark functions. Thus, the proposed HSATLBO can be applied for the minimization of the objective functions in connection with the problem of off-line single view surveillance video synopsis generation.

Algorithm 1 Proposed HSATLBO Algorithm

```

Input :  $V(x, y, t)$ ; //Original Video
         Tube; //Object Track
Output:  $f_{\mu}^{best}$ ; //A Set of Temporal Shifts

1 Initialization:
   objno; //Number of moving objects in original video
   objlen; //Corresponding Object's length
   maxlen; //Length of longest tube (Synopsis length)
    $T = 1000$ ; //Initial Temperature
    $e = 0.01$ ; //Change in fitness value
    $T_s = 0.99$ ; //Temperature schedule
   Gen = 1; //Iteration counter
   Genmax = 100; //Maximum number of iterations
    $P_{max} = 50$ ; //Population size

2 Generation of Initial Population:
    $x_j = \text{randi}([0 \text{ (maxlen - objlen)}]), \forall j = 1, \dots, \text{objno}$ 
    $f_{\mu k} = [x_1, x_2, \dots, x_j], \forall k = 1, \dots, P_{max}$ , where,  $\mu^T = [f_{\mu 1}, f_{\mu 2}, \dots, f_{\mu k}]$ ;

3 Evaluation of Initial Population:
    $E^T = [E_1, E_2, \dots, E_k]$ ;
   where  $E_k = E(f_{\mu k})$ ; using  $V(x, y, t)$  and Tube as per Equation (6);
    $E^{best} = \min(E^T)$ ;  $T^{best} = \mu^T(\text{index} - \text{of}(\min(E^T)))$ ;

4  $M_j = \sum_k x_j / P_{max}$ ;
5 for Gen = 1 to Genmax do
   // Teacher Phase:
6   for  $k = 1$  to  $P_{max}$  do
7      $T_f = \text{randi}([1 \ 2])$ ; // Teaching factor
8     for  $j = 1$  to objno do
9        $\mu_{j,k}^{new} = \text{round}(\mu_{j,k} + \text{rand}(\text{objno}) \cdot (T_j^{best} - T_f \cdot M_j))$ ;
10       $E_k^{new} = E(f_{\mu k}^{new})$ ; using  $V(x, y, t)$  and Tube as per Equation (6).
       // Selection for Teacher Phase:
11      if  $E_k^{new} \leq E^{best}$  then
12         $T^{best} = f_{\mu k} = f_{\mu k}^{new}$ ;  $E^{best} = E_k = E_k^{new}$ ;
13      else
14         $r = \text{rand}(1, 1)$ ;
15        if  $r \leq \exp(-((E_k^{new} - E^{best})/E^{best})/T)$  then
16           $T^{best} = f_{\mu k} = f_{\mu k}^{new}$ ;  $E^{best} = E_k = E_k^{new}$ ;

       // Student Phase:
17      for  $m = 1$  to  $P_{max}$  do
18        Select  $f_{\mu_n}$  randomly such that,  $f_{\mu_m} \neq f_{\mu_n}$ ;
19        if  $E(f_{\mu_m}) < E(f_{\mu_n})$  then
20           $f_{\mu_m}^{new} = \text{round}(f_{\mu_m} + \text{rand}(\text{objno}) \cdot (f_{\mu_m} - f_{\mu_n}))$ ;
21        else
22           $f_{\mu_m}^{new} = \text{round}(f_{\mu_m} + \text{rand}(\text{objno}) \cdot (f_{\mu_m} - f_{\mu_n}))$ ;
23         $E_m^{new} = E(f_{\mu_m}^{new})$ ; using  $V(x, y, t)$  and Tube as per Equation (6).
       // Selection for Student Phase:
24        if  $E_m^{new} \leq E^{best}$  then
25           $f_{\mu}^{best} = f_{\mu_m} = f_{\mu_m}^{new}$ ;  $E^{best} = E_m = E_m^{new}$ ;
26        else
27           $r = \text{rand}(1, 1)$ ;
28          if  $r \leq \exp(-((E_m^{new} - E^{best})/E^{best})/T)$  then
29             $f_{\mu}^{best} = f_{\mu_m} = f_{\mu_m}^{new}$ ;  $E^{best} = E_m = E_m^{new}$ ;

30    $T = T \times T_s$ ; Gen = Gen + 1;

```

Table 2 Single objective unconstrained (F_1 to F_{10}) and constrained (F_{11} to F_{15}) standard benchmark functions used for the analysis

Function name	Formula
Ackley [10] (F_1)	$-20 \exp(-0.2 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}) - \exp(\frac{1}{n} \sum_{i=1}^n \cos 2\pi x_i) + 20 + e$
Beale [10] (F_2)	$(1.5 - x + xy)^2 + (2.25 - x + xy^2)^2$
Booth [10] (F_3)	$(x + 2y - 7)^2 + (2x + y - 5)^2$
Goldstein-Price [10] (F_4)	$\left[1 + (x + y + 1)^2 (19 - 14x + 3x^2 - 14y + 6xy + 3y^2) \right]$ $\left[30 + (2x - 3y)^2 (18 - 32x + 12x^2 + 48y - 36xy + 27y^2) \right]$
Lévi (F_5)	$\sin^2 3\pi x + (x - 1)^2 (1 + \sin^2 3\pi y) + (y - 1)^2 (1 + \sin^2 2\pi y)$
Noisy quadric [11] (F_6)	$\sum_{i=1}^n (ix_i)^4 + \text{random}(0, 1)$
Rastrigin [11] (F_7)	$\sum_{i=1}^n (x_i^2 - 10 \cos 2\pi x_i + 10)$
Rosenbrock [11] (F_8)	$\sum_{i=1}^n (100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2)$
Sphere [11] (F_9)	$\sum_{i=1}^n (x_i^2)$
Step [11] (F_{10})	$\sum_{i=1}^n (x_i + 0.5)^2$
Rosenbrock (Disk) [10] (F_{11})	$(1 - x)^2 + 100(y - x^2)^2$, subjected to: $x^2 + y^2 \leq 2$
Rosenbrock (Cubic & Line) [11] (F_{12})	$(1 - x)^2 + 100(y - x^2)^2$, subjected to: $(x - 1)^3 - y + 1 \leq 0$ and $x + y - 2 \leq 0$
Simionescu [33] (F_{13})	$0.1xy$, subjected to: $x^2 + y^2 \leq \left[r_T + r_S \cos \left(n \arctan \frac{x}{y} \right) \right]^2$ where: $r_T = 1, r_S = 0.2$ and $n = 8$ where: $r_T = 1, r_S = 0.2$ and $n = 8$
Townsend (F_{14})	$-\left[\cos((x - 0.1)y)^2 - x \sin(3x + y) \right]$, subjected to: $x^2 + y^2 < \left[2 \cos t - \frac{1}{2} \cos 2t - \frac{1}{4} \cos 3t - \frac{1}{8} \cos 4t \right]^2 + [2 \sin t]^2$, where: $t = \text{Atan}2(x, y)$
Mishra's Bird [23] (F_{15})	$\sin(y)e^{[1-\cos(x)]^2} + \cos(x)e^{[1-\sin(y)]^2} + (x - y)^2$, subjected to: $(x + 5)^2 + (y + 5)^2 < 25$

4.4 Stitching

The stitching module is considered to be the final module of the video synopsis framework and used for the visualization of the final synopsis. The optimization framework generates the temporal shifts for all moving objects and according to the corresponding temporal shifts the objects are projected in the synopsis. Application of multi-layer background subtraction algorithm [46] for object detection and segmentation also produces the corresponding background frames parallel to the generation of foreground masks. These background frames are considered for making the background video for the synopsis onto which the objects are

Table 3 Implementation of various optimization techniques to standard benchmark functions $F_1 - F_5$ (single objective unconstrained functions)

Function name	Search domain	Optimization technique	Best	Worst	Mean	Standard deviation
F_1	$-5 \leq x, y \leq 5$	GWO	8.153×10^{-03}	9.491×10^{-01}	4.277×10^{-02}	6.519×10^{-02}
		NSGA II	1.695×10^{-03}	1.941×10^{-01}	5.002×10^{-02}	5.898×10^{-02}
		JAYA	8.889×10^{-05}	2.622×10^{-03}	7.472×10^{-04}	8.154×10^{-04}
		SA	4.214×10^{-03}	1.619×10^{-02}	9.108×10^{-03}	3.919×10^{-03}
		TLBO	9.148×10^{-03}	1.225×10^{-01}	4.833×10^{-02}	4.370×10^{-02}
F_2	$-4.5 \leq x, y \leq 4.5$	HSATLBO	2.220×10^{-15}	9.970×10^{-13}	1.417×10^{-13}	3.053×10^{-13}
		GWO	6.915×10^{-06}	1.296×10^{-03}	8.958×10^{-02}	2.339×10^{-02}
		NSGA II	3.645×10^{-07}	1.096×10^{-00}	1.583×10^{-01}	3.351×10^{-01}
		JAYA	4.588×10^{-02}	3.310×10^{-01}	5.890×10^{-00}	1.043×10^{-01}
		SA	4.287×10^{-06}	7.622×10^{-01}	7.669×10^{-02}	2.409×10^{-01}
F_3	$-10 \leq x, y \leq 10$	HSATLBO	7.896×10^{-13}	1.032×10^{-03}	2.179×10^{-04}	4.231×10^{-04}
		GWO	7.217×10^{-07}	1.273×10^{-03}	9.315×10^{-03}	4.479×10^{-03}
		NSGA II	4.810×10^{-06}	2.633×10^{-01}	5.878×10^{-02}	8.689×10^{-02}
		JAYA	7.751×10^{-10}	2.024×10^{-05}	2.503×10^{-06}	6.396×10^{-06}
		SA	1.807×10^{-05}	6.320×10^{-03}	1.072×10^{-03}	1.917×10^{-03}
F_4	$-2 \leq x, y \leq 2$	HSATLBO	2.408×10^{-18}	1.865×10^{-02}	2.637×10^{-03}	5.782×10^{-03}
		GWO	5.761×10^{-00}	5.734×10^{-04}	9.950×10^{-05}	1.796×10^{-04}
		NSGA II	3.000×10^{-00}	4.527×10^{-01}	2.637×10^{-01}	1.941×10^{-00}
		JAYA	3.050×10^{-00}	3.015×10^{-01}	5.722×10^{-00}	8.583×10^{-00}
		SA	3.000×10^{-00}	8.401×10^{-01}	1.150×10^{-01}	2.548×10^{-01}
F_5	$-10 \leq x, y \leq 10$	TLBO	3.001×10^{-00}	8.402×10^{-01}	1.110×10^{-01}	2.562×10^{-01}
		HSATLBO	3.000×10^{-00}	4.258×10^{-00}	3.258×10^{-00}	4.334×10^{-01}
		GWO	1.569×10^{-05}	3.000×10^{-00}	3.000×10^{-00}	1.138×10^{-04}
		NSGA II	1.105×10^{-08}	1.491×10^{-01}	5.292×10^{-02}	5.119×10^{-02}
		JAYA	4.550×10^{-09}	1.099×10^{-01}	1.730×10^{-02}	3.397×10^{-02}
		SA	2.543×10^{-05}	3.093×10^{-03}	3.407×10^{-04}	9.690×10^{-04}
		TLBO	1.591×10^{-05}	7.254×10^{-03}	1.101×10^{-03}	2.219×10^{-03}
		HSATLBO	2.736×10^{-10}	1.161×10^{-01}	3.396×10^{-02}	4.229×10^{-02}
						2.718×10^{-05}

Bold Signifies the optimized value (i.e. minimum)

Table 4 Implementation of various optimization techniques to standard benchmark functions $F_6 - F_{10}$ (single objective unconstrained functions)

Single objective unconstrained functions						
Function name	Search domain	Optimization technique	Best	Worst	Mean	Standard deviation
F_6	$-10 \leq x, y \leq 10$	GWO	1.956×10^{-02}	5.731×10^{-01}	5.927×10^{-02}	5.824×10^{-02}
		NSGA II	1.277×10^{-01}	5.250×10^{-01}	5.222×10^{-01}	1.179×10^{-01}
		JAYA	$4.914 \times 10^{+00}$	$5.987 \times 10^{+01}$	$2.852 \times 10^{+01}$	$2.048 \times 10^{+01}$
		SA	3.509×10^{-02}	1.351×10^{-01}	7.361×10^{-02}	3.160×10^{-02}
		TLBO	2.166×10^{-02}	9.194×10^{-02}	4.189×10^{-02}	2.208×10^{-02}
F_7	$-5.12 \leq x_i \leq 5.12$	HSATLBO	7.313×10^{-03}	3.285×10^{-02}	1.791×10^{-02}	6.944×10^{-03}
		GWO	2.425×10^{-03}	$6.352 \times 10^{+00}$	9.319×10^{-01}	$1.628 \times 10^{+00}$
		NSGA II	1.076×10^{-04}	4.522×10^{-01}	8.908×10^{-01}	1.414×10^{-01}
		JAYA	$1.779 \times 10^{+01}$	$6.043 \times 10^{+09}$	$1.016 \times 10^{+09}$	$2.006 \times 10^{+09}$
		SA	1.321×10^{-01}	$7.967 \times 10^{+01}$	$1.387 \times 10^{+01}$	$2.484 \times 10^{+01}$
F_8	$-\infty \leq x_i \leq \infty$,	TLBO	1.269×10^{-01}	$3.165 \times 10^{+01}$	$6.030 \times 10^{+00}$	$9.248 \times 10^{+00}$
		HSATLBO	2.141×10^{-05}	6.935×10^{-01}	2.727×10^{-01}	2.837×10^{-01}
		GWO	2.573×10^{-02}	$6.405 \times 10^{+00}$	$7.280 \times 10^{+00}$	4.051×10^{-02}
		NSGA II	1.930×10^{-01}	$4.138 \times 10^{+01}$	$1.551 \times 10^{+01}$	$1.562 \times 10^{+01}$
		JAYA	$1.779 \times 10^{+01}$	$6.043 \times 10^{+09}$	$1.016 \times 10^{+09}$	$2.006 \times 10^{+09}$
F_9	$1 \leq i \leq n$	SA	1.321×10^{-01}	$7.967 \times 10^{+01}$	$1.387 \times 10^{+01}$	$2.484 \times 10^{+01}$
		TLBO	1.269×10^{-01}	$3.165 \times 10^{+01}$	$6.030 \times 10^{+00}$	$9.248 \times 10^{+00}$
		HSATLBO	2.141×10^{-05}	6.935×10^{-01}	2.727×10^{-01}	2.837×10^{-01}
		GWO	2.704×10^{-03}	3.620×10^{-01}	9.061×10^{-02}	3.973×10^{-02}
		NSGA II	1.268×10^{-07}	4.607×10^{-01}	5.385×10^{-02}	1.439×10^{-01}
F_{10}	$-\infty \leq x_i \leq \infty$,	JAYA	4.304×10^{-02}	$5.366 \times 10^{+03}$	$1.349 \times 10^{+03}$	$1.731 \times 10^{+03}$
		SA	2.614×10^{-04}	1.438×10^{-02}	2.910×10^{-03}	4.260×10^{-03}
		TLBO	2.519×10^{-03}	6.503×10^{-01}	1.331×10^{-01}	2.154×10^{-01}
		HSATLBO	1.040×10^{-29}	1.137×10^{-23}	2.142×10^{-24}	3.812×10^{-24}
		GWO	4.691×10^{-03}	2.830×10^{-01}	4.229×10^{-02}	1.072×10^{-02}
	$1 \leq i \leq n$	NSGA II	9.146×10^{-07}	8.373×10^{-02}	3.107×10^{-02}	3.246×10^{-02}
		JAYA	4.891×10^{-01}	$9.429 \times 10^{+03}$	$2.326 \times 10^{+03}$	$3.667 \times 10^{+03}$
		SA	8.725×10^{-04}	7.442×10^{-03}	3.665×10^{-03}	2.100×10^{-03}
		TLBO	7.521×10^{-07}	$1.036 \times 10^{+00}$	1.790×10^{-01}	3.152×10^{-01}
		HSATLBO	3.102×10^{-15}	1.103×10^{-04}	2.919×10^{-05}	3.353×10^{-05}

Bold Signifies the optimized value (i.e. minimum)

Table 5 Implementation of various optimization techniques to standard benchmark functions $F_{11} - F_{15}$ (single objective constrained functions)

Single objective constrained functions						
Function name	Search domain	Optimization technique	Best	Worst	Mean	Standard deviation
F_{11}	$-1.5 \leq x \leq 1.5$,	GWO	7.501×10^{-03}	4.914×10^{-01}	5.093×10^{-02}	1.339×10^{-02}
		NSGA II	7.442×10^{-02}	5.740×10^{-01}	3.349×10^{-01}	1.999×10^{-01}
		JAYA	3.387×10^{-05}	1.174×10^{-02}	2.608×10^{-03}	3.627×10^{-03}
		SA	1.639×10^{-04}	2.039×10^{-02}	3.468×10^{-03}	6.240×10^{-03}
		TLBO	2.606×10^{-05}	2.041×10^{-02}	7.776×10^{-03}	6.683×10^{-03}
F_{12}	$-1.5 \leq x \leq 1.5$,	HSATLBO	7.994×10^{-07}	1.181×10^{-03}	7.043×10^{-03}	2.130×10^{-03}
		GWO	1.959×10^{-03}	2.629×10^{-01}	1.837×10^{-02}	4.380×10^{-02}
		NSGA II	1.713×10^{-02}	4.136×10^{-01}	2.637×10^{-02}	6.383×10^{-01}
		JAYA	2.954×10^{-05}	7.306×10^{-02}	1.142×10^{-02}	2.239×10^{-02}
		SA	1.837×10^{-05}	1.139×10^{-02}	2.518×10^{-03}	4.286×10^{-03}
F_{13}	$-1.25 \leq x, y \leq 1.25$	TLBO	3.146×10^{-04}	7.123×10^{-02}	1.315×10^{-02}	2.211×10^{-02}
		HSATLBO	2.659×10^{-08}	1.552×10^{-03}	2.735×10^{-04}	5.479×10^{-04}
		GWO	-7.172×10^{-02}	-6.372×10^{-02}	-7.390×10^{-02}	5.198×10^{-03}
		NSGA II	-7.198×10^{-02}	-6.921×10^{-02}	-7.113×10^{-02}	8.155×10^{-04}
		JAYA	-7.169×10^{-02}	-6.511×10^{-02}	-6.925×10^{-02}	1.955×10^{-03}
F_{14}	$-2.25 \leq x \leq 2.5$,	SA	-7.199×10^{-02}	-7.117×10^{-02}	-7.181×10^{-02}	2.423×10^{-04}
		TLBO	-7.186×10^{-02}	-6.508×10^{-02}	-7.025×10^{-02}	2.073×10^{-03}
		HSATLBO	-7.200×10^{-02}	-7.194×10^{-02}	-7.198×10^{-02}	1.912×10^{-05}
		GWO	$-1.734 \times 10^{+00}$	$-1.381 \times 10^{+00}$	$-1.619 \times 10^{+00}$	4.194×10^{-02}
		NSGA II	$-1.840 \times 10^{+00}$	$-1.639 \times 10^{+00}$	$-1.669 \times 10^{+00}$	6.075×10^{-02}
F_{15}	$-2.5 \leq y \leq 1.75$	JAYA	$-1.889 \times 10^{+00}$	$-1.639 \times 10^{+00}$	$-1.677 \times 10^{+00}$	7.484×10^{-02}
		SA	$-1.660 \times 10^{+00}$	$-1.640 \times 10^{+00}$	$-1.654 \times 10^{+00}$	9.579×10^{-03}
		TLBO	$-1.921 \times 10^{+00}$	$-1.649 \times 10^{+00}$	$-1.722 \times 10^{+00}$	1.061×10^{-01}
		HSATLBO	$-2.018 \times 10^{+00}$	$-1.640 \times 10^{+00}$	$-1.693 \times 10^{+00}$	1.142×10^{-01}
		GWO	$-1.067 \times 10^{+02}$	$-1.165 \times 10^{+02}$	$-1.291 \times 10^{+02}$	4.943×10^{-02}
$-10 \leq x \leq 0$,	$-6.5 \leq y \leq 0$	NSGA II	$-1.068 \times 10^{+02}$	$-1.487 \times 10^{+00}$	$-8.605 \times 10^{+01}$	$4.372 \times 10^{+01}$
		JAYA	$-1.050 \times 10^{+02}$	$-9.323 \times 10^{+01}$	$-1.014 \times 10^{+02}$	$3.583 \times 10^{+00}$
		SA	$-1.068 \times 10^{+02}$	$-5.274 \times 10^{+00}$	$-8.646 \times 10^{+01}$	$4.279 \times 10^{+01}$
		TLBO	$-1.068 \times 10^{+02}$	$-1.064 \times 10^{+02}$	$-1.067 \times 10^{+02}$	1.022×10^{-01}
		HSATLBO	$-1.068 \times 10^{+02}$	$-1.068 \times 10^{+02}$	$-1.068 \times 10^{+02}$	2.732×10^{-06}

Bold Signifies the optimized value (i.e. minimum)



Fig. 2 Experimental setup

stitched. The background video length is dependent on the synopsis length, where the synopsis length is user dependent or as per the length of the largest tube to preserve all activities. Uniform temporal sampling is employed on the background frames to generate time lapsed synopsis background. This scheme helps to preserve the lightning conditions successfully in the generated synopsis background video. The corresponding time stamp for each tube is generated and stitched together with the tubes on the synopsis background frames applying the temporal shifts using Poisson Image Editing [25].

5 Experimental results and discussions

Exhaustive experiments are carried out on numerous surveillance videos for the performance evaluation of the proposed framework. MATLAB (version 2017a) with Intel Core I7 3.2 GHz CPU with 32 GB RAM is employed in the development and testing phase of the proposed approach. Figure 2 shows the experimental setup for the proposed work. The description of the set of video data, as well as the experimental results and analysis are discussed herein.

5.1 Experimental video data set

Videos having different frame rates, video lengths (i.e. number of frames) and number of objects are considered for the validation of the proposed methodology. In Table 6 various features of the videos considered are presented along with a snapshot of the first frame.

Videos considered for the testing purpose have different types of objects. The 1st video (Atrium) observes human beings walking around the field taken from [15]. The 2nd video (Pedestrians taken from ChangeDetection.NET [42]), captured within a park, involves walking as well as cycling action of humans. The 3rd video is a scene from MIT campus (PETS 2001 [5]) with a group of people moving around; it also contains tree branches and car movement. The 4th one is a real generated video taken from the IIIT Bhubaneswar surveillance dataset. The 5th video is captured in a railway station, projecting various human beings including long activity (PETS 2006 [6]) and the 6th one is depicting an outdoor crowded scenario of University of Minnesota (DS1/UMN/1 taken from UMN Dataset [38]).

Table 6 Parameters of experimental surveillance video

Video number	Video length (# Frames)	Frame rate (fps)	Number of objects	First frame snap shot
1	600	30	6	
2	1066	30	5	
3	2688	30	12	
4	3810	10	30	
5	1200	10	12	
6	625	30	16	

5.2 Simulation results

5.2.1 Evaluation of the preprocessing modules (object detection & segmentation and tube formation)

In the suggested scheme, Multi-layer background subtraction algorithm is employed in the object Detection and Segmentation phase. Figure 3 shows a detailed visual analysis of various object detection and segmentation techniques applied to the considered experimental data set. Here, A(1), B(1), C(1), D(1), E(1), and F(1) are the input frames from original video (1st row); A(2), B(2), C(2), D(2), E(2), and F(2) are the resultant frames for background subtraction method (2nd row), where background is modeled by executing the temporal median for the video clip of 30 sec (15 seconds before and 15 seconds after); A(3), B(3), C(3),

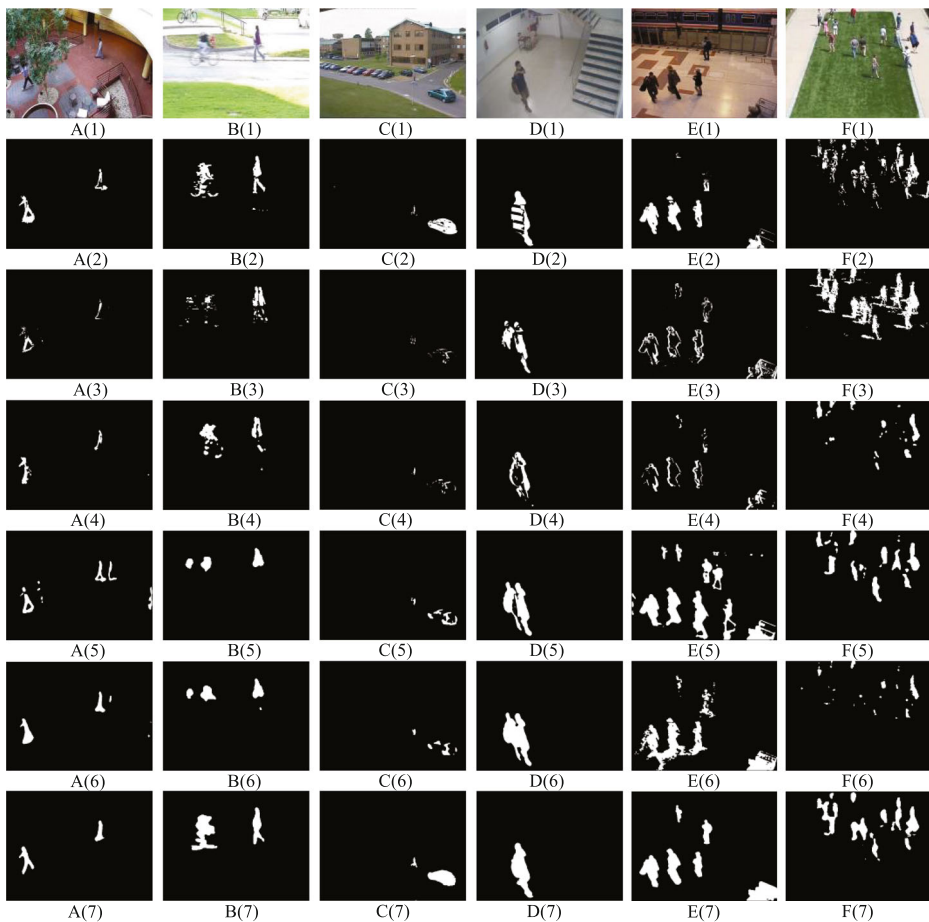


Fig. 3 The input frame from the considered videos and the resultant foregrounds obtained from various object detection and segmentation algorithms. Columns A, B, C, D, E, and F corresponds to video 1, 2, 3, 4, 5, and 6, respectively. Rows 2-7 presents the resultant foreground frames while the 1st row depicts the input frames

D(3), E(3), and F(3) are the resultant frames obtained by Gaussian mixture model (GMM) [14] (3rd row); A(4), B(4), C(4), D(4), E(4), and F(4) are the resultant frames obtained by PBAS [8] (4th row); A(5), B(5), C(5), D(5), E(5), and F(5) are the resultant frames obtained by LOBSTERBGS [34] (5th row); A(6), B(6), C(6), D(6), E(6), and F(6) are the resultant frames obtained by SuBSENSE [35] (6th row) and finally, A(7), B(7), C(7), D(7), E(7), and F(7) are the resultant frames obtained by applied multi-layer background subtraction algorithm [46] (7th row) for the proposed framework.

It can be observed from Fig. 3 (A(7), B(7), C(7), D(7), E(7), and F(7)) that the resulted foreground masks obtained through the application of multi-layer background subtraction algorithm outperforms in terms of detected results using the other approaches (results depicted in Fig. 3 (A(2-6), B(2-6), C(2-6), D(2-6), E(2-6), and F(2-6))). Additionally, it is also able to remove noise and artifacts like ghosting and shadows as compared to its counterparts. Due to the fact that the performance of the tube formation is greatly affected by the results of object detection and segmentation phase, the selection of the foreground mask

generation algorithm is crucial. The presence of shadow and noise in the foreground lead to sudden appearance of insignificant activities in the synopsis video. On the other hand, broken object trajectories on the foreground drives a single object extraction as multiple. This phenomenon affects the tracking process and restricts the tracker from generating a continuous tube, leading to unexpected disruption of appearance of objects in the synopsis video.

In the tube formation phase, like other approaches [16, 24, 27], the widely-applied Kalman filter-based multi-object tracker is employed. It is noted from the literature that in the process of tracking related to video synopsis, Kalman filter-based multi-object tracker is widely accepted, which in turn successfully addresses the occlusion problem. Thus, similar to other object-based video synopsis generation techniques [16, 24, 27], this work exploits Kalman filter-based multi-object tracker for tracking process. It should be noted that if the original video contains partial or full occlusion of moving objects, it will be reflected in the synopsis video as well, because of the fact that the occluded visual information of that moving object is missing in the original video.

5.2.2 Evaluation of proposed algorithm for the optimization framework

The proposed algorithm (HSATLBO) is implemented for the energy minimization (given by (6)), which generates the optimized video synopsis. The corresponding values of the biasing co-efficients ω_0 and ω_1 are determined to be 0.8 and 0.2, respectively, through extensive experimentation. Collision cost is assigned a higher bias as per sheer choice of the user to eliminate collision which causes loss of object activity data.

A comparative experimental study is performed with various optimization techniques (i.e. SA, TLBO, JAYA, NSGA II, and GWO) to authenticate the potential and precision of HSATLBO for the optimization of (6) and the results presented in Table 7. The chosen population size is 10 and maximum number of cost function evaluation is 100. In this table, the individual costs and the corresponding fitness values along with the execution time are compiled. The length of the synopsis produced is assigned to be equal to that of the longest object tube, thus a zero activity cost is obtained.

A statistical analysis in terms of Best, Mean, Worst, Standard Deviation and average Execution Time is presented in Table 8 for all the optimization techniques applied to videos 1, 2, 3, 4, 5, and 6, respectively. This analysis is performed using 10 independent runs for each algorithm and the best convergence plot is represented in Fig. 4. Additionally, Table 9 reflects a comparison between the proposed HSATLBO algorithm, and SA to solve the optimization framework proposed in the best state-of-the-arts [32] and [16]. The results obtained are presented in terms of the fitness value, time of execution, and final synopsis length (in terms of number of frames). During simulations of [32] and [16], the weights associated to collision and temporal consistency costs, are assumed to be 0.8 and 0.2, respectively.

In summary, from Tables 7 and 8, it is evident that the proposed algorithm outperforms the other optimization techniques considered for solving the proposed objective function. Moreover, it is observed from Table 9 that the proposed scheme is superior for solving the state-of-the-arts optimization frameworks in [32] and [16]. The visual quality assessment in terms of collision avoidance is shown in Fig. 5. The resultant synopsis frames present in all the odd rows of Fig. 5 are the outcome of the methodology proposed in [32], whereas those present in all the even rows are the outcome of the proposed methodology. In Fig. 5, the red circle and yellow circle signify the resultant collision from [32] and the proposed method, respectively. Except for the second row of Fig. 5, there is no yellow circle present in other rows indicating less number of collisions obtained from the proposed scheme. Thus, the

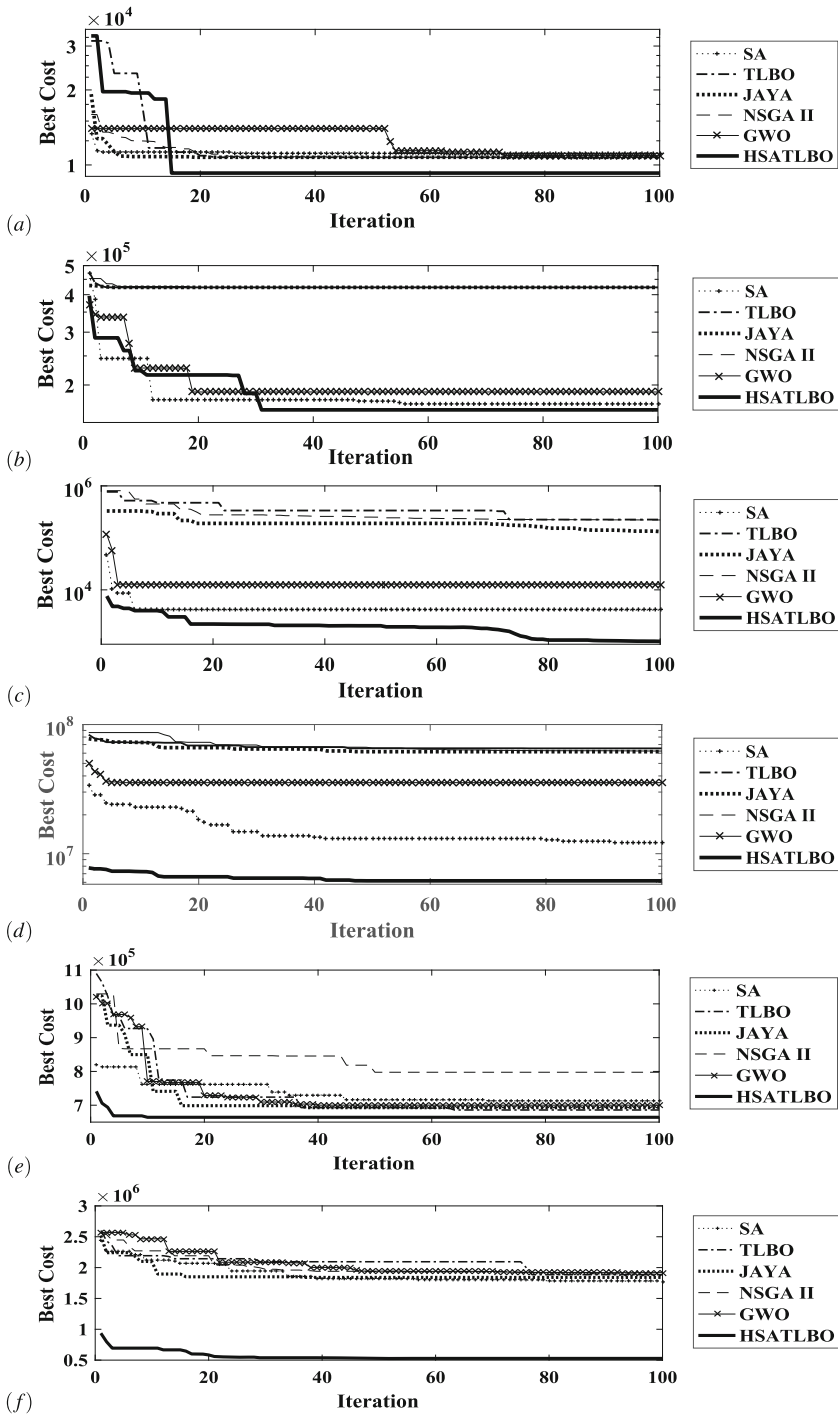


Fig. 4 Comparison of convergence characteristics for various optimization techniques (SA, TLBO, JAYA, NSGA II, GWO, and HSATLBO) applied to **a** Video-1 **b** Video-2 **c** Video-3 **d** Video-4 **e** Video-5 **f** Video-6

Table 7 Performance comparison among various optimization techniques

Video-1	Optimization techniques	Activity cost	Collision cost ($\times 10^3$)	Temporal consistency cost	Fitness value ($\times 10^3$)	Time of execution (Sec)
Video-1	SA	0	00013.77	009.69	00011.19	0207.18
	TLBO	0	00013.21	008.40	00010.73	0094.13
	JAYA	0	00013.21	008.40	00010.73	0041.28
	NSGA II	0	00013.21	008.38	00010.73	0047.09
	GWO	0	00013.44	008.63	00010.92	0074.90
Video-2	HSATLBO	0	00011.42	004.32	00009.27	0303.64
	SA	0	00212.85	009.77	00172.90	0120.39
	TLBO	0	00520.75	011.53	00423.01	0053.23
	JAYA	0	00520.75	011.53	00423.01	0026.89
	NSGA II	0	00520.75	011.53	00423.01	0033.12
Video-3	GWO	0	00233.43	010.90	00189.62	0046.54
	HSATLBO	0	00203.43	008.89	00165.25	0170.79
	SA	0	00005.01	036.75	00004.08	1213.30
	TLBO	0	00276.72	032.51	00224.78	0512.52
	JAYA	0	00162.33	048.37	00131.87	0269.04
Video-4	NSGA II	0	00272.19	041.29	00221.10	0266.65
	GWO	0	00015.81	028.39	00012.85	0397.49
	HSATLBO	0	00001.25	000.36	00001.02	0756.27
	SA	0	14965.34	294.99	12156.40	1015.00
	TLBO	0	80458.35	256.35	65356.36	0442.28
Video-5	JAYA	0	07770.55	322.65	61885.00	0149.50
	NSGA II	0	76742.08	308.15	62337.65	0247.75
	GWO	0	43681.65	312.48	35482.67	0280.53
	HSATLBO	0	07586.81	334.71	06162.77	1661.23
	SA	0	00877.52	066.32	00712.82	0694.32
Video-6	TLBO	0	00844.32	049.46	00685.85	0298.85
	JAYA	0	00852.67	087.63	00692.64	0148.48
	NSGA II	0	00981.85	092.28	00797.57	0278.95
	GWO	0	00861.89	079.96	00700.13	0293.76
	HSATLBO	0	00817.72	058.33	00664.25	0987.43
Video-6	SA	0	02185.36	081.65	01775.18	1886.27
	TLBO	0	02320.97	072.94	01885.34	0560.75
	JAYA	0	02266.31	064.60	01840.93	0270.56
	NSGA II	0	02326.40	090.98	01889.75	0598.24
	GWO	0	02361.68	081.24	01918.41	0395.23
	HSATLBO	0	00649.87	096.30	00527.91	1967.87

Bold Signifies the optimized value (i.e. minimum)

Table 8 Comparative statistical analysis of various optimization techniques

	Optimization techniques	Statistical measures					Average execution time (Sec)
		Best ($\times 10^3$)	Mean ($\times 10^3$)	Worst ($\times 10^3$)	Standard deviation ($\times 10^3$)		
Video-1	SA	00011.19	00014.30	00019.39	0002.86	0202.69	
	TLBO	00010.73	00012.48	00014.40	0001.53	0096.52	
	JAYA	00010.73	00014.30	00020.43	0003.57	0044.68	
	NSGA II	00010.73	00015.02	00020.94	0004.33	0045.63	
	GWO	00010.92	00014.45	00019.50	0005.69	0081.29	
	HSATLBO	00009.27	00014.11	00025.76	0045.75	0295.35	
Video-2	SA	00172.90	00276.80	00364.36	0054.02	0117.81	
	TLBO	00423.01	00424.83	00440.89	0005.64	0051.94	
	JAYA	00423.01	00426.61	00442.68	0007.64	0026.55	
	NSGA II	00423.01	00454.01	00583.30	0041.11	0032.96	
	GWO	00189.62	00423.82	00435.38	0029.84	0047.49	
	HSATLBO	00165.25	00214.49	00281.72	0035.22	0169.21	
Video-3	SA	00004.08	00009.18	00018.78	0005.45	1150.33	
	TLBO	00224.78	00298.06	00341.17	0037.12	0508.85	
	JAYA	00131.87	00226.47	00389.91	0089.27	0259.01	
	NSGA II	00221.10	00303.37	00454.78	0069.52	0270.13	
	GWO	00012.85	00015.71	00019.58	0021.31	0372.59	
	HSATLBO	00001.02	00002.49	00003.42	0000.87	0801.98	
Video-4	SA	12156.40	17750.41	22076.02	2886.88	1033.79	
	TLBO	65356.36	68308.66	69779.14	1223.55	0451.72	
	JAYA	61885.00	66565.01	69183.41	2315.49	0114.43	
	NSGA II	62337.65	68689.44	72483.18	2913.78	0250.90	
	GWO	35482.67	36219.17	41297.73	1594.25	0297.17	
	HSATLBO	06162.77	06487.64	06923.66	0235.23	1929.89	
Video-5	SA	00712.82	00764.73	00797.57	0034.07	0654.92	
	TLBO	00685.85	00727.28	00805.34	0049.49	0285.65	
	JAYA	00692.64	00712.74	00731.04	0016.28	0140.78	
	NSGA II	00797.57	00829.46	00857.28	0040.47	0270.69	
	GWO	00700.13	00706.89	00713.06	0004.79	0287.47	
	HSATLBO	00664.25	00674.07	00680.57	0008.89	0952.53	
Video-6	SA	01775.18	01816.46	01860.92	0034.01	1684.36	
	TLBO	01885.34	02009.86	02113.29	0090.08	0490.46	
	JAYA	01840.93	01963.34	02137.45	0114.13	0252.40	
	NSGA II	01889.75	01964.18	02034.66	0063.57	0577.86	
	GWO	01918.41	02036.62	02100.29	0083.85	0378.64	
	HSATLBO	00527.91	00740.66	01041.15	0204.44	1956.24	

Bold Signifies the optimized value (i.e. minimum)

Table 9 Performance comparison for solving state-of-the-art optimization frameworks

	Optimization techniques	Fitness value [32] ($\times 10^3$)	Time of execution (Sec) [32]	Fitness value [16] ($\times 10^3$)	Time of execution (Sec) [16]	Video length (# Frames)	Synopsis length [16, 32] (# Frames)
Video-1	SA	0011.19	0207.18	0.666	0350.85	600	221
	HSATLBO	0009.27	0303.64	0.552	0414.20	600	221
Video-2	SA	0172.90	0120.39	4.514	0194.18	1066	234
	HSATLBO	0165.25	0170.79	4.314	0235.47	1066	234
Video-3	SA	0004.08	1213.30	0.248	1866.05	2688	858
	HSATLBO	0001.02	0756.27	0.062	1163.14	2688	858
Video-4	SA	12156.40	1015.00	738.92	1561.06	3810	1200
	HSATLBO	06162.77	1661.23	374.59	2554.96	3810	1200
Video-5	SA	00712.82	0694.32	042.43	1175.80	1200	467
	HSATLBO	00664.25	0952.53	039.55	1613.15	1200	467
Video-6	SA	01775.18	1886.27	105.65	3194.38	625	470
	HSATLBO	00527.91	1967.87	031.43	3332.56	625	470

Bold Signifies the optimized value (i.e. minimum)

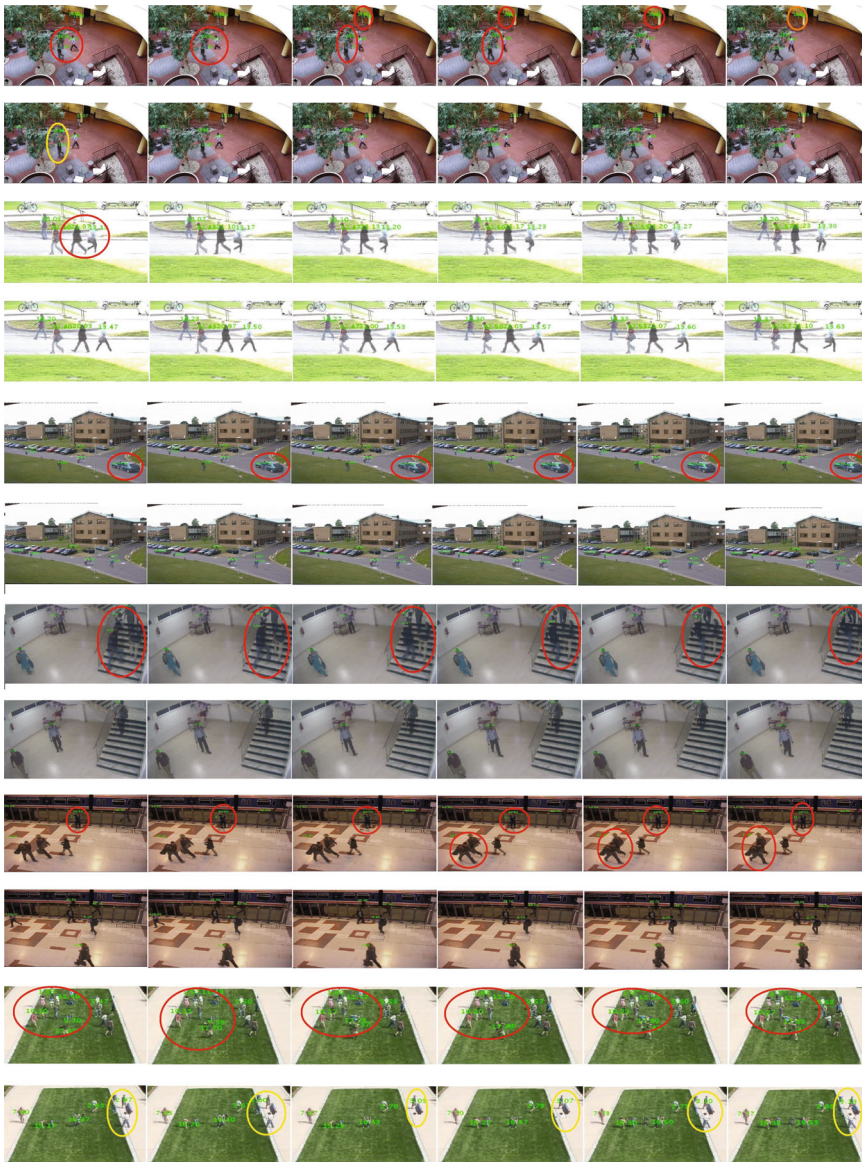


Fig. 5 First row: Video Synopsis results obtained in [32] of Video 1 (155th - 160th frames). Second row: Video Synopsis results obtained in HSATLBO of Video 1 (155th - 160th frames). Third row: Video Synopsis results obtained in [32] of Video 2 (140th - 145th frames). Fourth row: Video Synopsis results obtained in HSATLBO of Video 2 (140th - 145th frames). Fifth row: Video Synopsis results obtained in [32] of Video 3 (425th - 430th frames). Sixth row: Video Synopsis results obtained in HSATLBO of Video 3 (425th - 430th frames). Seventh row: Video Synopsis results obtained in [32] of Video 4 (271th - 276th frames). Eighth row: Video Synopsis results obtained in HSATLBO of Video 4 (271th - 276th frames). Ninth row: Video Synopsis results obtained in [32] of Video 5 (110th - 115th frames). Tenth row: Video Synopsis results obtained in HSATLBO of Video 5 (110th - 115th frames). Eleventh row: Video Synopsis results obtained in [32] of Video 6 (175th - 180th frames). Twelfth row: Video Synopsis results obtained in HSATLBO of Video 6 (175th - 180th frames). Red circle signifies the presence of collision in the resultant synopsis obtained in [32], where as yellow circle signifies the collision in resultant synopsis obtained in HSATLBO

final video synopsis obtained from the proposed scheme outperforms that of the approach used in [32].

6 Conclusion & future scope

Video synopsis is a potential tool to assess long surveillance video in a shorter retro of time by projecting multiple objects concurrently. In this paper, a hybrid approach (HSATLBO) is proposed for energy minimization problem. The hybrid optimization scheme offers an optimum object tube placement, which minimizes the energy function in terms of activity, collision and temporal consistency costs. Exhaustive experimental results and analysis reveal that the proposed scheme outperforms the benchmark schemes in terms of minimizing the overall fitness value for three different standard surveillance videos and one real generated surveillance video. Further, the proposed scheme has potential applications for home/office security, traffic control, detection of criminal activity, unusual alarming and civic protections etc.

Development of an automated and intelligent video synopsis framework remains a challenging problem. There exist several future directions which might further improve the overall efficacy of video synopsis framework. (1) To include action recognition module in single as well as multi-camera video synopsis framework to classify wide range of actions to obtain better action recognition accuracy. Moreover, it would help the synopsis generation steps to retain only the important actions in final synopsis video. (2) To design and formulate alternate multi-objective optimizations to reduce the synopsis length, number of collisions, and action loss. (3) To develop an efficient feature based object tracking techniques to form fluent activity tubes. (4) Further, to devise improved scheme for best view selection and mapping the multiple-view to a single view could be thought of as another area of extension.

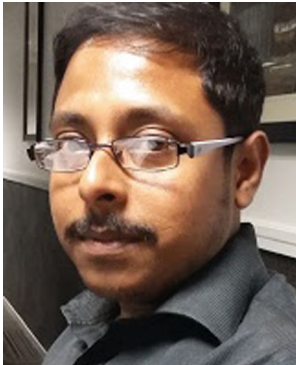
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Bishop G, Welch G et al (2001) An introduction to the Kalman filter. Tech. Rep. 27599-23175 Chapel Hill, NC, USA
2. Boykov Y, Veksler O, Zabih R (2001) Fast approximate energy minimization via graph cuts. *IEEE Trans Pattern Anal Mach Intell* 23(11):1222–1239. <https://doi.org/10.1109/34.969114>
3. Chou CL, Lin CH, Chiang TH, Chen HT, Lee SY (2015) Coherent event-based surveillance video synopsis using trajectory clustering. In: 2015 IEEE international conference on multimedia & expo workshops (ICMEW). IEEE, pp 1–6
4. Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput* 6(2):182–197. <https://doi.org/10.1109/4235.996017>
5. Ferryman JM (2001) Second IEEE international workshop on performance evaluation of tracking and surveillance (pets 2001). IEEE Computer Society
6. Ferryman JM (2006) Ninth IEEE international workshop on performance evaluation of tracking and surveillance (pets 2006). IEEE Computer Society
7. He Y, Gao C, Sang N, Qu Z, Han J (2017) Graph coloring based surveillance video synopsis. *Neurocomputing* 225:64–79
8. Hofmann M, Tiefenbacher P, Rigoll G (2012) Background segmentation with feedback: the pixel-based adaptive segmenter. In: 2012 IEEE computer society conference on computer vision and pattern recognition workshops (CVPRW). IEEE, pp 38–43

9. Huang CR, Chung PCJ, Yang DK, Chen HC, Huang GJ (2014) Maximum a posteriori probability estimation for online surveillance video synopsis. *IEEE Trans Circ Syst Vid Technol* 24(8):1417–1429. <https://doi.org/10.1109/TCSVT.2014.2308603>
10. Jamil M, Yang XS (2013) A literature survey of benchmark functions for global optimization problems. arXiv:13084008
11. Khadanga RK, Kumar A (2016) Hybrid adaptive ‘gbest’-guided gravitational search and pattern search algorithm for automatic generation control of multi-area power system. *IET Gen Transmiss Distrib* 11(13):3257–3267
12. Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. *Science* 220(4598):671–680
13. Kolmogorov V, Zabini R (2004) What energy functions can be minimized via graph cuts? *IEEE Trans Pattern Anal Mach Intell* 26(2):147–159. <https://doi.org/10.1109/TPAMI.2004.1262177>
14. Lee DS (2005) Effective gaussian mixture learning for video background subtraction. *IEEE Trans Pattern Anal Mach Intell* 5:827–832
15. Li K, Yan B, Wang W, Gharavi H (2016) An effective video synopsis approach with seam carving. *IEEE Signal Process Lett* 23(1):11–14. <https://doi.org/10.1109/LSP.2015.2496558>
16. Li X, Wang Z, Lu X (2016) Surveillance video synopsis via scaling down objects. *IEEE Trans Image Process* 25(2):740–755. <https://doi.org/10.1109/TIP.2015.2507942>
17. Liao W, Tu Z, Wang S, Li Y, Zhong R, Zhong H (2017) Compressed-domain video synopsis via 3d graph cut and blank frame deletion. In: Proceedings of the on thematic workshops of ACM multimedia 2017. ACM, pp 253–261
18. Lin W, Zhang Y, Lu J, Zhou B, Wang J, Zhou Y (2015) Summarizing surveillance videos with local-patch-learning-based abnormality detection, blob sequence optimization, and type-based synopsis. *Neurocomputing* 155:84–98. <https://doi.org/10.1016/j.neucom.2014.12.044>
19. Liu Y, Li X, Wang L, Niu Y (2014) Interpolation-tuned salient region detection. *Sci China Inf Sci* 57(1):1–9
20. Lu M, Wang Y, Pan G (2013) Generating fluent tubes in video synopsis. In: 2013 IEEE international conference on acoustics, speech and signal processing, pp 2292–2296. <https://doi.org/10.1109/ICASSP.2013.6638063>
21. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21(6):1087–1092
22. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. *Adv Eng Softw* 69:46–61
23. Mishra SK (2006) Some new test functions for global optimization and performance of repulsive particle swarm method. MPRA Paper
24. Nie Y, Xiao C, Sun H, Li P (2013) Compact video synopsis via global spatiotemporal optimization. *IEEE Trans Vis Comput Graph* 19(10):1664–1676. <https://doi.org/10.1109/TVCG.2012.176>
25. Pérez P, Gangnet M, Blake A (2003) Poisson image editing. *ACM Trans Graph (TOG)* 22(3):313–318
26. Pritch Y, Rav-Acha A, Gutman A, Peleg S (2007) Webcam synopsis: peeking around the world. In: IEEE 11th international conference on computer vision, 2007. ICCV 2007. IEEE, pp 1–8
27. Pritch Y, Rav-Acha A, Peleg S (2008) Nonchronological video synopsis and indexing. *IEEE Trans Pattern Anal Mach Intell* 30(11):1971–1984. <https://doi.org/10.1109/TPAMI.2008.29>
28. Pritch Y, Ratovitch S, Hendel A, Peleg S (2009) Clustered synopsis of surveillance video. In: 2009 advanced video and signal based surveillance. IEEE, pp 195–200
29. Rao R (2016) Jaya: a simple and new optimization algorithm for solving constrained and unconstrained optimization problems. *Int J Ind Eng Comput* 7(1):19–34
30. Rao R (2016) Review of applications of tlbo algorithm and a tutorial for beginners to solve the unconstrained and constrained optimization problems. *Decis Sci Lett* 5(1):1–30
31. Rao RV, Savsani VJ, Vakharia D (2011) Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems. *Comput Aided Des* 43(3):303–315
32. Rav-Acha A, Pritch Y, Peleg S (2006) Making a long video short: dynamic video synopsis. In: 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR’06), vol 1, pp 435–441. <https://doi.org/10.1109/CVPR.2006.179>
33. Simionescu PA (2014) Computer-aided graphing and simulation tools for AutoCAD users. Chapman and Hall/CRC
34. St-Charles PL, Bilodeau GA (2014) Improving background subtraction using local binary similarity patterns. In: 2014 IEEE winter conference on applications of computer vision (WACV). IEEE, pp 509–515
35. St-Charles PL, Bilodeau GA, Bergevin R (2014) Flexible background subtraction with self-balanced local sensitivity. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 408–413

36. Sun L, Xing J, Ai H, Lao S (2012) A tracking based fast online complete video synopsis approach. In: Proceedings of the 21st international conference on pattern recognition (ICPR2012), pp 1956–1959
37. Tian Y, Zheng H, Chen Q, Wang D, Lin R (2016) Surveillance video synopsis generation method via keeping important relationship among objects. *IET Comput Vis* 10(8):868–872
38. UMN (2006) Umn: unusual crowd activity dataset. <http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi>
39. Vural U, Akgul YS (2009) Eye-gaze based real-time surveillance video synopsis. *Pattern Recogn Lett* 30(12):1151–1159. <https://doi.org/10.1016/j.patrec.2009.03.002>
40. Wang S, Yang J, Zhao Y, Cai A, Li SZ et al (2011) A surveillance video analysis and storage scheme for scalable synopsis browsing. In: 2011 IEEE international conference on computer vision workshops (ICCV Workshops). IEEE, pp 1947–1954
41. Wang Sz, Wang Zy, Hu Rm (2013) Surveillance video synopsis in the compressed domain for fast video browsing. *J Vis Commun Image Represent* 24(8):1431–1442
42. Wang Y, Jodoin PM, Porikli F, Konrad J, Benezeth Y, Ishwar P (2014) Cdnet 2014: an expanded change detection benchmark dataset. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 387–394
43. Wang W, Shen J, Li X, Porikli F (2015) Robust video object cosegmentation. *IEEE Trans Image Processing* 24(10):3137–3148
44. Wang WC, Chung PC, Huang CR, Huang WY (2017) Event based surveillance video synopsis using trajectory kinematics descriptors. In: 2017 fifteenth IAPR international conference on machine vision applications (MVA), pp 250–253. <https://doi.org/10.23919/MVA.2017.7986848>
45. Xu M, Li SZ, Li B, Yuan XT, Xiang SM (2008) A set theoretical method for video synopsis. In: Proceedings of the 1st ACM international conference on multimedia information retrieval. ACM, pp 366–370
46. Yao J, Odobez JM (2007) Multi-layer background subtraction based on color and texture. In: 2007 IEEE conference on computer vision and pattern recognition, pp 1–8. <https://doi.org/10.1109/CVPR.2007.383497>
47. Yao T, Xiao M, Ma C, Shen C, Li P (2014) Object based video synopsis. In: 2014 IEEE workshop on advanced research and technology in industry applications (WARTIA), pp 1138–1141. <https://doi.org/10.1109/WARTIA.2014.6976479>
48. Yildiz A, Ozgur A, Akgul YS (2008) Fast non-linear video synopsis. In: 23rd international symposium on computer and information sciences, 2008. ISCIS'08. IEEE, pp 1–6
49. Yuan Y, Lu X, Chen X (2015) Multi-spectral pedestrian detection. *Signal Process* 110:94–100. <https://doi.org/10.1016/j.sigpro.2014.08.003>. <http://www.sciencedirect.com/science/article/pii/S0165168414003661>, machine learning and signal processing for human pose recovery and behavior analysis
50. Zhong R, Hu R, Wang Z, Wang S (2014) Fast synopsis for moving objects using compressed video. *IEEE Signal Process Lett* 21(7):834–838. <https://doi.org/10.1109/LSP.2014.2317754>
51. Zhu X, Liu J, Wang J, Lu H (2012) Key observation selection for effective video synopsis. In: 2012 21st international conference on pattern recognition (ICPR). IEEE, pp 2528–2531



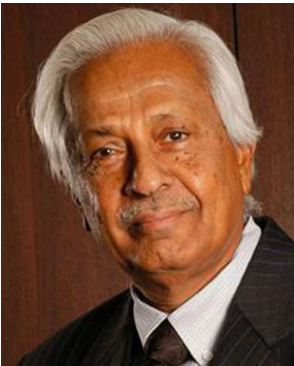
Subhankar Ghatak received the B.Sc. (Hons.) degree in mathematics from University of Calcutta, India, in 2005 and the MCA and M.Tech. in Information Technology degrees from the same University, in 2008 and 2011, respectively. He is continuing his Ph.D. degree in Computer Science and Engineering from IIIT Bhubaneswar, India. His research interest includes Advanced Image and Video Processing, Video Synopsis, Computer Vision, Visual Cryptography and Steganography.



Suwendu Rup received his M.Tech. degree in Computer Science and Engineering, from Jadavpur University, Kolkata, India. He received his Ph.D. degree in Computer Science and Engineering from National Institute of Technology (NIT), Rourkela, Odisha, India. Since 2010, he is with the Department of Computer Science and Engineering, International Institute of Information Technology Bhubaneswar (IIIT-Bh), India, and currently serving as an Assistant Professor. His research interest includes Image and Video Processing, Computer Vision, Distributed video coding, and assistive computing.




Banshidhar Majhi received his M.Tech. degree and Ph.D. in Computer Science and Engineering in the year 1998 and 2003, respectively, from National Institute of Technology (NIT), Rourkela, Odisha, India. Currently, he is serving as the Director of IIIT-Kanchipuram, Chennai, India. Since 1991, he was with the Department of Computer Science and Engineering, NIT, Rourkela, and served as the Professor and Dean (Academics). His research interest includes Image and Video Processing, Data Compression, Soft Computing, Bio-metrics and Network Security.



M. N. S. Swamy (S'59–M'62–SM'74–F'80–LF'01) received the B. Sc. (Hons.) degree in mathematics from Mysore University, India, in 1954, the Diploma in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1957, the M. Sc. and Ph. D. degrees in electrical engineering from the University of Saskatchewan, SK, Canada, in 1960 and 1963, respectively, and the Doctor of Science degree in Engineering (Honoris Causa) from Ansted University, Penang, Malaysia, in 2001. He is presently a Research Professor and the Director of the Center for Signal Processing and Communications, Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada, where he served as the Chair of the Department of Electrical Engineering from 1970 to 1977 and Dean of Engineering and Computer Science from 1977 to 1993. Since July 2001, he holds the Concordia Chair (Tier I) in Signal Processing. He has also taught in the Electrical Engineering Department, Technical University of Nova Scotia, Halifax, NS, Canada, and the University of Calgary, Calgary, AB, Canada, as well as in the Department of Mathematics, University of Saskatchewan. He has published extensively in the areas of number theory, circuits, systems, and signal processing and holds four patents. He is the coauthor of two book chapters and three books: *Graphs, Networks and Algorithms* (Wiley, 1981), *Graphs: Theory and Algorithms* (Wiley, 1992), and *Switched Capacitor Filters: Theory, Analysis and Design* (Prentice-Hall, 1995). A Russian Translation of the first book was published by Mir Publishers, Moscow, in 1984, while a Chinese version was published by the Education Press, Beijing, in 1987.

Affiliations

Subhankar Ghatak¹  · Suvendu Rup¹ · Banshidhar Majhi² · M. N. S. Swamy³

¹ Image and Video Processing Laboratory, Department of Computer Science and Engineering, International Institute of Information Technology, Bhubaneswar, 751003, India

² School of Computer and Electrical Engineering, IIITDM Kancheepuram, Chennai, India

³ Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada